Paolo Agostini

University of Padova
pagostini@tin.it

LANGUAGE RECONSTRUCTION – APPLIED TO THE URALIC LANGUAGES[*]

SUMMARY

After pointing out the shortcomings and methodological weakness of the general theory of linguistic reconstruction, the author disputes the alleged antiquity of Uralic. Proto-Uralic as reconstructed by the scholars seems to be the sum of a set of features belonging to several distinct language families. The paper examines a number of lexical concordances with historically attested languages and comes to the conclusion that the Proto-Uralic word-stock is the result of a sum of borrowings that took place from the most disparate languages: Balto-Slavic, Old Swedish, several Turkic dialects, Mongolic, Tunguz, Aramaic, Hebrew, Arabic, late Middle Persian dialects, Byzantine Greek and Latin. Yet, other languages may also come into account: Chinese, Caucasian languages as well as languages unknown in present day are possible candidates. A large number of bases of the Uralic word-stock can be easily identified by following a few phonological constraints. The linguistic features of the Uralic daughter-languages seem to show that they originated from a pidgin language spoken along the merchant routes that connected the Silk Road to North- and East-European trade. It is a well-known phenomenon that sometimes, when groups of people speaking different languages come into contact for the first time, a new restricted language system (*lingua franca* or *pidgin*) comes into being in order to cater to essential common needs. For this reason, pidgins tend to arise along trade routes. Taking into account the characteristics of the original word-stock as well as the report of the Byzantine Emperor Constantinos Porphyrogennetos, according to which the Magyars learned the language they speak from the Xazars, the place of origin of the Proto-Uralic pidgin is to be identified with the Xazarian Qaɣanate. The Xazarian Qaɣanate succeeded in gaining full control over the trade in the Caspian and Black Sea regions during the three hundred years 650–950 of our era. The Xazars are known to have established their trade posts from the Talas Valley in Kazakhstan up to the very heart of Sweden. Xazaria was one of the main trade routes in the early Middle Ages, and the most likely place where a new language might have developed. The Uralic languages are very probably the creolised offspring of an ancient pidgin that developed around the Xazarian trade posts and along the trade routes controlled by the Xazars.

KEY WORDS: language reconstruction, comparative linguistics, Uralic languages, Finno-Ugric languages, pidgin languages, creole languages, Xazars (Khazars), mediaeval trade, Silk Road

---

Section I: Prospects and Methods

Nihil est hominum inepta persuasione falsius
— Petronius (*Satyr.* 132,16)

Uralistics

Uralistics is a discipline that seems to elude any exact definition of its own essence. Although it is supposed to be the science that purports to study the so-called "Uralic" languages, Uralistics developed *de facto* and, as an obvious consequence, the way it should deal with the object of its study was not cleared yet. Uralists seldom questioned themselves about the definition of this discipline, thus getting entangled in a series of divergences, which became more dangerous, being implicit. The arising problems were not brought forward and discussed when they were *in statu nascendi*. It follows that some scholars write about the proto-Uralic people while others maintain it did not really exist; some authors investigate the social and religious life of the U ancestors whereas some others leave out of consideration such researches; some scholars try to trace the U migrations on a linguistic-genetic basis while others keep on demonstrating the absurdity of such quest. One might ask: what is Uralistics then? By definition and common consensus, Uralistics is the science studying the U languages. Nevertheless, some researchers seem to think of it as of a paleo-ethnographic discipline or as the missing key to prehistory and proto-history. For example, J. Szinnyei wrote:

> "*History begins with the first written data. Nevertheless, bygone events and the past are not completely concealed to us. With the help of anthropology, ethnology, archaeology and linguistics we can put together a number of data concerning the prehistoric period. Many of these data are much more reliable than the reports of certain ancient writers. I do not want to depreciate the other sciences. Yet, I dare declare that, among the sciences mentioned afore, linguistics furnishes the largest amount of data and the most reliable ones. I would add that linguistics is superior to the others because its data belong to more ancient ages than the data of other sciences.*" [1]

---

[1] Szinnyei, József: *A magyarság eredete, nyelve és honfoglaláskori műveltsége.* Budapest, 1919: "*A tulajdonképpeni történelem az első írott adattal kezdődik. Hanem azért az sem marad előttünk teljesen rejtve, ami annakelőtte volt és történt, amiről semmi föjegyzés sem szól; mert az antropológia, az etnológia, az archeológia és a nyelvtudomány segítségével összeállíthatunk egy sereg adatot a történelem előtti korból, s ez adatok jó része még hitelesebb is, mint egynémely régi írónak ránk maradt tudósítása. Anélkül, hogy a többit kisebbíteni akarnám, bátran kimondhatom, hogy az említett tudományok között a nyelvtudomány az, amely a legtöbb adatot szolgáltatja, és a leghitelesebbeket; s hozzátennem még, hogy abban is fölötte áll a többinek, hogy adatai sokkal régibb korból valók, mint amazokét.*"

Naturally, these are not the tasks a linguistic discipline should cope with. Besides, the concept of "Uralic people" originates from a merely linguistic issue. If ever, we could wonder what an U language is. The question might look strange at first sight, but it is justified by extant facts.

Genetic Affinity of Languages

We consider as belonging to a certain group or family a language that, from the structural, grammatical and lexical point of view shares a statistically relevant whole of facts with other languages of the same language family. W. S. Allen remarked: "*The origin of the linguistic comparison is not difficult to state: it arises in the first place from an intuitive, impressionistic recognition, requiring no depth of linguistic analysis, that two or more languages exhibit certain mutual similarities, such that one could not reasonably attribute them to chance*" (Allen, 1953). The obvious lexical similarities existing between the various U languages were already recognized by the first scholars who were familiar with them. It was only in the first half of the nineteenth century that the actual development of a recognizable comparative philology and the growth of a concept of linguistic affinity saw the light. Rasmus Rask (1787–1832) for example, showed that it was not enough to allude to the intuitive linguistic similarity between various languages as was the practice of the earlier linguistic antiquarian; he argued that these similarities must be demonstrated systematically[2].

The object of comparative linguistics is to find out whether there is a *genetic relationship* between languages. This implies that other linguistic relationships exist which are not genetic. Not every linguistic relationship can be qualified as genetic as there are many cases of related languages, the relationship of which is not genetic. There are two types of non-genetic linguistic relationships generally acknowledged. One of these is usually indicated as "typological". For the second there is no universally accepted label. Yet, since it is the result of linguistic loans, whatever this may mean, we shall call it "contact relationship" or "loan relationship". While genetic and contact relationships are defined by the way they came about, typological relationship is determined by the way it manifests itself. A genetic relationship is an affinity by origin (viz. a primary relationship), while a loan relationship is an affinity by contact (that is a secondary relationship). It is quite usual in linguistics to regard languages as related in more than one way. English and French *e.g.* are genetically related, but they are related also by loan and typologically. Genetic and typological relationship differ in the kind of correspondences by which they are

---

[2] In I-E, not merely the similarities of sounds are striking but the structure of the languages as well. The Sanskrit and Latin words for fire, *agnis* and *ignis* respectively, are not only similar in sound, but display similar changes in different grammatical cases: NomSg: Sk *agnis*, Lat. *ignis*; AccSg.: Sk *agnim,* Lat. *ignem;* Dat/Abl.Pl.: Sk *agnibhyas,* Lat. *ignibus*. Nothing similar can be found in the U languages.

established. In comparative linguistics it is usual to establish such correspondences for which it may be assumed that they are a consequence of a common origin. In typology, only equivalent terms are acknowledged as correspondent and they may or may not be of common origin.

The Concept of Proto-Language

> When anyone says "theoretically" they really mean "not really"
> — David Parnas

Lexical items in two or more languages that display similarities in sound and meaning and which appear frequently enough to exclude accidental similarities indicate that these languages are genetically related, that is they have a common origin. These relationship take the form of regular phonological correspondences, that offer the surest proof of the genetic relationship. The highest degree of validity is reached when the phonological characteristics of one or more languages can be predicted from the information contained in one member of the group. The comparative method can be applied before languages are known to be related; that is, as a procedure for discovery. Any two or more languages can be subjected to comparative analysis in order to ascertain whether or not they reflect a common heritage.

The comparative method is only possible because languages change within a framework of more or less universal principles. Many of these principles relate to phonological change which is an instrumental factor in language modification. Sound change has loomed as the dominant factor in language change: since the neogrammarians, emphasis has been placed on its regularity which, in their doctrinaire view of language, allowed no exceptions. Opponents of the view that sound change is always regular have labelled as sporadic, occurrence or non-occurrence of sound changes which do not show up as regular correspondences between divergent languages. For the proponents of regularity in change, the notion of sporadic modification implied unscientific and mentalistic[3] concepts which denied the very foundations upon which the major nineteenth century contributions to linguistics were founded. The regular and systematic course along which phonological change proceeds is often disrupted, however, by other linguistic and non-linguistic forces which play an important role in language change. The irregularities are often due to the co-existence of several sub-systems the mutual relation of which is 'irregular'. Rather than incompatibility or logical impossibility, their irregularity consists in the fact that the phenomena of the single sub-systems follow different constraints. If we regard one of these sub-systems as the rule, we shall be forced to consider irre-

---

[3] Concerned with discovering a mental reality underlying actual behaviour (Chomsky, N.: *Aspects of Theory of Syntax.* 1965. I. 4.)

gular those phenomena that take place in other linguistic sub-systems. The researcher should always grant due consideration to such irregularities.

Usually, one of the greatest methodological difficulties of the linguistic reconstruction is the impossibility to test the result of the proposed theories. We can overcome such epistemological restraint in two ways. The first way is to regard reconstruction as a language that existed for a certain period and in a certain area, although we shall be unable to tell exactly the epoch when it existed and the place where this language was spoken. The second way is to admit that such a language never existed, thus meaning that the reconstructed forms are nothing more than procedures that enable us to express the set of relationships presumed to exist between two or more languages.

Most authors usually keep an intermediate position, thus adopting sometimes the first, some other times the second position, according to the problems or criticisms they have to face. In practice, the horns of the dilemma are the following:

- If we consider the reconstructed forms as being endowed with a certain degree of reality, the theory of reconstruction becomes extremely complex but, on the other hand, it becomes possible to use the reconstructed forms of the proto-language in linguistics and other close domains.

- On the contrary, by considering the reconstructed forms as merely theoretical (either assuming that they did never exist in reality or claiming that the question is not yet solved), we remove all the methodological problems of language reconstruction, thus avoiding the criterion of authenticity. Since the reconstructed forms are situated out of time, our proto-language turns into a metalanguage enabling us to easily move from one linguistic system to another. That is to say that our proto-language will behave much like a metalanguage used for transpositions. We should then forcedly admit that there can be an infinite number of metalanguages enabling us to pass from language A to language B, and the only criteria needed to appreciate the transposition rules would be economy, simplicity, comprehensiveness and effectiveness. Thus, the best metalanguage shall be the one capable of generating the concerned languages in the simplest, most complete and most effective way. However, since the scope of our investigation has changed, there would be no reason to limit the range of comparisons to some – genetically related – languages only, thus yielding a certain degree of confusion between language reconstruction and linguistic typology. Neither would it be possible to explain why we choose one system of relationships instead of another, if it was not for its capability to generate the compared systems.

The theory of reconstruction assumes that any proto-language shares the same universals of natural languages (the principle of non-restrictivity). Therefore, proto-languages are endowed with all the characters of natural languages, thus meaning that the same universals of natural languages must have existed also in proto-languages, whatever their geographical location or chronology might have been. We have

therefore to consider proto-languages as representatives of a certain kind of historical reality, lack of which we would miss our aim. Yet, since it is impossible to prove that proto-languages represent a former phase of development of any attested language, we have to consider the consequences of this position on the methodological level.

A *caveat* is necessary. If in principle non-restrictivity is valid for every language category, it is also true that the sub-category of dead languages becomes analysable only through the corpus that can be found in the whole of the preserved texts, whatever its extension be. Although they are a particular category of languages, proto-languages participate in this characteristic since they are accessible only through the direct or indirect evidence supplied by their offsprings. Since the former phases of natural languages (including the former phases of living, dead and reconstructed languages) are available only within the limits of a corpus we have to forcedly assimilate these languages to restricted languages. This happens to a greater or lesser extent according to the size of the corpus. For example, Latin can be situated closer to non-restricted languages than Phrygian or Illyric.

When analysing a living language, by taking into account the chronological aspect of two variants, we can tell which of the two is older. This is also possible when we study the former stages of a living or dead language, if there are documents available and they are distributed along the time axis. On the contrary, it is impossible to tell the absolute chronology of reconstructed languages. In order to establish the absolute chronology of a proto-language we cannot go any further but to assume that, if the concerned proto-language really existed and if there was a common phase, the proto-language should be situated in a epoch reasonably preceding the first testimony of the attested language, thus allowing a certain lapse of time for the break-up to take place. The internal reconstruction methods can help us in situating the relation of the linguistic phenomena to each other on the chronological axis but they will never be able to supply us the relative chronology of a language.

The primary assumption underlying internal reconstruction is that many events in the history of a language leave discernible traces in its design. An examination of these traces can lead to a reconstruction of linguistic processes of change and thus to a reconstructed form of the language prior to the events which changed it. There are difficulties inherent in the internal method of reconstruction, being based, as it is, on the assumption that all complex systems originate from a formerly simpler system. Thus, internal reconstruction may yield erroneous results when chronological sequences of related events are undifferentiated by the method. Another major obstacle is *analogical levelling*. The internal method must rely to a greater or lesser extent (depending on the available data) on what was most likely to have occurred. Some phonological changes are more common than others within given phonological circumstances. Recourse to comparative data, if available, may substantiate the reconstruction. Among languages the written history of which is recent or non-existent, internal reconstruction may be the only method of gaining some understanding of their past. Used in conjunction with the comparative method, internal reconstruction may furnish information about events in related languages

which reveals the processes by which they diverged, and may divulge a certain order of phonological events by which the underlying forms are connected to the derived forms. Yet, every internal reconstruction is subject to confirmation or rejection by further comparative studies or by written historical documentation. Morphophonemic alternations reflect historical events, but in the case of a proto-language we lack any clue as to the events that originated the change. In addition, the methodology of internal reconstruction is based on the assumption that all complex systems originate from a formerly simpler system.

Be it as it may, linguistic reconstruction will *never* be able to supply us with the relative chronology of a language. The internal reconstruction methods can help us in situating the relation of the linguistic phenomena to each other on a chronological axis.

The greatest methodological weakness of the internal method of reconstruction is that it must rely to a greater or lesser extent on what *most likely* occurred. Some phonological changes are more common than others within given phonological circumstances, yet this does not mean that rare phonological changes are not liable to take place. It follows, hence, that internal reconstruction, in the case of proto-languages, is based on assumptions that, although grounded on probabilistic data, cannot be proved. Internal comparison of linguistic features may divulge a certain plausible order of phonological events by which the underlying forms are connected to the derived forms. Yet, every single bit of information gained by internal reconstruction is subject to confirmation or rejection by further comparative studies, or by written historical documentation. This, though, seems not to be the case in Uralic.

Beyond the limitations of universals imposed on us by the status and nature of the languages of the past, either attested or not, there are other limits that are a direct consequence of the methodology used for the reconstruction. Languages are (or should be) reconstructed on the basis of certain principles:

- Every human language consists of a set of more or less complex sub-systems. The complexity of a system depends on the number of its sub-systems. In turn, the complexity of a sub-system depends on the number of elements by which it is set up. It also depends on the set of relationships that might tie either the elements of each sub-system or the sub-systems themselves one to the other.

- Every natural language can be considered as a balanced mixture of simple and complex sub-systems. Languages undergo evolutive modifications. Evolution consists in a simultaneous struggle for the simplification of complex sub-systems and a slackening causing the complexification of simple sub-system. Therefore language evolution is a combination of simplification of certain sub-systems and complexification of other sub-systems. Moreover, some sub-systems are bound to be suppressed while new ones may appear.

- If we think that proto-languages really existed, we have to consider every proto-language as endowed with the same characteristics of every living language.

From these rules it follows that the elaboration of a proto-language rests on the study of the general consequences of the universals of simplification or complexification of linguistic sub-systems. In the reality, though, while it is possible to reconstruct a former, simpler sub-system from a later and more complex sub-system, the contrary appears to be impossible (unless we discover some univocal evolutive constraint enabling us to state that a certain situation proceeds from a certain other situation). From this methodological weakness as well as from the chronological characters of proto-languages mentioned above, it follows that every reconstructed proto-language is made up by a set of simpler features which are presented as synchronic.

Loan-words

Lexical similarities or lack of them between languages may in part be due to lexical borrowings. Words are often borrowed from language to language, to a greater or lesser extent, depending on the type and amount of contact between them. Even linguistic communities with fairly close-knit cultural ties often find it expedient to incorporate foreign words into their language. When cultures come into contact with one another, borrowing takes place primarily in the realm of lexical items. Some speakers may adopt loan words to show their superior learning over other members of their culture, or the lexical item in question may fill a definite need in that it is imported along with a new idea or object. In the former the loan word replaces or partially replaces a native word while in the latter it represents a new concept. A given group may borrow words which reflect a more remote period of its own cultural history, yet these are borrowed with little or no change in form. There is still a further type of intra-cultural borrowing of vocabulary items which move along the ladder of social stratification in both directions. In unstratified societies there would be some borrowing back and forth on a regional basis, but in stratified societies this would be further compounded by borrowings between social levels.

Languages in contact show a marked tendency to increase the number of equivalent units in the system. The concordances among languages that developed in strict contact one with the other is investigated by areal linguistics.

In most cases it is possible to neatly distinguish genetic and loan relationship. But there is at least one case when such distinction becomes sometimes difficult. That is the case of pidgin and creole languages. Pidgin languages represent extreme borrowing. The entire language is borrowed from the so-called lexifier languages, but considerably modified in the process.

Differences

The "scientific" period of Uralic studies initiated approx. 150 years ago, when the first lexical concordances were established, and the researchers took for granted that the Finno-Ugric languages are genetically related. It was in recent times only that cracks on the wall of the traditional doctrine begun to show, thus revealing the weakness of the foundations.

There is no consensus about the actual number of Uralic languages. The extreme minimum count, supported by many viewers, would give us no more than eighteen languages, viz. Lappish (*i.e.* Sami); five Finnic languages (*i.e.* Livonian, Estonian, Votian, Finnish and Vepsian); Mordvin; Cheremis (Mari); two Permian languages (Votyak or Udmurt and Zyrian or Komi); three Ugrian languages (Hungarian, Vogul or Mansi and Ostyak or Khanty); five Samoyed languages (Tavgi or Nganasan, Yenisey Samoyed or Enets, Yurak or Nenets, Ostyak Samoyed or Selkup and Kamas). In contrast, the most comprehensive list of Uralic languages, thus involving a separate linguistic description for each language, would include forty-six languages, *i.e.* eleven Sami, nine Finnic, two Mordvin, two Mari, three Permian, eight Ugrian and eleven Samoyed languages.

There is even less consensus on the consistence of the U wordstock. Bjorn Collinder in his *Fenno-Ugric Vocabulary* (1955) drew up a list of 277 Uralic etymologies. Yet – according to Juha Janhunen (SUSA 77; 1981) – only 140 of them are acceptable. Károly Rédei in his UEW (1988) listed 472 U etymologies, 80 of which are regarded as uncertain. Gyula Décsi (1990) presented 472 words, 36 of which labelled as uncertain.

Péter Hajdú in 1975 spoke of 1000–2000 F-U words, while Loránd Benkő wrote "*Den bisherigen Untersuchungen nach, verfügt unsere Sprache* [= das Ungarische] *über fast 1000 finnisch-ugrische Grundwörter*" (Virittäjä, 1964), a figure criticized by Aulis J. Joki (1988). For the sake of truth we have to say that Finnish scholars reduce dramatically the total number of accepted roots because they believe that Finnic is the language that better preserved the structure of P-U and P-F-U. Although there is nothing to support such a view, Finnic is regarded as the key language for linguistic reconstruction, while Ugric and Samoyedic are taken into little or no consideration.

From the very beginning of U studies, the genetic relationship of the U languages was never put in issue. Katičić correctly pointed out that:

"*It is a well-known fact that the marked interest in the genetic classification of languages prevailing in the last century and at the beginning of the present one has its roots in European nationalisms. The exact knowledge of dialects and languages was supposed to strengthen the national individuality and to align nations in 'natural' alliances*" (Katičić, 1970).

Linguistics was (is) much too often viewed not as a science but rather as a way to acquire greater "national prestige". In the case of U studies, nationalistic views often prevailed over scientific research and introduced several prejudices. For example, Károly Rédei wrote:

"*Die indouralischen Etymologien wurden kritisch gesichtet. Als Beweisstücke haben wir nur die sicheren und sehr wahrscheinlichen Gleichungen aufgenommen. Die nach unseren heutigen Kenntnissen als irrtümlich zu betrachtenden Etymologien und die auf Zufall beruhenden Gleichklänge haben wir nicht berücksichtigt. Es ist nicht ausgeschlossen, dass auch manche richtige Gleichungen der strengen Kritik anheimgefallen sind*" (Rédei, 1988).

Namely, ethnocentric views prevent scholars from accepting the idea that the U wordstock could partly be of Indo-European origin. Some other nationalistic prejudices were pointed out by T. Salminen when he wrote:

"*In practically all textbooks, the standard claim is that the Uralic family is a union of two very distantly related groups of languages, called Finno-Ugrian and Samoyed. The standard view is originally based on the classification presented by Otto Donner, the founder of the Finno-Ugrian society. While dissolving the so-called 'Ural-Altaic' unit – established by the Finnish scholar M. Alexander Castrén – he also excluded the Samoyed languages from the family. This error was later remedied but the Samoyed branch remained as 'the first branch to have left the Uralic unity'. As a consequence of this, the similarities between Samoyedic and Lappo-Finnic languages have been regarded as a result of a better survival of the Proto-Uralic heritage at the extreme peripheries of the expansive zone of occurrence of Uralic languages, thus in Lappo-Finnic languages in the West and in Samoyedic languages in the East. The standard classification continues to split the 'main' branch, i.e. Finno-Ugrian, into Finno-Permian and Ugrian, Finno-Permian further into Finno-Volgaic and Permian, Finno-Volgaic into 'Early Proto-Finnic' and Volgaic (Mordvin and Mari), and finally 'Early Proto-Finnic' into Sami and 'Late Proto-Finnic'. This practice is also unfounded, and originally based on a nationalistic Finnish view which wanted to see the Finnish language literally as the highest sprig of the 'sacred' family tree*" (Salminen, 1997).

Once Max Planck said: "A scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up that is familiar with it". Regrettably, nationalism is hard to die so that the situation of U studies probably is not bound to change within short.

The "family tree" of the Uralic languages

```
                    Proto-Uralic Language
                          4000 BC


              Proto-Finno-Ugric
                  2500 BC           Samoyedic Languages


        Finno-Permian                      Ugric
          1500 BC                          1500 BC

                    Permian
                    500 AD
                                    Proto-Hungarian
        Finno-Volgaic
          400 AD
                                                     Ob-Ugric
                    Volgaic                          1000 AD
                    600 AD

    Early Proto-Finnic
    Beginning of our Era

                    Proto-Saami


    Late Proto-Finnic
    (Balto-Finnic) 1000 AD
```
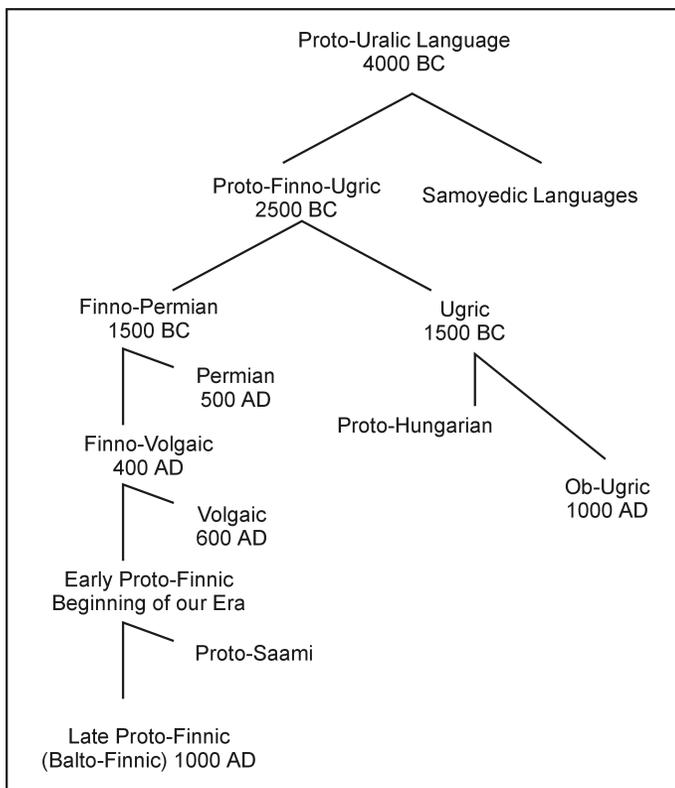
Branching off

A distinction between Proto-Uralic and Proto-Finno-Ugric becomes necessary only from a theoretical point of view, so as to keep the two — assumedly distant — "prehistoric" phases distinct the one from the other. In practice, though, there is no way to accomplish such distinction. As a matter of fact, there seem to be almost no separation between the linguistic conditions of the U and F-U ages, as P. Hajdú concisely expressed: "*...Azt is meg kell mondanunk, hogy túlságosan nagy különbség az uráli és finnugor kor nyelvi állapota között nincsen: éles határral nem választható el a kettő*" ["We have to say that there is no great difference between the linguistic situation of the Uralic and Finno-Ugric period and they cannot be differentiated in a neat way"] (Hajdú, 1978: 46). The distinction is accomplished by partitioning the fragmentary word-stock among the assumed sub-nodes of the family tree, but there is no fundamental phonological variance between the situation of P-U and P-F-U, the only difference consisting in the relics of the proto-vocabulary, as they appears in the va-

rious languages, and in some coincidences of the grammatical structure (but not of grammar). Oddly enough, by comparing the cognate languages there is no way to show systematic structural divergences that could enable a neat linguistic distinction between language phases that, according to the standard claim, ought to be spaced by not less than 2 – 3,000 years the one from the other. To a certain extent, this circumstance alone makes questionable the correctness of the theory.

The situation is further complicated by the fact that it is impossible to derive from Proto Uralic the same occurrences in all cognate languages. A concordance may appear in some of the languages while some other languages may lack it. As a consequence of this, linguists decided to hypothesize the existence of secondary proto-languages (the so-called *sub-nodes*). In spite of the fact that there was no linguistic evidence to support the choice, Uralic and Hungarian were assumed to be tied through three proto-languages, that is Finno-Ugric, Ugric and Proto-Hungarian. At the same time Uralic was assumed to be related to Finnish through at least five intermediate proto-languages: Finno-Ugric, Finno-Permian, Volga Finnic, Early-Proto-Finnic and Proto-Finnic. These proto-languages were presumed to be related to the daughter languages more or less *geographically*. If there is a concordance between a Zyryen and a Votyak word, that is if these words are cognates, they are assumed to go back to the hypothetical proto-language called "Permian". If the same concordance can be found in a third cognate language of the Finnic branch too, they postulate that the concerned word goes back to a former proto-language called "Finno-Permian". Should it be possible to match the word with a Samoyedic cognate too, then it is assumed to belong to the Proto-Uralic word-stock, and so on. Naturally, such procedure does not account for lexical loss that might have taken place in the daughter languages and depends exclusively on the presence or absence of a vocabulary item in a certain language and/ or group of languages. It follows that, should the assumption that Samoyedic was the first language to branch off be false, the whole set of relationships established among the daughter languages is also false. Hajdú confirms that "*the* [fact of attributing an] *origin from the inserted proto-languages practically depends from the quantity (or extension) of available data. Secondary proto-languages inserted between Proto-Uralic and present-day languages must be considered a category motivated rather by prehistoric than linguistic reasons*" [4] (Hajdú, 1987: 180).

A historical linguist may be considered primarily a rational descriptivist who imposes upon the data hypotheses to account for the facts. As the tools of historical studies are refined and brought into line with empirical data from descriptive linguistics, the historian must also substantiate old hypotheses or dismiss them by examining the facts in conformity with these data. Therefore, one could wonder which kind of "prehistoric" considerations might prompt a linguist to disregard lin-

---

[4] "*…die Herkunft aus diesen zwischengeschobenen Grundsprachen praktisch von dem Maß der Dokumentierungsmöglichkeit (oder ihrer Verbreitung) abhängt und man die sekundären Grundsprachen zwischen der uralischen Grundsprache und den heutigen Sprachen nicht so sehr als sprachwissenschaftlich, sondern eher als* urgeschichtlich motivierte *Kategorie betrachten muß*" (Hajdú, 1987, 180) (my italics).

guistic facts and favour non-linguistic hypotheses instead. Facts are quite simple, though. From a linguistic point of view there is no way to neatly separate any of the concerned languages to represent earlier, different phonological stages. There are no linguistic facts and data which could allow to infer that splitting of the "main branch" took ever place, with the only exception of a few phenomena that could easily brought back to different "dialects" of one and the same language. Language reconstruction is unable to reveal any intermediate proto-language, since phonological and lexical changes can be referred to P-U only. Linguistic phenomena are spread, although not uniformly, throughout the cognate languages. Cognate words may occur in languages that are situated at a great distance the one from the other and that do not belong to the same sub-group, while the same cognate words cannot be found in languages that, by definition, should go back to the selfsame intermediate proto-language. For example, Sami shows many similarities with Samoyedic notwithstanding the geographical and alleged temporal distance. Yet, it does not share any of these occurrences with the Finnic language group from which it is said to originate. The situation is substantially identical in all the concerned languages of the U family. At the same time, each single language retains individuality and peculiarities of its own.

"Archaic" vocalism and consonantism

> He that knows least commonly presumes most
> — Thomas Fuller

Vocalism always represented a major problem in the reconstruction of Uralic. Up to now it has been impossible to find general consensus on a vowel scheme. The first who tried to unsuccessfully reconstruct the vocalism of Uralic was Toivo Lehtisalo (FUF 21, 1933). In 1944, W. Steinitz published his work (Steinitz, 1944) in which he maintained that the best representatives of the vocalism of Proto-Uralic were Cheremissian and Ostyak. The Finnish scholar Erko Itkonen criticized this position (Itkonen, 1946). In fact, one of the most deep-rooted tenets of the Finnish school has been that the vocalism of the proto-language tallies with that of Early Proto-Finnic. The works of Itkonen mirror most consistently this view and he motivated the traditional conception with ideas that can be succinctly expressed as it follows:

> "*Generally speaking, the phonological structure of Baltic-Finnic roots, when compared against the bases of other F-U languages, is seemingly very conservative, i.e. "more ancient" than the roots of any other F-U language. Hence it follows that the common ancestor of the Baltic-Finnic languages and resp. its daughter-languages (Finnish, Estonian and Votiac in the first place) preserved most faithfully the original vocalism of the proto-language*".

If we took these ideas at their face value, they would be a typical example of circular thinking. When we reconstruct a proto-language by using one of its daughter languages as a key for reconstruction, the reconstructed proto-language will necessarily look very similar to the daughter language that we employed. Yet, the theory is accompanied by some details worth further consideration:

- Among all F-U languages, reconstructed I-E loans in Finnic are the most "similar" to the reconstructed forms of Proto-Indo-European.

- Germanic borrowings in Early-Proto-Finnic were preserved in the assumed proto-Germanic form.

- The vocalism of the Sami languages can be satisfactorily explained from an Early-Proto-Finnic vocalism where the quantity (*viz.* short and long vowels) was relevant.

- Both the Sami and Finnic languages are strongly conservative from several points of view. Conversely, the phonological structures of other F-U languages show the marks of great changes which they underwent. Palato-velar vowel harmony was obscured to a lesser or greater extent; final vowels faded away (apocope) and often the same lot fell to those consonants that, as a consequence of this, got in the final position; in many cases medial vowels were lost; the difference between short and geminated consonants was levelled; some consonantic clusters underwent metathesis. These changes become more evident the farther we go from the Baltic-Finnic languages.

In his reconstructions Steinitz took into consideration all the U languages, while Itkonen based his assumptions principally on the linguistic phenomena of Finnic since, as we have seen before, he prejudicially considers Finnic "the most conservative" U branch. B. Collinder (1960), Gy. Lakó (1974) and K. Rédei share the views of Itkonen. According to György Lakó though, Mordvin and Permian seem to be the "most conservative" languages from the point of view of consonantism (Lakó, 1974). In practice, this means that in most cases the reconstructions of proto-Uralic words of the UEW and MSzFgrE were made taking into account the consonants of Mordvin and Permian.[5]

One might unavoidably wonder what does the word "conservative" mean in the above context. The conservativism of a language in respect to a reconstructed proto-language cannot be defined by linguistic instruments, since we simply do not know how the proto-language looked like. As a consequence of this, conservativism is a very subjective criterion that cannot be used as an argument in linguistic research. However, many – if not all – of the above statements are false. When

---

[5] "*A konszonantizmus szempontjából a fgr. nyelvek közul viszonylag a mordvin nyelvi és a permi nyelvek a legkonzervatívabbak. Gyakorlatilag ez azt jelenti, hogy a fgr. alapnyelvre az esetek többségében olyan mássalhangzót rekonstruálunk, amilyen a mai mordvin és permi nyelvekben van, illetőleg fordítva: a mordvin nyelvi és a permi nyelvekbeli hangképviselet a hangminőség szempontjából az esetek többségében megegyezik az alapnyelvre rekonstruált mássalhangzóval*" (Gy. Lakó, *ibidem*).

comparing the Finnic roots with the original words from which they were borrowed, it is possible to realize that the Finnic word-stock was generally affected by as many phonological changes as any other U language.

Attempts to reconstruct the U proto-language

> The proof of the pudding is in the eating.
> — Miguel de Cervantes

Once the alleged affinity between Hungarian, Finnish and the others was accepted, the linguistic community working in the field started the reconstruction of the proto-language, *i.e.* the hypothetical "original" and "ancient" mother-language from which all the daughter-languages are thought to originate through a process of successive separations and diversifications. It is received knowledge that the genetic classification of the languages, purporting to show their affinity with a certain linguistic family, proceeds from the existence of meaningful phonological, morphological and lexical concordances in all or almost all the concerned languages. In fact, genetic relationship in the traditional sense can only be posited when systematic correspondences can be found in all linguistic sub-systems: vocabulary, phonology, morphology and – I would add – syntax as well. Proto-Uralic has been reconstructed according to the procedures of comparative methodology, that is for each linguistic level considered (phonemic, lexical, morphological), by postulating the (phonetic, lexical, morphological) item which should best represent, phonetically, the actual, attested items of the real languages. These attested items are therefore considered, by definition, as derived from their corresponding reconstructed ones. Linguistic reconstruction is not always easy or possible, particularly when we consider the morpho-syntactic level. This is even more true if, as in the case of U languages, there are no records old enough to assist the linguist in his reconstruction activity. Such is the peculiar nature of the realm of Uralistics: we lack any information about the early U languages and we can consider them only after the alleged U unity has long dissolved. This is also the essence of our problem, since the definition of "Uralic languages" reaches here its extreme logical limit: they appear as something that is historically unperceivable. We do not know anything of them at an earlier phase, and we learn of their existence only when they have long since settled in their present locations. The only historically perceptible reality – with very few exceptions – is present-day linguistic reality[6].

---

[6] Texts written in Hungarian are available from the XI–XII century onward. There are some sporadic relics of Finnish and Estonian going back to the XIII century, but the first available texts in these languages came about in the XVI century. The earliest Zyrien linguistic records were written in the XVI century. Literary data of other Uralic languages go back only to the XVIII–XIX century. This is why we do not know anything of the historical development of most of the U languages and linguistic comparison relies almost exclusively on present-day linguistic data.

It goes without saying that it is impossible to consider proto-Uralic as the sum of all the historically attested characteristics of the U languages. In general, there are profound differences between the U languages presently spoken. It would be inconsiderate to add up every phenomenon that can be found among the U languages in order to reconstruct proto-Uralic, since most of these phenomena are the result of external borrowings. For example, in a fine article Denis Sinor concluded that "... *I am quite certain that if from all the Uralic and Altaic languages only the Northern Tunguz and Ob-Ugric were known, no one would deny their genetic relationship. In fact Northern Tunguz and Uralic are in many respects closer than Mongol and Tunguz... there are fundamental differences between Mongol and Turkic, which shows many links with Finno-Ugric. It is clear that the relationship between any of these groups is much more involved than traditionally supposed... A meticulous study of Central Eurasian isoglosses cannot but reveal the existence of linguistic areal units which, whether or not related genetically with the neighbouring regions, share with them a number of morphological and lexical elements*" (Sinor, 1988: 738). A major problem in Uralistics therefore arises from the fact that these languages share, although asystematically, a statistically relevant number of common characters with other linguistic families. Theoretically speaking though, an "independent" language family is expected to share its main features with no other language groups. This issue raises a problem of legitimation: proto-Uralic as reconstructed by scholars seems to be the sum of a set of features belonging to several distinct families. Can we realistically consider it an "independent" language family?

In the early '50es, Bubrih claimed that "*every attempt to reconstruct the Finno-Ugric protolanguage was unsuccessful*". E. N. Setälä proved the impossibility to reconstruct a verbal conjugation (Setälä: 1899), while the Hungarian scholar J. Szinnyei maintained that the F-U languages had no common "declension" (Szinnyei, 1922). Different scholars reconstructed the "proto-phonemes" in different ways, but they were forced to give up reconstruction of the morphology of the protolanguage: as a matter of fact Uralic seems to *lack any kind of morphology*. In spite of this, some scholars went as far as presenting a comprehensive reconstruction of Proto-Uralic (Décsy, 1990).

Let us consider as an example the case suffixes. The U languages manifest to a certain extent the tendency to replace cases by postpositions or other analytic expressions. According to W. Tauli (1966: 12), "*There was naturally a time when the cases did not exist, and in the course of the gradual development of cases these languages showed a tendency to increase the number of cases. We lack data concerning the older stages of development of the U case system*". According to A. Sauvageot (Lingua 2, 36 ff.), the present case system emanates from a more or less *analytic-isolating structure* where the word was almost inflexible and the word order was fixed. Several researchers expressed the opinion that in the U languages the syntactic relations that correspond to cases were at an earlier stage expressed by means of postpositions, whereas later on these postpositions agglutinated into case

suffixes. As a matter of fact, a number – if not all – of these postpositions were probably *free morphemes,* yet the majority of the U case suffixes have not originated from postpositions by way of agglutination but some of them were borrowed from other languages during the stage of early grammaticalization, probably during the separated life of the single U languages. Let us take a quick look at the functional suffixes which are believed to go back to Uralic times.

Accusative. It is normally assumed that P-U or P-F-U had an accusative suffix **-m* (e.g. Hajdú 1981: 136; Collinder 1960: 294). Alo Raun expressed himself with caution: "*the suffix* -m *appears in Cheremis and dialectically in Lapp, Vogul and Samoyed. Correspondences or eventual traces of the same are found in Finnic, Permian, and perhaps also in Mordvin".* Ostyak has no accusative and the Hungarian *-t* is thought to be a "special" development. In his fine article, D. Sinor pointed out that "*The examination of Tunguz accusative…* [shows that] *…all Tunguz dialects use one or several of the following morphemes: -m, -b, -w, -u. The accusative of Evenki* herkan '*knife'… is* herkam. *I think that the evidence would amply warrant the reconstruction of a Proto-Tunguz *-*m *accusative which could be equated with the Proto-Uralic form*" (Sinor 1988: 714–5). Naturally, Sinor left out consideration of the fact that in the Indo-Iranic languages, from which a number of U and F-U loan-words originate, the mark of the accusative is *-m.*

Genitive. The existence of a U or P-F-U genitive is a debated question. It is absent in Permian and Ugric, but Finnish, Sami, Cheremis and Selkup do have a genitive, the ending of which is *-n.* In the article quoted above, Sinor stressed the fact that the Turkic languages have a genitive which in Old Turkic and Turkmen is *-Vŋ,* in the south-western, Oghuz languages *-Vn.* In most other Turkic languages the suffix is *-nVn* or *-nVŋ.* Moreover, it is generally agreed that the Proto-Mongol genitive, still attested in some languages, was **-n.*

Local suffixes. Both U and Altaic have a great variety of local suffixes. In the same article quoted above, Sinor showed how they equate with Turkic, Mongol or Tunguz forms.

We have to add that all these suffixes are not shared by every U language. The situation in the field of the U languages is extremely fragmentary and while a certain language may possess one or more of these suffixes, another may not. This probably means that many if not all of these suffixes were introduced during the separate lives of the single languages and are very likely the result of a late grammaticalization induced by the influence of other languages.

Reconstructing the proto-society

Le language est source des malentendus
— Antoine de Saint-Exupéry

So far, so good. Except that, once they "discovered" the linguistic relationship of the U languages, the linguists begun their quest for the proto-Uralic people. Uralistics is a linguistic discipline, but a linguistic community implies almost necessarily an ethnic community that finds one of its most evident demonstrations in the linguistic manifestation. Yet, linguistic affinity does not mean *per se* cultural or ethnic affinity. By accepting the opposite view, we would infer that the Spanish-speaking Indios in South America are related to the Spaniards both culturally and ethnically. Is the "social" aspect of language reason enough to justify a historical treatment of the group it characterizes? We can answer this question with another question: did a proto-Uralic language really exist? We know Finnish and Hungarian, Sami and Samoyedic, but in the course of our studies we never met any proto-Uralic: "*I see the horses but I cannot see the horsedom, o Socrates*".

Be it as it may, there are two ways to reconstruct a former phase of a human group characterized by a common language: we can either identify a linguistic family with a certain *prehistoric culture* or have recourse to the methods of *linguistic palaeontology*.

As far as the first method is concerned, the early researchers located the U homeland in a relatively small area. One of the principal reasons of this is to be sought in the fact that they tended to identify the homeland with an archaeological culture or region that, at that very moment, was well known. Nowadays, the Swiss pile-dwellings, the megalithic tombs of Southern Scandinavia, the late Neolithic cord-ware culture as well as the kurgan culture of South-Russian steppes, that were formerly thought to be related to the Indo-Europeans, are slowly sinking into oblivion, in spite of the noteworthy attempts of C. Renfrew and T. Gamkrelidze to partly revive the old hypotheses of V. Ivanov. In the field of F-U studies, E. Itkonen, P. Ariste and other Finnish and Estonian scholars accepted the views of R. Indreko according to which the Mesolithic Kunda-culture was related to F-U speaking peoples. On the other hand, some researchers – like Moora, Jäänits and Vilkuna – maintained that Finno-Ugrians inhabited the Baltic region many thousand years before our era. At the same time, the Neolithic and Bronze Age cultures of the river Kama were not laid aside. Since researchers were practically unable to harmonize the theory of a homeland located on the Kama river area with the theory that situates the U homeland on the Urals or in an area between the Kama and the Urals, there had been a continuous growth in the number of those who believe that the Neolithic *Kammkeramik, i.e.* comb-ware culture of the Baltic area was developed by F-U populations.

The most appropriate criticism of the notion of a relatively small homeland was given by the Hungarian archaeologist Gy. László: "*Let's assume that the* Urvölker *really lived in three distinct areas in conformity with our conceptions. Well then, but who lived in the remaining territories where findings show that they were inhabited with an equal density of population? If the linguists' hypotheses were correct, we should had found three large areas with a high density of population and three homogeneous cultures, one different from the other, while the immense territories lying amidst these areas should had been uninhabited*" (László, 1987).

Besides, we do not need much imagination to conjure how effectively we could succeed in matching a given language with a particular form of vascular handle or, even worse, with the relics of a fisher-hunter culture – not to mention the fact that several different languages may be related to one and the same archaeological culture. As an example we can cite here the Villanova culture in Italy, which was said to be the result of an immigration because of its metallic manufacts, a view contradicted by the urnfield of Pianello, showing the continuity of this culture. The end of the Villanova culture coincided with the appearance of texts written in several different languages: Etruscan, Ligurian, Retian, Picenian, etc. As a matter of fact, one archaeological culture may accumulate peoples speaking many different languages and, at the same time, an ethnic group speaking one and the same language may develop several different archaeological cultures based on their geographical context or as a consequence of contacts with other cultures.

Another fact contradicts the prehistoric theories that some linguists put forward. The alleged Uralic Age and the milieu of the U homeland were established by inferences grounded on unreliable linguistic data. The U proto-language was assumed to be very ancient so as to fulfil nationalistic expectations and prejudices. Since the proto-language included words for *bee* and *honey*, linguistic palaeontologists assumed that it came about when men domesticated the bee. They fixed an arbitrary age for bee domestication and, after tracing by ruler and compasses the hypothetical geographic centre from which the expansions of the U peoples might have started, they sought for a homeland in its vicinity. Owing to the fact that the European honeybee allegedly did not spread beyond the Urals before the assumed U age, they inferred that the homeland was located on the European side of the Urals. This is the reason why we are still talking about "Uralic" languages. Actually, a Soviet scholar realized that the "Uralic" name of the bee was a plain borrowing from Iranic (late Iranic, in addition). Bees and honey fell into oblivion, but the scholars kept on calling these languages "Uralic", a most unhappy appellation.

The Hungarian historian J. Makkay wittily remarked: "*While at the beginning of our century there was general consensus on the hypothesis postulating that the situation of I-E protolanguage was in existence at some time in the course of the 3rd millennium, it is now usual to date it back at least to the 5th millennium. Yet, generally speaking, the datation is pushed back to even earlier millennia. The situation is much the same in the case of the datation of U and F-U linguistic phenomena. The circumstance is worsened by the fact that the philologists, in the first place, did not make clear on which datation system (traditional, standard C_{14}, corrected or cali-*

brated C_{14}) *their esteemed chronologies are based. As a consequence of this, a puzzling confusion developed in the absolute and relative chronology of Indo-Europeanistics and Uralistics so that it is impossible, for the time being, to put things right*" (Makkay, 1991: 5 – my translation).

Regarding the method of linguistic palaeontology, it sets off from the assumption that semantic changes often leave relics of former meanings in the language. The lack of cognate forms of a particular word in related languages may suggest that the earlier and common stage of the language had no such word and linguistic differentiation occurred before such a word was needed to represent a cultural entity. There are, for example, no common words for *silver, gold* or *iron* in I-E, thus P-I-E probably had no such terms. It has been concluded from this that knowledge of these metals came about in the various cultures independently and at different times. The general term for metal found in Latin *aes* 'bronze', Sanskrit अयस् *ayas* 'iron' and Old English *ār* (Modern English *ore*) seem to suggest that the break-up of the common language took place at the end of the Neolithic period (that is during the last centuries of the 3rd millennium before our era).

Did "Proto-Uralians" know iron?

As far as the U languages are concerned, they *do* have a common word for *iron*, and namely:

- P-U form (as reconstructed by K. Rédei): *waśke
- Finnish *vaski* 'ore, copper, metal'
- Estonian *vaske* 'copper, bronze'
- Sami N *væi'ke* 'copper'; L *vei'kē* 'bronze'; K T *vieške*, Kld. *vie$_i$šk*, Not. *via$_i$šk* 'copper'
- Mordvin E *uśke, viśkä*, M *uśkä* '(iron) wire, (iron) chain'
- ?Cheremiss K B *waž* '(white) ore': *kərtńi-waž* 'iron ore', *ši-waž* 'silver ore' (*kərtńi* 'iron', *ši* 'silver')
- Votyak S K G *veś* : *azveś* 'silver'
- Zyryen S P *iś* : *eziś* 'silver'
- Ostyak V *wăγ*, DN *wăχ*, O *ŏχ* 'iron, metal, money'
- Vogulic TJ *küš*, KO *wəs*, P *wəš*, So. *wəs* 'lead', K *khwės* 'lead'
- Hungarian *vas* 'iron'
- Samoyedic Yur. O *jeśe*, Lj *ẃeśe* 'iron; money'; Yen. *bese* 'iron'; Tvg. *bása* 'iron, metal' Selkup Ta. *kezi*, Tur. *kēzä*, Ke. *kwez*, Ty. *kwezi* 'iron'; Kam. *baza, waza* 'iron'; Koib. *bazĕ* 'iron'; Mot. *baze* 'iron'; Taig. *beiše* 'iron'

The UEW, following Toivonen (JSFOu. 56/13:12), Aalto (UAJb. XXXI, 33), Itkonen (UAJb. XXXII,15), Hajdú and the MSzFgrE (III:675–7), tries to explain

the presence of a word for *iron* in the allegedly six thousand years old linguistic layer as follows: "*Es handelt sich um den einzigen Metallnamen aus uralischer Zeit. Das uralische Urvolk bezeichnete damit wahrscheinlich das in seinem Wohngebiet in der Natur vorkommende Kupfer, wobei es sich nicht unbedingt auf seine Bearbeitung zu verstehen brauchte. Möglich ist auch, daß das uralische Urvolk mit diesen Namen die in den Mooren und Seen vorkommende Brauneisenschichten (Limonit) oder andere auf der Oberfläche vorkommende Eisenerze der Uralgegend bezeichnete.*" (UEW: 560).

The explanations supplied are hardly credible. Even the extensors of the UEW realized that the primary meaning of the P-U word is indeed *iron*, and they tried to explain it by assuming that it meant "natural iron" (limonite). Now, is it reasonable to assume that so many daughter-languages scattered from Northern Norway across Northern Europe to Siberia and the Taymir peninsula, clustered on the North-East Baltic seaside and around rivers Ob, Yenisei, Pechora and Volga (that is in a number of environments where native iron is unfrequent), kept alive a rare, seldom used word designating a surface layer of an unknown material, thus preserving its name intact and with unchanged meaning for at least four thousand years until someone discovered what the iron was good for?

According to the SKES, "*A corresponding metal denomination exists among others in the proto-Indo-European language, for example Tokharian* wäs *'gold',* wsāyok *'gold-coloured',* yasa *'gold' (where* y- < *v–); proto-Tokharian *vas- ~ *vās, Armenian* oski *(< ? *vask) 'gold' (perhaps from the I-E root *wrd > *āus etc. 'reddish'; it is probably a very ancient international loanword that appears, among others, also in the Sumerian compound word* gusking *'gold'*" (SKES, s.v. *vaski*). The relationship of the U word with Tocharian A *wäs- ~ yasā-, ysä,* Armenian *oski, voski,* Sanskrit *ayas*[7] 'gold' was discussed, among others, by Moór (ALH. VII, 366), Aalto (ibid.), Schrader and Nehring (Reallex. I:404), Munkácsi (Ethnol. V.7, Akért. V.130) and mentioned, last but not least, by the IEW (87).

The shift in meaning of the Baltic-Finnic languages might be due to the fact that iron had always been rare in the area. Metallurgical analyses carried out on some archaeological findings in Finland show that iron was imported until the late Middle Ages from the Black Sea area.

However, the reconstructed U form *waśke* is provided with the Armenian and/or Iranic *-ki* suffix that fulfils exactly the same functions as the *-en* suffix in the English word *golden*. It means that the P-U borrowing did not refer to *gold* but to some kind of object that was "gold-coloured" or "made of gold". To explain its development in some of the concerned languages, we have to assume that the original meaning of the I-E loanword was not 'gold' *tout court*, but rather *'golden (coin)'*. In fact, money only was minted in gold, silver, copper, bronze and iron. The development of the original meaning could be posited as follows: *'golden coin'* ⇒ *'silver, copper, bronze or iron coin'* ⇒ *'silver', 'copper', 'bronze', 'iron'* ⇒ *'any meta'*.

---

[7] As we have seen before, the Sanskrit word अयस् *ayas* does not mean 'gold' but 'iron' (Benfey, 1991).

This solution, though, poses a datation problem. Gold was used since the second millennium BOE as a means of payment. Yet, the first coins in history were minted at the end of the 7th – beginning of the 6th century BOE in Minor Asia and later in Greece (Vincenzini, 1996). This indicates that the concerned U word, should the derivation prove true, was borrowed at least three millennia after the supposed "Uralic Age". Nevertheless, such explanation seems to be too far-fetched and does not explain why the meaning of 'iron' can be found in such distant languages as Hungarian and Samoyedic.

The Iranic word for 'gold' originates from an I-E base meaning 'red'. However, another I-E word could be taken into consideration as a possible source for the borrowing. Cfr. Anglo-Saxon *baso, basu* 'red, purpure', Gothic *basi* 'berry', OHG *beri* – originating from a former **bhəs-ko* [8], as in Middle Irish *basc* 'red' (IEW: 105 s.v. *bhō–s*).

Germanic loan-words in the proto-language

Libenter homines quod volunt, credunt
— Caesar, *De bello gal*. III.28.2.

In light of the current theories on the origins of U languages, though, a Germanic etymology would be unacceptable since it does not fit the hypotheses concerning the ancient homeland. Nevertheless, there are some words in the "most ancient" layer of proto-Uralic that are very likely of Germanic origin. Let us see a few examples:

- **U *aŋt3 ~ *oŋt3** 'horn' (UEW: 12), occurring in Ostyak, Vogulic and Samoyedic. Compare against Old High German *andi* 'forehead', Danish (dial.) *and*, Old Swedish *œnne < *œnde*, Icelandic *enni < *endi* 'forehead', Middle English *auntelere* 'antler; the brow-antler or lowest branch of a deer's horn' (cfr. also Old French *andouiller* 'antler; the brow-antler' and Low Latin *antiae* 'capilli demissi in frontem; hair on the forehead').

- **U *lekka-, *lejkka-** 'gap, breach, rift, slit, fissure, crack, split; to cut, split', occurring in Sami, Hungarian and Samoyedic (UEW: 244). Compare against Old High German *lech ~ leck* ; Middle Low German *lak* ; Low German *lek* ; Dutch *leck*, (adj.) *lech ~ leck* ; Anglo-Saxon *hlec* (with a spurious *h*); Old English *lec*, English *leak*; Old Icelandic *lekr* (adj.) 'not water-tight, cracked, fissured', *leki* m. 'leak', *laka* (strong verb) 'to leak, to let water pass through, to be cracked, fissured'. Among all the derivates of the I-E root **leg-* a meaning comparable with that of the U word occurs in the Germanic languages only (IEW: 657).

---

[8] According to Pokorny, the origin of this base is possibly to be sought in the I-E root **bhā–s* 'light, glow, magnificence, might'.

- **U *maća** 'fishing net' (UEW: 263), occurring in Cheremiss and Samoyedic. Cfr. against Old Swedish *max (< *masc),* Dutch *maas*; Old Icelandic *mǫskvi* > Icelandic *mǫskvi*; Danish *maske*; German *masche*, Anglo-Saxon *max, mæscr*, Middle English *maske* > English *mesh* 'the opening between the threads of a net'. Cfr. also the following cognated forms: Lithuanian *mezgù* 'to knit, tie', *mãzgas* 'knot'; Latvian *mazgs* 'knot', etc. (IEW: 746 s.v. *mezg- 'to knit, to tie; stricken, knüpfen', *mozgo- 'Knoten; knot').

- **F-U *leśe (liśe-)** 'to strip something of leaves, to hull, husk; entblättern, enthülsen' (UEW: 246), occurring in Finnish, Estonian and Hungarian. Compare against the Germanic root *les– (MHG *lesen*, OHG *lesan*, Swedish *läsa*, Gothic *lisan*, Old English *lesan*), in the meaning 'to pick, glean, clean, weed, peel, winnow, husk, hull, shell; to choose, select, sort', as well as Lithuanian *lèsti* 'to peck, pick, to choose, select, sort'. This meaning cannot be found in other I-E cognates.

- **F-V, ?F-U, ?U *lowkk3** 'hole, orifice, opening, cave; Loch, Öffnung, Höhle' (UEW: 252), occurring in Finnish, Estonian, Cheremiss, Hungarian and Samoyedic. Compare against the common Germanic base from which the following originate: Gothic *us–luk* 'opening, hole, cavity'; Swedish *lock* 'fastener, lock; cover, lid'; MHG *loch*, OHG *loh* 'fastener, hideaway'; German *Luke* 'hole or opening in the wall, ground or roof (which can be closed with a flap)', English *lock*. The Germanic base belongs together with other cognated words to the I-E root *leug– 'to bend, wind, turn; biegen, winden, drehen' like *e.g.* Lithuanian *lùgnas* 'pliable, flexible; biegsam, geschmeidig' and Latin *luxus* 'luxated, sprained; verrenkt' or even German *Lauch* 'the plant with leaves turned downwards' ⇒ 'garlic'. The meaning of 'opening, hole (that can be closed/fastened with some kind of device)' developed in the Germanic languages only. This means that the borrowing could take place from a Germanic language only.

- **U *mol3** 'piece, fragment, crumb, little bit; to break, crack, smash, crumble', a *nomen-verbum* occurring in Sami and Samoyedic (UEW: 278). Cfr. I-E *mel-, smel- melə-, etc. 'to crush, break, hit, beat, grind' (IEW: 716). The U forms can possibly be traced in Norwegian *smola* 'to crush', Old Swedish *smola, smula, smule* 'piece, hunk, morsel, bit', Old Icelandic *moli* of same meaning, etc.

By examining these loan-words we can draw *prima facie* two conclusions: 1. Proto-Uralic borrowed the Germanic words from historically attested languages, that is in *historic* and not in *prehistoric* ages; 2. The standard claim of an Uralic homeland is false.

Linguistic palaeontology and homeland

Returning to the question of linguistic palaeontology, conclusions concerning the origin of the I-E people based on cognate forms such as those for *birch* and *beech* have suggested to some observers that their homeland was located in the zone where these trees grow. Deductions of this kind are risky, however, for at least three reasons: there is no assurance that these words formed part of the vocabulary during the period in which the homeland is sought; lexical items are easily borrowed and these words may have originated in one branch of the language family and at an early time spread to the others; in deductions of this sort one must also consider the climatic factor – the geographical limitations of these trees today may not match those of five or six thousand years ago, a period corresponding to the alleged age of P-U.

Semantic considerations have also a role in inferences of this nature. Lexical items change in meaning, making it difficult to determine if a particular name referred to the same physical object in prehistoric times as it does now, and various dialects may contain the same lexical item with different meanings. The wide-range researches carried out in a masterly way by H. Krahe in the I-E field do not seem to have yielded much more success than the early and naïve quest of I. Guidi for the Semitic homeland. All the I-E comparisons of linguistics palaeontology in restricted semantic fields came to a deadlock or, what is even worse, they were and are constantly modified as a consequence of the new discoveries made in the fields of botany and zoology.

The hypotheses concerning the names for *horse* in the field of I-E studies are the best example. One of the most popular theories concerning the I-E homeland is based on the knowledge, use and name of the horse, in particular of the domesticated horse, in the steppes between Carpathians and the Volga in the 5th – 3rd millennium BOE. For example, W. Meid tries to establish the I-E homeland almost exclusively on the base of the occurrence of bones of domesticated horses: "*die frühe Sitze der Indogermanen eher im westlichen als im östlichen (asiatischen) Domestikationsgebiet des Pferdes anzusetzen*" (Meid, 1989). On the other hand, though, J. Untermann claims that "*the existence of the word* *ekwos *proves only that before the break-up of proto-Indo-European there was an animal that had this name and that the same animal was known even later. On the contrary, there is no evidence that the Indo-Europeans knew the domesticated horse or that they made use of domesticated horses. Every attempt to discover the Indo-Europeans among the prehistoric finds with the help of the domesticated horse is a vain hope*" (Untermann, 1985). The question under debate is whether the P-I-E reconstructed form *ekwos meant "horse" or definitely "domesticated horse". Should the latter meaning be proved, it could be theoretically possible to identify the geographic area and the archaeological culture where I-E developed. Nevertheless, much data concerning the I-E names of domesticated animals show that these names were simply handed down from wild species to their domesticated progeny. The I-E name **ekwos*

existed probably even *before* the horse was domesticated somewhere in the steppes of Southern Russia. Not to mention the fact that as far as the first millennia of horse domestication are concerned, it is often practically impossible to differentiate the osteological findings of wild and domesticated horses.

As far as the U languages are concerned, Pekka Sammallahti confirmed what we already knew, that is: "*in spite of many attempts with different approaches, it has not been possible to identify a restricted homeland from where the spread to the present areas would have begun.*" [9]

Absolute and relative chronologies

> If the facts don't fit the theory, change the facts.
> — Albert Einstein

A gross error of the Uralists was to adopt acritically the methods and technical terms of the Indo-Europeanists working at linguistic reconstruction. An investigation into proto-Indo-European or proto-Semitic can be prehistoric indeed, since written records of these languages go back in time to about four thousand years before the common era (but then only in limited cultural and geographical areas of the Middle East). Yet, the earliest relics of an U language go back to the 11th century of our era only. The research, in the case of U, ought not be *pre*–historic but rather *pan*–historic, thus it should be situated out of the axis of time.

In analysing a living language, we can verify the chronological aspect of any two variants thus identifying the one that is more ancient. This is also possible in the study either of a dead language or of the former phases of living languages, whenever we are in possession of evidence distributed along the axis of time. We are conversely unable to specify the chronology of a reconstructed language as there is no way to obtain its absolute chronology (unless we can have recourse to non-linguistic methods like archaeology). In establishing the absolute chronology of a proto-language we cannot go any further than stating that *if* such a proto-language existed and *if* there was a common period when it was spoken, this should be posited fairly long before the first testimonies of the attested languages in order to allow time enough for the diversifications to take place. Practically, we have a *terminus ante quem* (*i.e.* the proto-language was spoken *before* the daughter languages came to light), while it is impossible to fix any rough or approximate *terminus post quem*, since we cannot demonstrate how long it took for the proto-language to differentiate and break up into two or more branches.

In the case of relative chronology the situation is even worse, since we can rely only on the methods of internal reconstruction that might be of help in localizing the phenomena by encompassing their reciprocal relationship on the axis of

---

[9] Sammallahti, Pekka: *"Language and roots",* Papers of the 8th CIFU, 143–152.

time. Now, these methods are based on the axiom that every complex system originates from a former, simpler system. This limitation is due to the fact that *we are unable* to reconstruct a more complex system from a simpler one. Actually, this position is inconsistent with the general theory of reconstruction, since the latter assumes that any proto-language shares the same universals of natural languages (the principle of non-restrictivity). Natural languages are bound to change both by complexifying or simplifying their systems. The words we are able to reconstruct, though, are one-way only.

The age of Proto-Uralic

The question of the age of the U proto-language, assumed to be "at least" six thousand years old (sic), was never an issue. The researchers were so struck by the great divergences of the U daughter languages that they assumed, more or less tacitly, that this phenomenon was due to the great antiquity of the proto-language. No one ever argued about the assumed "great age" of the U proto-language, since they thought that the time depth separating each single language from the others obscured any salient feature of relationship. As a matter of fact, the U languages show no one of the regularities the Indo-Europeanists are used to. They do not share, or share just to a very limited extent, personal pronouns, numerals, case suffixes, verbal declinations, etc. As a consequence of this, no one ever doubted that the assumed time depth might be true. At the very most someone added or subtracted one or two millennia to the assumed age, but no prejudice-free attempt was ever made to check if the "great antiquity" stands closer examination. P. Hajdú explained the reasons for choosing such chronology as follows:

> "*The chronology of the U and resp. F-U epoch can be established by taking into consideration several factors. – 1. We can ascribe an absolute chronological value to the fact that all the trees of the taiga (northern conifer forest), which are peculiar to the proto-language, did not come in touch with the eastern expansion border of the elm-tree before the 6th millennium BOE. In other words, the F-U epoch cannot be posited before the 6th millennium. – 2. We have to take into account the fact that the oak-tree expanded toward the Urals more slowly than the elm-tree, and it appears on the area of river Pechora in the second half of the Middle Holocene (approx. 3rd – 2nd millennium BOE). From the different Permian and Magyar names for the oak-tree it is possible to infer that the break-up of the Finno-Ugrians took place before the expansion of the oak trees in the area between Kama and Pechora. – 3. In the F-U languages there are some I-E loanwords. The borrowings took place from proto-Indo-European or from a language very close to it, while some other loan-words originate from Indo-Iranic… Owing to the fact that the Samoyedic languages lack these borrowings, we assume that these words were borrowed after the break-up of proto-Uralic during the F-U epoch. We cannot say exactly when the break-up of proto-Indo-European took place. In determining*

it… we do not think to be far from the mark if we repute that the I-E unity – between the 6th and the 4th millennium BOE – became loose enough for some groups to start the development of a language of their own. If we accept such dating and harmonize it with dendrological chronology there is ground enough to maintain that the cohabitation of the ancestors of the Finno-Ugrians and Samoyeds broke up between the 6th and the 4th millennium BOE. – 4. The relative chronology established by linguistics tallies with the archaeological ascertainment that the Neolithic culture, attributed to Finno-Ugrians, which developed between river Kama and the Ural mountains, at the beginning of the 3rd millennium did not yet expand to the other side of Volga. This culture begins its large-scale westwards spreading only at the end of the 3rd millennium BOE. The assumedly F-U population invades the area between Volga and Oka rivers, thus subjugating and driving away or incorporating the natives. As a consequence of this process, at the turn of the 3rd and 2nd millennium, in the Neolithic age, it settles in Northern Europe from Volga. The expansion from the Kama area towards West-North West can be posited in the 3rd millennium BOE with the F-U break-up. – 5. Lexico-statistical methods were applied so as to periodise the U prehistory… Presently there are controversies as concerning the serviceability of lexico-statistical methodologies in order to determine the break-up of languages… – 6. To establish the absolute chronology of the U history we can have recourse to the C_{14} method insofar archaeological findings which can be undoubtedly attributed to the Finno-Ugrians should ever come to light" (Hajdú, 1988: 10 – my translation).*

Dendronyms

Hajdú points to the question of the U names of conifers and elm-tree. The names for the taiga trees which – in Hajdú's view – are important in order to locate the U homeland are: Picea excelsa: **kus3 ~ *kos3* ; Pinus cembra: **soks3 ~ *saks3 ~ *seks3* ; Abies: *ńulk3* ; Larix: *näŋ3*. The name of the elm tree (Ulmus), reconstructed as **śala*, is the only common F-U name for a broad-leaved tree[10].

The famous Swedish naturalist Carolus Linnaeus (Carl von Linné, 1707–1778) claimed that plant names are extremely variable "*because in the same region there are many different names for the same plant; for in a single village, and even within the same house the inhabitants do not agree on the names of the selfsame plants, it is clear to everybody how difficult it is to collect dialectal plant names*"[11].

---

[10] If it depended on the species mentioned, the U homeland could be located anywhere on the lower Alps (!)

[11] "*Cum in una eademque regione tot tamque diversa sint nomina eiusdem plantae; cum in singulo pago, immo in una eademque doma saepe dissentiant in nominibus earundem plantarum incolae, patet cuique, quam difficile est vernacula nomina colligere.*" (C. Linnaeus, *Philosoph.Bot.* §324 in Critic.Bot.)

It is quite important to note that, on the contrary, in the U daughter languages these plant names show very little or no variation in meaning, as would be typical of languages the break-up of which took place not too long ago.

The same point was raised also by the Hungarian historian E. Molnár in 1955: "*I maintain that palaeolinguistics is a metaphysical method since it is based on the hypothesis that plant and animal species many millennia ago lived there where we find them today, without any change, and that the meaning of the words is the same in the present as it was many millennia ago, without any change*" [12] .

In order to date Proto-Uralic, Hajdú claims that "*the trees of the taiga, the names of which are peculiar to the proto-language, did not come in touch with the eastern expansion border of the elm-tree before the 6th millennium before C.E.*". This might be true, yet the problem is the same one raised by J. Harmatta in the early 70's: "*What if these tree names were loan-words?*". Indeed, the "Uralic" names for these trees could be borrowings from the Altaic languages. For example, D. Sinor wrote: "*L. V. Dimitrieva's thorough study of Altaic tree names (1972) failed to trace one single tree name common to Turkic, Mongol and Tunguz… Even more disturbing is the fact… that such common tree names as those referring to various kinds of conifers are identical in Tunguz and Finno-Ugric*" (Sinor, 1988: 737). Even the "Finno-Ugric" name of the elm-tree, *śala*, is very likely an I-E loan-word.

Owing to the fact that they do not belong to the independent U word-stock, the names of the taiga trees could have been borrowed at any point in time. Thus, they cannot be of any use in order to establish an absolute chronology. Taking into account this detail, it is *impossible* to infer, on the basis of these dendronyms only, that the U proto-language is 6,000 years old. The second point raised is the name of the oak-tree that, being different in Permian and Magyar, should enable – in Hajdú's view – to posit the break-up of the Finno-Ugrians in the 2nd millennium BOE. In order to attain his scope though, he rebutted – certainly not on phonologic or semantic grounds – the concordance with the Permian word for "forest, fir-wood". The Hungarian name for the oak-tree (*tölgy*) is a late Iranic loanword (cfr. Eastern Ossetic *tūldz*, Western Ossetic *toldže*), and the same origin can be assumed for the Permian word.

The earliest pioneers in the U and F-U studies realized very soon that in these languages, beyond a noteworthy amount of Altaic *Sprachgut*, there is a num-

---

[12] "*Én azt állítom, hogy az életföldrajz módszer metafizikus módszer, mert alapja az a feltevés, hogy a növény- és állatfajták évezredekkel ezelőtt is változatlanul ott éltek, ahol ma és hogy a szavak értelme ma is változatlanul az, ami évezredekkel ezelőtt volt. […] Hajdú elvtárs például munkájában 12 növény- és állatfajta földrajzi elhelyezkedésére állapítja a finnugorok európai eredetének elméletét. Ezek közül azonban csak kettőnél tud megállapítani bizonyos helyváltoztatást, kettőnél feltesz ilyet, de nem tudja, hogy a helyváltoztatás hol és mikor játszodott le, nyolcnál pedig, tehát a nagy többségnél, abból indul ki, hogy ezek ma is ott élnek, ahol négyezer évvel ezelőtt. Ehhez járul, hogy bár teljesen lehetetlen megállapítani, hogy ezek a növényeket és állatokat jelölő szavak négyezer évvel ezelőtt mit jelenthettek, — abból indul ki, hogy jelentésük ma is ugyanaz, mint volt négyezer évvel ezelőtt*" (Molnár, Erik: *"A magyar őstörténetről"*, in: Nyelvtudományi Értekezések 5. Budapest, 1955).

ber of I-E loanwords. The fact that these borrowings should have taken place from Proto-Indo-European (or from a language very close to it) is still far from having been proved. As a matter of fact, many of these loan-words – like Finnish *nimi* 'name', *suoni* 'sinew, vein', *vesi* 'water', *tuo-* 'to bring', *mehiläinen* 'bee', *orpo* 'orphan', *porsas* 'pig' and so on, might belong to much later linguistic phases of I-E cognate languages.

For example, the U word ***mura** 'bramble, huckleberry' was compared to Armenian *mor, mori* 'dewberry, bramble', Greek μῶρον 'dewberry, bramble', Latin *morus –us* f. (> Low Latin *mora* f.) 'dewberry, bramble'. It is important to stress the fact that this root does not occur with the same meaning in the Indo-Iranic branch of I-E. It follows that the borrowing could have taken place from Greek, Armenian or Latin *only*.

The U word ***śala** 'elm-tree' mentioned before was compared to Latin *salix* 'willow, osier'; etc. Naturally, comparatists assumed that the borrowing went back to times immemorial when Latin, Greek and Armenian did not yet exist. Thus the Latin or Greek word comparisons purported only to show that the word concerned might be a borrowing from an early P-I-E form. But let us take a closer look at the Latin word *salix*. It originates from I-E **sal-* 'grayish; schmutziggrau'[13] and developed the meaning of 'willow' in the westernmost I-E daughter languages only. Beyond Latin, we can find it in Celtic (cfr. Middle Irish *sail*, Cymric *helyg-en*, Old Breton *Salico-dūnon* [place-name], Gallic *Salicilla* [place-name]) and Germanic (OHG *sal(a)ha*, MHG *salhe*, NHG *Sal-weide*, Old Icelandic *selja* and Anglo-Saxon *sealh*). No trace of continuity of the I-E root in the sense of 'a species of trees' was ever found in the Indo-Iranic languages, thus meaning that if U **sála* is an I-E loanword, the borrowing took place from a western I-E language.

Latin and Greek loan words in the proto-language

One might except that six thousand years ago the U homeland was possibly situated next to the homeland of the ancestors of the Teutons, Celts and/or Latins. This is not the case, though. We shall illustrate our point with the help of another "Uralic" word, that is

- **U *är3** 'year' (UEW: 26), cfr. Votyak *ar* 'year'; Zyrien *ar* 'autumn'; Vogulic *oårėm, årėm* 'time'; Samoyedic Yur. *ŋäerū* 'autumn', Yen. *narra*, Tav. *narro*; Selkup Ta. *ărà*, Kam. *ere* 'autumn', Koib. *ire*, Mot. *iriu*, Karag. *iriu*. We can compare this word against Latin *aera* (plural of *aes, aeris* 'brass, bronze, coin')[14]. In particular, the Latin word *aera* was used to de-

---

[13] Cfr. also Latin *salvia* `sage', a herb the leaves of which also have a greyish colour.

[14] The Romans called *aera* the bronze markers or token money used for reckoning. It also meant 'given number', in particular the number from which they started counting, whence the meaning of 'account' in bookkeeping (> Lat. *aerarium* 'the Treasury').

signate *a fixed point of time or a date starting from which the years were counted*. Only later the word came into use to indicate a series of years (> Spanish *era*, Italian *era*, French *ère*, English *era* and so on).

There is another possible source for the borrowing, though, namely Old Norse *ár* 'year', with a change in vowel quality (from back vowel to front vowel). This brings us back to the question of the Germanic loan-words in the Uralic languages.[15]

The Latin and Old Norse words are certainly not as old as U is assumed to be. What is more important, we can hardly consider these words a sound-alike or a fortuitous coincidence only. In fact, the "most ancient" layer of proto-Uralic has *a number of words* the most likely origin of which is Latin (Low Latin) and Greek (Byzantine Greek). We shall detail hereunder just a few of these loan-words.

- **U *aśkel3, aćkel3** 'step' (UEW: 19), appearing in Finnish (*askel*), Estonian (*askelda*), Mordvin (*aśkiľa, eśkiľa, iśkila, aśkoľa*), Cheremiss (*aškəl-, oškəl-*), Votyak (*učkil*), Zyrien (*vośkol, ośkel*), Vogulic (*ūsəl, ūsil*) and Samoyedic (Selk. Ta. *aasel*). The word originates, beyond any reasonable doubt, from Greek σκέλη 'step' (cfr. Modern Greek σκέλι ~ σκελιά) in its Hellenistic and Byzantine form ἀσκελιὰ, *ἀσκέλη 'step; long step; step made to walk over something' (cfr. Modern Greek ἀσκελιά of same meaning).[16]

- **U *jama-** 'to fall ill' ⇒ 'to die' (UEW: 89). The U base can be found in Sami, Mordvin, Cheremiss and Samoyedic. The word is a borrowing from Greek ἰῶμαι 'to fall ill' (cfr. also ἴαμα 'medicine, remedy, healing').

- **U *lampa** 'snow-boots, arctics', occurring in Sami, Samoyedic and Zyryen (UEW: 234). Compare against Greek κλάπαι 'crutches; clogs, wooden shoes' and German *klumpen* 'clogs, wooden shoes'. The U word could be either a Germanic or Greek loan that underwent the usual simplification of the initial

---

[15] The connection with the I-E base **i̯ēro- : i̯ōro- : i̯əro-* 'year, Summer' (IEW 296–7 s.v. *i̯ē-*) is not possible. In fact, the trailing *j-* belongs organically to the I-E base (cfr. Avestan *yārə-*, Gothic *jēr*, Anglo-Saxon *geār* 'year', Old Slavonic *jara* 'Spring', etc.). If it was a borrowing from this root, the *j-* would had been preserved in the U languages.

[16] The ultimate origin of the Greek word is the I-E root **(s)kel–* 'austrocknen, dörren' (IEW, 927), whence Greek σκέλλω 'to dry up, essiccate; make thin' > σκέλος 'thigh, leg; bone' (from which originate words like σκελέαι 'article of clothing used to cover the thighs or legs' → 'trousers' and σκελε-τόν 'dry, embalsamed corpse; mummy; skeleton'). The Greek verb was connected to a notion of strength, cfr. the adverb ἀσκελὲς ~ ἀσκελὲως 'strengthfully, powerfully, inesorably' as well as the classical sentence ἐπί σκέλος ἀνάγειν 'to withdraw while looking the enemy straight in the face'. This is probably the reason why the Byzantine word ἀσκελιά, *ἀσκέλη 'step, long step' became a military term, the meaning of which was very likely '(to march with) long, strengthful step(s)' and, as such, was borrowed in Arabic as عسكرية ['askariya] 'army' and عسكر ['askar] 'soldier' (with the rhotacism of post– and intervocalic *l* which is typical of Late Byzantine and Mediaeval Greek, cfr. Classical Greek ἀδελφός > Byzantine Greek ἀδερφός 'brother', Low Latin *fanale* 'lamp, light' > Byzantine Greek Φανάρι (a quarter of Constantinaples named from its lighthouse).

consonantal cluster, but – if it was Greek – we should assume that it underwent secondary nasalization. The Hungarian dialectal word *klumpa* 'wooden shoes, clogs' is a modern borrowing from German.

- **U *kur3** 'knife' (UEW: 218), cfr. Finnish *kuras*, Sami *korr*, *gūrra*, Samoyedic *kar*, *kōru*, *kuru*, *kolu*, *kōrdo*, *kuro*, *kura*, *kurru*. The UEW suggests that this base might be an "ancestral, international loanword" *("uraltes Wanderwort")* and compares it with Yukagir *kiŕe* 'stone weapon'; Turkic *qïr-* 'to scrape, to strip hair', *qïrq-* 'to shear (sheep)'; Osmanli Turkic *qïrgi* 'knife, blade'; Mongol. *kïrɣa-* 'to shear' > Tung. *kirg-* 'to shear'; Hettite *kuruzzi-* 'tool for cutting', Gothic *hairus* 'sword'. Taking into account both the meaning and phonological form of the U cognates, the only language from which the U word could have been borrowed is Greek κουρίς 'razor, knife'.

- **U *kur3** 'curved, bent, askew, awry; to bend, curve, to make crooked, warp'. Cognates of this base can be found in Votyak, Zyrien, Ostyak and Samoyedic (UEW: 220). Cfr. the Latin adjective *curvus* of same meaning (but see also Albanian *kurrús, kërrús* 'to bend, to bow', Slavic *kriv* 'crooked, warped', etc.).

- **U *kama** 'fish-scale, scale, (crust, bark, rind, shell)' (UEW: 121), occurring in Finnish, Estonian, Cheremiss, Votyak, Zyrien, Vogulic, Hungarian, Samoyedic. Cfr. Latin *squama* 'fish-scale, scale'. The U base underwent the typical simplification of the initial consonantal cluster.

- **U *num3** 'the upper, superior, heaven, god' (Ostyak, Vogulic, Samoyedic) (UEW: 308). Cfr. Latin *numen* 'hint, will, command (in particular: heavenly command); heaven, god' (cognate of Greek νεῦμα 'hint, will, command').

- **U *kewe** 'female, she-animal', occurring in Livonian, Sami and Samoyedic. (UEW: 152). Possibly a borrowing from Greek κυέω 'to conceive, to be or become pregnant, gravid'. If so, we can assume that the in the U languages its meaning developed from 'pregnant female' to ⇒ 'female'.

- **U *kanta** 'burden, load, charge' (UEW: 124). Cognates of this base can be found in Finnish, Estonian, Sami, Mordvin, Cheremiss, Ostyak, Vogulic and Samoyedic. Cfr. Post-Biblical Hebrew קנתל *qanthel* 'pack-saddle; pack-baskets; panniers on both sides of a pack saddle' < Greek κανθέλια of same meaning < Greek κάνθον 'pack-animal; ass, donkey'.

- **U *piľm3** 'dark; to grow dark'. Cognate words can be found in Finnish, Estonian, Votyak, Zyrien and Samoyedic (UEW: 381). Cfr. Greek πελίωμα 'black-and-blue spot, ecchymosis; something becoming/turning blue or black; dark-coloured; the colour of a dead body' < πελιός 'farblos, bleich, grauschwarz, schwarzblau', going back to I-E **pel–* (IEW: 804). This word shows several distinct phenomena, and namely:

  1. *Palatalisation*. Greek *l* was palatalised into Uralic *ľ*.

  2. *Change of vowel quality* in the first syllable, *i.e.* Greek ε > Uralic *i*. We shall discuss this issue later on.

3. *Loss of medial vowels*. The Greek diphthong –ιω– disappears in U: *pel–ío–ma > piľ–∅–m3*. possibly as a consequence of the U stress falling on the first syllable.

4. *Loss of the palatalised consonant* in Balto-Finnic, cfr. *piľm3* > Finnish *pi–∅–meä*, Estonian *pi–∅–me*, etc.

The **F-U** word ***pilwe ~ *pilŋe** 'cloud' (UEW: 308) very likely originates from the same I-E root, *i.e.* *pel– 'in Ausdrücken für unscharfe Farben wie »grau, fahl«' (IEW: 804), possibly from Greek πελιός ~ πελλός ~ *πελϝός ~ *πελνός 'grey'. Unluckily there is no written evidence to confirm the meaning of 'cloud' of these words, but it might be due to a sense development that took place in the U languages from a general meaning of 'grey (weather)'. One might wonder why palatalising did not occur in this case. A likely hypothesis could be that the borrowing took place from πελλός, where a geminate *lambda* occurs, thus the geminate might have prevented palatalising. Yet, other variations can also come into account.

There is another problem, though. The Finno-Permian word ***pelma, *peľma** 'dirt, filth' ('ashes'), occurring in Finnish, Zyrien and Votyak (UEW: 728), is very likely related to the Greek word mentioned afore, *i.e.* πελίωμα 'any dark-coloured thing'. From a semantical point of view, the meaning of the Greek word overlaps both the meaning of 'dark' and 'dirt' (or 'ashes') of the U words. The vowel in the first syllable makes the difference: in the one case we find an /i/ and in the second case an /e/, the latter seems to be closer to the older Greek pronunciation. Should we infer from this that Steinitz was right when he proposed a vocalism based on alternances? Is the divergence due to multiple borrowings that took place at different points in time, *i.e.* the Finno-Permian word first, when the Greek pronunciation of the letter was still *eta* and the Uralic word at a later point in time (!), when it was already pronounced *ita*? In *extrema ratio* we could have recourse to the hypothesis of a semantic dichotomy producing the divergence. Yet, none of these hypotheses are likely to be true. The phenomenon of alternance of several roots is not unique. In the Etymological Dictionary of Uralic (UEW) we can find several words that, although very similar in meaning and phonic form, have been put under different entries. As an example we could cite the following:

U ***lapa** 'plainness, surface; flat surface of sg, plane surface' (UEW: 236);

U ***lappa** 'clasp' (UEW: 236);

U ***lapp3** 'smooth, plain' (UEW: 237);

U ***lapta** 'flat, plain' (UEW: 238);

U ***lap3** 'oar, paddle' (UEW: 238);

U ***l8mp3** 'surface (palm of the hand or sole of the foot)' (UEW: 255).

It goes without saying (even if it seems that no uralist realized it up to now) that all of these U cognates originate from an I-E root, that is **lep-, *lop-, *ləp-* 'to be plain; palm of the hand or sole of the foot; shoulder blade, shovel, paddle and similar' (IEW: 679). Cfr. Kurdistani *lapk* 'instep; wrist; paw'; Gothic *lōfa* m., Old

Icelandic *lōfi* m. 'flat hand; palm of the hand'; MLG and Middle English *lōf* 'wind side', *i.e.* 'a large paddle, with the help of which it is possible to keep a ship to the wind'. See also the forms with Ablaut and gemination: OHG *laffa* f.; MHG and dial. *laffe* 'plain hand, shoulder blade', *lappo* 'plain hand, oar, paddle'; Norwegian and Swedish *labb* m.; Danish *lab* 'paw'; Icelandic *löpp* f. of s.m., and so on. Lettish *ļepa* f. 'paw', *lãpa* 'leaf', *lâpsta* 'shovel, spade, shoulder-blade'; Lithuanian *lãpas* 'leaf', *lopetà* 'shovel'; Old Prussian *lopto* 'spade'; Russian *лána* 'paw, sole', *лопáта* 'shovel', *лопáтка* 'shoulder blade', *лопáтина* 'steer-paddle'; Polish *łápa*, *łába* 'paw, sole', *łápiniec* 'bear paw'; Czech *tlapa*, Slovak *dlaba* 'sole'; Old Bulgarian **лопата** 'driving shovel', etc.

The reconstruction of the U wordstock made by Rédei shows the contemporaneous presence of a large range of synchronic variants, exactly as we would expect from the continuum of a set of isolects (*i.e.* systems of idiolects of individual speakers).

- **U *käl3** '(coagulated) blood' ('clot of blood', 'blood stain') (Sami, Ostyak, Vogulic, Samoyedic) (UEW: 134). Cfr. Greek κηλίς 'stain'. The sense development in the U cognates might have been 'stain' ⇒ 'bloodstain'.

- **U *muna** 'testicle, egg' (Finnish, Estonian, Sami, Mordvine, Cheremiss, Ostyak, Vogulic, Hungarian, Samoyedic) (UEW: 285). Cfr. Byzantine and Mediaeval Greek μοῦνος m. 'testicle(s); scrotum; male sex organ' (as well as μουνή f. 'female sex organ', whence the word *mona* of same meaning in the dialect spoken in Venice, Italy, that was heavily influenced by Greek)[17]. Neo-Hellenic preserved the feminine form only, but the meaning of 'testicle, scrotum' survived in a number of compound words, cfr. μουνου-χάρι 'castrated, evirated' (χαρέω = 'to make a cut, cut, incise'); μουνού-χος 'eunuch, evirated (man or animal)', μουνου-χίζω 'to evirate, castrate' (σχίζω = 'to cut, cut asunder'); as well as μουνό-ψειρα 'crab-louse' (ψεῖρα = 'flea')[18]. The U word was borrowed from Late Hellenistic or rather Byzantine Greek.

- **U *kuč3** 'ant'. Cognate words can be found in Ostyak, Vogulic and Samoyedic (UEW: 192). Cfr. Mediaeval Greek and Neo-Hellenic κουτσί 'anything very small; small grain; crumb; morsel, bit'. The Neo-Hellenic word is often used as a hypocoristic prefix in compound words. It is very interesting to note that the reconstructed U form has an affricate *č* exactly where Greek

---

[17] According to Filinda, the origin of the Greek word can be traced in archaic Greek **mni–* 'moss; downy hair' > classical Greek μνοῶς m. of same meaning > Mediaeval Greek μνίον 'moss; down-feather; dawny hair; fur of young animals; pile; pubic hair' through the hypocoristic form μνοδίον. (Ἀνδριώτης, 1983, 214 s.v. μουνί).

[18] It is to be noted that Slavic has also a similar word, namely **mǫdo* (< **mǫ–* [=*mon*] + *–d* abstract suffix) in the meaning 'egg, testicle' (cfr. Russian *мудó*, Ukrainian *мудо*, Old Slavonic **мѫдо**, Bulgarian *мъдó*, Croatian *múdo*, Serbian *мýдо*, Slovenian *módo*, etc.).

has its affricate τσ[19]. Beyond confirming the concordance, the phoneme can be of help in dating P-U to Byzantine times.

- **U *lampe** 'marsh, bog, swamp; puddle, plash; mud, silt, warp' (Finnish, Estonian, Sami, Samoyedic) (UEW: 235). Cfr. Greek λάμπη ~ λάπη 'surf, foam, scum, spume, froth; mildew, must (on water or wine); slime, mucilage, mud, silt, warp'[20].

In connection with the Hungarian word láp 'fen, bog, moor', the UEW writes: "Das ung. Wort wurde wegen der Bedeutung 'Moor, Sumpfwiese' von mehreren Forschern zur Wortfamilie *lampe 'Pfutze; Teich; Moor' U gestellt. Diese Zusammenstellung ist aus lautlichen Grunden unwahrscheinlich (die Annahme von *mpp ist unbegrundet) und auch semantisch nicht einwandfrei." Beyond the fact that a secondary nasalisation can never be excluded, in this very case Greek makes no difference between the forms *lámpe* and *lápe*. If the U word is a Greek borrowing, Rédei is probably wrong in rebutting the concordance.

- **U *kura** 'hoar-frost, thin snow' (UEW: 215). The word is an I-E borrowing, cfr. Scythian **xrohu-kasi* 'Caucasus', *i.e.* 'ice-shining', Old High German *(h)rosa, (h)rose* 'ice, crust', etc. The borrowing possibly originates from an Iranic form or from Greek κρύος 'frost, ice', with the insertion of a euphonic (metathetic) vowel to avoid the *kr–* consonantal cluster at the beginning of the word.

- **U *aśe-** 'to put, place, lay, set; to pitch, put up a tent' (Finnish, Sami, Estonian, Mordvin, Samoyedic) (UEW: 18). Compare the U form against Greek ἄεσα 'to overnight', ἄσω 'to sleep', ἄστυ ~ ἄστεα 'town, city', ἄεσα 'oven, altar, heart of the house; house, family', etc.[21]

- **U *puj3** 'back, rear, posteriors' (UEW: 401). Cfr. Greek πυγή 'podex, rear, back, posteriors'. When the loan took place, the pronunciation of the ypsilon was not yet been affected by the typically Greek phenomenon of itacism, thus the vowel sounded /u/ as it is shown by words like Hungarian *turó* 'cheese-curd' < Greek τυρός 'cheese, cheese-curd'.

---

[19] Since in the phonemic stock of Greek there is neither /tʃ/ nor /ʃ/ nor /ts/, the [τσ] digraph, an affricate phoneme the sound of which is a good approximation of both *č*, is used to represent the affricate sound *č* in foreign names. In Byzantine times and in the Middle Ages the digraph τζ [tz] was also used for the same purpose.

[20] This word occurs also in Georgian *lopo* 'Saft (z.B. in Bäumen); schleimige Flüßigkeit auf abgeschälten Baumstämmen' as well as in Basque and Spanish *lapa* 'marshy, muddy soil'.

[21] See the IEW under the following entries: *au–, aw–es–, au–s– `to overnight, to sleep' (p. 72); \*wes– `to stay, rest, reside, overnight; verweilen, wohnen, übernachten', *w–es–ti–s `rest, stay; Aufenthalt' (p. 1170). The I-E word is possibly a loan from Akkadic *ašābu* `to dwell'.

Iranic loan-words in the proto-language

We have seen how some authors consider the Samoyedic branch as the first to have left the alleged Uralic unity. As a proof thereof they mention the fact that it does not include loanwords of Iranic origin. It would be strange enough, however, if it included Greek and Latin words, i.e. relatively "late" languages, but it lacked "ancient" Iranic borrowings. In reality, this claim is false.

- **U *natta** 'snot, slime, catarrh' (Finnish, Sami, Estonian, Samoyedic) (UEW: 299). According to Rédei the word is "onomatopoeic" (?!). Cfr. the I-E base **snā-, snə-(t), snā-sn-eu-, sn-et-* 'fließen, Feuchtigkeit; to flow, damp(ness), moisture' (IEW: 97): Sanskrit *snāuti*, participle *snuta* 'drip, run (eyes, nose); to release any liquid from the body, in particular mother milk'; Avestan *snayeitē* (past participle: *snāta–* 'to wash', etc. Cognate words can be found in many I-E languages, *e.g.* Umbrian *snata, snatu* acc. Pl. 'ūmecta'; Greek voτίς f. 'humidity, moisture'; Norwegian *snott*, Anglo-Saxon *gesnott* n. 'slime, snot'; Old Icelandic *snŷta*, OHG *snūzen*, Norwegian *snūt*, NHG *Schnauze* 'snout, nose', etc.

- **U *jäje** 'strap, band, belt' (UEW: 90) occurring in Votyak (*je*), Zyrien (*ji*) and Samoyedic (*nī, ńi*), etc. The disyllabic reconstruction supplied by Rédei is very likely incorrect, since almost all U cognates are monosyllabic. Compare the U base against Avestan *yāh* 'strap, belt'.

- **U *majδ'a** 'forest, wood' (UEW: 263), occurring in Sami and Samoyedic. Here Rédei and others assumed the presence of the phoneme δ', the existence of which is strongly debated. In this very case, though, the original phoneme was very likely a *δ* , *i.e.* dh. The word can be usefully compared to Crimean-Tartaric *maidan* 'clearing in a forest; forest meadow; forest area' < Persian-Arabic ميدا ن [maidān] of same meaning, originating from Arabic ن ميدا [maidān] 'field, area'[22]. If the concordance proposed by Rédei is correct, this loanword provides further evidence about the real age of the so-called Uralic proto-language. As a matter of fact, the sense of the Arabic word has been extended in Persian to include the notion of "wood, forest". It is common knowledge that Arabic loanwords begun to enter Persian at the time of the Omayyad dynasty (661–750 C.E.), in particular during the reign of Abd el-Malik (685–705). As a consequence of this, the U word might go back to the end of the 7th or beginning of the 8th century of our era only.

- In connection with this entry, we would like to remind readers that the phonemes /δ/ and /θ/ in syllables others than the first one usually yield /z/ in Hungarian:

---

[22] Cfr. the Polish place-name Majdanek (Lokotsch, 1927, #1354; Brückner, 1927, 318 s.v. *majdan*; Staszewski, 1968, 264).

- Akkadian *qātu*, Aramaic and Syriac קתא *qatâ*, Hebrew קת *qath* 'hand' > Hung. *kéz* of s.m. (> **F-U *käte*** 'hand', see UEW, 140).

- Any I-E language (Sanskrit, Iranic, Greek, Germanic, Slavic): **metu–, *medu–, *meδu–* 'mead, hydromel, honey' > Hung. *méz* 'honey' (> **F-U *mete*** 'honey, mead' – UEW, 273);

- Arabic لذع [laδˁ] '(sg) burning, afire, flaming' > Hungarian *láz* 'feever', etc.

This is why the Hungarian word *mező* 'field, area' is very likely a borrowing (either direct or through Persian) from the Arabic word ميدان [maidān] 'field, area'. Therefore, the Hungarian word has nothing to do with the meaning 'to put on clothes'– as the UEW (869) and its forerunners wrongly maintain – but it is rather the offspring of the "Uralic" word **majδ'a* through the following passages:

1. Simplification of the original diphthong *ai*, exactly as in the case of Hebrew מעשה [ma'ăśah] 'story, tale, fable, legend' (> Yiddish [maise]) > Ugric **meće* (UEW: 867) : Hung. *mese* 'story, tale, fable, legend', Ostyak *maś, mańť* 'story, fable'.

2. Levelling of vowel harmony in the second syllable.

3. Loss of the final consonant –*n* as a consequence of the well-known propension of Hungarian for open syllables.

Semitic loan words

As we have seen before, there are Semitic loan words in the Uralic languages. Were they present in the linguistic phase which is called Proto-Uralic? One of the most handy ones seems to be the U word **ńor3* 'swamp, moor' (UEW: 324), Finnish (*noro* 'brook, little stream') and Estonian (*nõru, nõrg* 'water discharge; little, slow-flowing brook; water draining device') preserved the original meaning of the Semitic word. Cfr. Hebrew נהר *nāhār* 'river, (artificial) canal', Aramaic and Syriac נהרא *naharā* of same meaning; Biblical Aramaic נהרא *nĕharā,* Akkadic *nāru,* Arabic نهر *nahr,* and so on. It is to be noted that we can find a phonologically similar base in Mongolic, cfr. *nuur* 'lake'.

On the other hand, the Hungarian word *nyár* 'summer' – that the UEW puts under the same entry – originates from a Semitic homonym, cfr. the verb נהר *NHR* 'to shine, beam', whence Arabic نهر *nahr* 'daylight' and نهار *nahār* 'day, daytime' – but also نار *nār* 'fire', نار *nāra* 'to burn, to be hot', نور *nūr* 'light', نور *nūra* 'to shine, to give light, to burn'; Aramaic and Syriac נהר *nĕhar* 'to shine, beam', נהרא *nehōrā* 'light' (Syriac: נוהרא *nūhrā* 'light'), נור *nūr* and נורא *nūrā* 'fire'; Hebrew נהר *nāhār* 'to shine, beam', נהרה *nĕhārā* 'light, daylight', נור *nūr* 'light, fire'[23], etc.

---

[23] As far as meaning is concerned, a typical feature of *lingua franca-s* and pidgins is that each word is endowed with a number of "meanings". When the language crystallises into a creole, though, one

Conclusion of Section One

The "relative chronology" established by some linguists is very likely wrong. The attempt of matching a false chronology with the absolute chronology of archaeological cultures that developed on assumed homeland-areas had the effect only of leading researchers astray. However, under item 6, Hajdú admits that, up to now, no archaeological culture could be attributed with a minimum degree of certainty to the Finno-Ugrians, let alone the "Uralians".

Anyhow, even Hajdú realized that the arguments he brought forward in order to establish the age of the language did not supply sufficient ground. This is why in a later work he only claimed that: "*Soll die Geschichte der frühesten – nicht über Schriftzeugnisse verfügenden – Periode eines Volkes mit einer bestimmten Sprache erforscht werden, dann ist es eine der sichersten Arten, um die in den Nebel der Geschichte hineinführenden Geschehnisse kennenzulernen, daß man die Sprache des Volkes einem Verhör unterzieht. Die Sprache gestaltet sich als Ergebnis einer jahrtausendelangen Entwicklung heraus.*" (Hajdú, 1987: 274). This view is also wrong. It does not always take "thousands of years" for a language to develop. A *lingua franca* or a *pidgin* may get "independent language" status within *a few years* (sic!) from the time when it first comes to light. It might be that the glottogenesis of the U languages is not "lost in the mists of time", as Hajdú maintains, but took place in an epoch close enough to the present.

In almost all textbooks dealing with the U language we can find an approximate age of six thousand years. One should always wonder from where some figures come from, and in which way they were obtained. By which criteria was the author able to carry out such a calculation? Those who cite an approximate age usually read about it in the works of some other author; then they repeat and amplify the figure, as usually happens with urban legends. The "great age" of Uralic is also a legend, a scholarly myth handed down from one generation of scholars to the next.

Even riskier conclusions beguile the linguistic palaeontologist who attempts inferences about social organization, religion and race of such prehistoric peoples from cognate sources. In connection with this method, J. Anderson wrote: "*A reconstruction of the social organization of the Roman people, for example, based upon cognates in the Romance languages such as those for* bishop, beer, war *and* horse*, would suggest that the Romans were beer-drinking Christians who fought on horseback – all of which is false*" (Anderson, 1973: 27).

The questions of protolanguage and homeland are far removed from the field of the phonological context of language reconstruction and they may often prove false. Thus, it seems that we have forgotten the wise words with which F. de

---

meaning or only a few meanings are preserved. This is the reason why in the present study I tried to present only concordances the meaning of which match in full the original meaning in the lexifier language. In this case, though, the concordance of the Semitic homophones versus Hungarian *nyár* : *nyár* was too important not to be given attention.

Saussure concluded his Course de Linguistique Générale: "*La linguistique a pour unique et véritable objet la langue envisagée en elle-même et pour elle-même*".

Section II: The Finno-Ugric Wordstock

> It is not enough to be Hungarian; you must have talent too
> — Alexander Korda

From the mid-XVIII century, scholars dealing with the Uralic languages were aware of the presence of Altaic and Indo-European *Sprachgut* in the common vocabulary of P-U, to such an extent that many researchers (e.g. Austerlitz, Menges, etc.) maintained that there might be a genetic relationship between the Altaic and the Uralic language families, while Collinder and others put forward the hypothesis of a genetic relationship between Proto-Uralic and Proto-Indo-European. As a matter of fact, a number of words belonging to the U worstock correspond to Indo-European and Altaic bases. The main problem, though, arises from the fact that the matches are asystematic, thus they do not allow us to infer a genetic relationship, but rather a loan relationship. It is moreover possible to show that the concerned loanwords were not borrowed from the most ancient layer of Proto-Indo-European or Proto-Altaic (if they ever existed), but much rather from later daughter languages of these families.

We have now seen that the so-called "Uralic" wordstock includes also words of *Germanic, Iranic* (Middle Persian), *Greek, Latin* and *Semitic* (Arabic, Hebrew, Neo-Akkadic, Aramaic) origin. The Samoyedic languages are very independent and have replaced much of their U lexemes with words of other origins. The Uralic borrowings that we shall present hereunder show that much the same constraints were functionally operating in all the concerned languages synchronically, that is *in the same historical period*. What we mean is namely that there were no intermediate nodes or sub-proto-languages, and that all the offsprings of P-U came about more or less at the same time. To show this, though, we need some more data from the so-called Finno-Ugric layer.

Transpositional rules and phonological constraints

In the following section, we shall use the reconstructed forms of the *Uralisches Etymologisches Wörterbuch* as a metalanguage that shall enable us to check the accuracy of the proposed concordances. We have to stress, though, that some of the phonemes as reconstructed by the UEW are very likely incorrect. This may be considered a direct consequence of the shortcomings of the reconstruction theory. Nevertheless, comparison with the reconstructed "Uralic" metalanguage enables us

to recognize in a straightforward way some of the constraints which led to the birth of the so-called U languages, without the need to cite data for every single daughter language concerned.

In reality, however, a unitarian P-U – as it has been reconstructed from the offspring languages – *never did exist*. Present-day languages originate from a continuum of lects that ranged from the acrolects (that is the forms nearest to the original forms of the lexifying languages) to the basilects (that is those languages the forms of which were the farthest from the lexifiers). Each "daughter language" underwent a complex process of sociolinguistic changes, involving both approximation and expansion of the linguistic resources, thus developing rules and constraints on its own. The borrowings presented in these pages have been singled out for two reasons. First of all, they do match almost exactly the meaning of their counterparts in the lexifier languages. Secondly, the phonological changes between the lexifiers and the reconstructed "metalanguage" can be predicted on the base of a few simple and straightforward rules.

Consonantism & vocalism

From the loan words presented herewith it is possible to realize that some noticeable linguistic phenomena took place in the course of the borrowing. The most common ones are – of course – the same constraints that are valid also to date for almost all Uralic (and Altaic!) languages.

1. **Consonantic clusters** in the initial syllable are not tolerated. This means that consonantic clusters like *sn-* or *sk-* are simplified into *n-* or *k-*.

2. **Diphthongs** are not accepted, and they are usually simplified into one vowel only.

3. **Vowel harmony.** Since the accent falls always in the first syllable, the type of vowel of the first syllable causes an analogical levelling of the vowel in the following syllable(s). If the first syllable contains a front vowel, the following syllable will change its vowel into a front vowel. If the first syllable contains a back vowel though, the vowel of the following syllable(s) will be a back vowel.

4. **Levelling of word length.** The U languages show a very strong tendency to borrow mainly bisyllabic words. When words are composed by more than two syllables, middle vowels or even extra final syllables may easily disappear.

5. **Open syllables**. Only open syllables are allowed in the finale. If the captioned word in the lexifying language has a closed syllable, the extra consonant is very often dropped.

6. **Palatalisation**. A strong palatalisation affects some consonants, in particular when they occur in an intervocalic position: *s* > **ś (š), l >*ľ, n > *ń (ň)*. At the same time, though, geminates in the lexifier tend to remain unpalatalised.

7. **Consonantal simplification:** *č, *ć. Some phonemes in the lexifiers underwent a peculiar treatment consisting in a phonological simplification, thus yielding consistently *č– or *ć– in the reconstructed forms. The involved phonemes are the following ones:

   A. The Greek digraph *ts* (τσ) that, according to many phoneticians, has to be considered a single phoneme[24]. Greek lacks the postalveolar velar sibilant *š* (ʃ) and its counterpart *tš* (tʃ), so that this sound is used since Byzantine times to render the foreign affricate *č* [25].

   B. The Arabic emphatic consonant ص *ṣād* – This is usually pronounced with the blade of the tongue against the teeth ridge, the tip being behind the lower teeth. It is called "emphatic" in relation to the corresponding consonant *s,* yet *ṣād* is articulated in the region farther back in the mouth called the soft palate and with greater tension of the articulatory organs[26]. In Arabic, the emphatics give to the vowel *a,* when it precedes or it follows them, special "dark" qualities, very similar to the Hungarian short labial *a* (ɒ) vowel. Throughout the Middle Ages it was used to render the foreign affricate *č*.

   C. The Hebrew emphatic consonant צ *ṣādhḗ* – The precise nature of this consonant in Ancient Hebrew (and all the dead Semitic language for that matter) is an unresolved question[27]. Its value, that has no equivalent in our languages, is precisely that of the corresponding emphatic sound *ṣād* in Arabic. In Modern Hebrew it is pronounced as an alveolar affricate *ts*. Throughout the Middle Ages and up to the present it is used as representative of foreign *č ,* e.g. כרץ = *Kerč* (place-name) .

   D. The alveolar affricate *c.* These constraints were still productive in Hungarian during the XVII–XVIII century: cfr. the German place name *Leipzig* and its Hungarian counterpart *Lipcse,* where the diphthong *ei* is simplified to *i,* the affricate *c* (=z) turns into *č* (=cs) and the final *-g* is dropped to leave the ending syllable open.

   E. The consonantal cluster *ps* (Greek ψ).

   F. The sibilants alveolar *z* and postalveolar *ž.*

   G. According to the testimony of the reconstructed forms, we should in-

---

[24] See Mirambel, A.: "Le groupe *ts* en grec moderne", in: Bull. Soc. Ling. de Paris, 1942–45 (42), 145–164; Householder, F. W.: "Three dreams of Modern Greek Phonology", in: Papers… Pappageotes, 1964, 17–27; Romeo, L.: "Toward a phonological grammar of Modern Spoken Greek", in: Papers… Pappageotes, 1964, 60–77; Mioni, A.: Fonematica constrastiva. Bologna, 1973, 395.

[25] For example, Greece is called in Italian "Grecia" [pr.: *greča*]. A native Neo-Hellenic speaker renders it as *gretsa*..

[26] See Laufer, A.: "Descriptions of the emphatic sounds in Hebrew and in Arabic" (in Hebrew), in: M. Bar-Asher (ed.): *Language Studies* II–III (Jerusalem, 1987), 423–38.

[27] For a well-balanced exposé of the question, see J. Cantineau in *Semitica* 4 (1951–2), 91–93.

clude here also the postalveolar velar sibilant *š*. Yet, this inclusion might be due to the reconstruction techniques, and a refinement of the method applied could possibly yield different results. In fact, Rédei does not reconstruct consequently all *š* as **č* or **ć*.

1. **Devoicing *f* > *p**. The unvoiced labial spirant *f* (*ph*), that did not occur in ancient Turkic, is replaced by *p*

2. **Hiatus filler**. The voiced labio-dental *v* is replaced by **h*

3. **χ > *k**. The unvoiced velar spirant *x* (χ) in the first syllable is usually reconstructed with **k*, a peculiarity that is possibly due to a Slavic influence. The constraint often does not apply to monosyllabic words.

4. **Spirantization: *g* > (*γ >) *j**. The velar occlusive *g* is usually replaced by the continuous spirant **γ* > **j*.

5. **F-U **hk*–** is the representative of the emphatic *q* of Semitic

6. **Geminates** in the original language undergo a peculiar treatment.

7. ***d* > *δ' and θ > *t**. The dental occlusive *d* is often reconstructed as **δ'*, while the interdental consonant θ is often reconstructed as **t*.

Many of these constraints do not apply to Hungarian that, in this respect, represents an *acrolectic* variety.

As concerning the vocalism, the situation is a bit more complex. We have seen the peculiar phenomenon of the *a* turning to *ä*, cfr. Avestan *yāh* 'strap, belt' > U **jä*– 'strap, band, belt'; Akkadian *qātu* :: Aramaic and Syriac קתא *qatâ* :: Hebrew קת *qath* 'hand' > F-U **käte* 'hand'; Hebrew מעשה *ma'ăsah* 'story, tale, fable, legend' > Ugric **meće* 'story, tale, fable, legend'. In connection with the Iranic borrowings, É. Korenchy described this phenomenon as follows:

> "*...Was nun die Lehnwörter anbelangt, so wurde bei der Untersuchung der finnisch-ugrisch-iranischen Lautentsprechungen ein Merkmal festgestellt [...] An der Stelle des *a in den erschlossenen iranischen Wörtern steht auf finnisch-ugrischer Seite in vielen Beispielen immer wieder *ä. Es ist auch möglich, daß in einer gewissen Periode der iranischen Sprachgeschichte neben den a auch das bis dahin fehlende ä erschienen ist. Mit anderen Worten, man kann annehmen, daß in der Sprache eines untergegangenen iranischen Volkes das Phonem ä vorhanden war, obwohl es in den meisten belegten iranischen Sprachen keine Beweise dafür gibt*" (Korenchy, 1988: 669).

The vowel shift is not limited to the /a/, but it effects also the vowel /e/ that in some occurrences turns into /i/. The phenomenon was observed also by Harmatta, who wrote: "In the case of the vowels, we find in Finno-Ugric a double correspondence: *a* and *o* are the representatives of Proto-Iranic *a*, while *e* and *ä* are the representatives of *e*" (Harmatta, 1977: 173).

This phenomenon seems to be due to an *areal* aspect of vowel pronunciation.

Let us examine the scale of Hebrew vowel timbres as registered by the Massoretes [נקדנים *Naqdanim* "punctuators, vocalizers"] of Tiberias (Palestine):

i	e	ε	a	ɔ	o	u
					•	
.	..	˸	–	˸		˙.
ḥireq	*ṣeré*	*seghol*	*pataḥ*	*qameṣ*	*ḥolem*	*qibbuṣ*

Yet, Lambert in his *Traité de la grammaire hébraïque* makes an interesting observation of the fact that the Babylonian Massoretic vowel system of Hebrew has a single sign *pataḥ* (called *pitḥa'* in the Babylonian massorah) representing both the Tiberian *pataḥ* (that is the vowel /a/) and *seghol* (that is the phoneme /ε/)[28]. This means that in the seventh century the pronunciation of Babylonian Jews did not make any distinction between /a/ and /ε/ (or /ä/, if you so desire) and – as a consequence of a push chain – between closed /e/ and /i/.

Arabic has only three vowels notated in its vocalization: *a, i* and *u*. A simplified diagram could be the following:

Vowel timbres	i	e	ε	a	ɔ	o	u
Arabic		i		a		u	

Much the same phonological phenomenon can been observed in Middle Persian, where there is a peculiar shift of /a/ and /e/ to a more frontal pronunciation:

Vowel timbres	i	e	ε	a	ɔ	o	u
Iranic		i	e		a		o	u

This might explain the reasons for the vocalic shift in the U languages. It is interesting to note that the same self-constraint can be found in the late Hungarian loans, where the vocalic shift remained productive a long time:

- Turkic *baqa* > Hungarian *béka* 'frog';
- Ťuvash *ʒarta* (Vogulic *šarta, tšarta*) > Hungarian *gyertya* 'candle';
- Greek κάστανον (>Latin *castanea*) > Hungarian *gesztenye* 'chestnut';
- Greek κάμινος (>Latin *caminus*) > Hungarian *kémény* 'chimney';
- Latin *Castellum* > Hungarian *Keszt-hely* (place-name)

---

[28] Joüon, 1993: 34.

- Bulgarian Turkic *kap* (> Old Slavonic **капъ**) > Hungarian *kép* 'shape, image, picture', and so on.

From the point of view of **semantics,** it is quite interesting to note that – generally speaking – not every U language preserved the original meaning of the loaned words. The meaning supplied by the UEW often represents just one of the possible choices. Yet, by checking the various meanings of the U languages *there is always at least one language that preserved the original meaning*, while the other U cognates might show a lesser or major semantic shift.

Examining the "primitive" wordstock

Hereunder we shall take a closer look at some of the neater correspondences of the F-U wordstock.

- **F-U *al3** 'to cast a spell' (UEW: 7). Here the UEW supplies a meaning that does not match in full the various meanings of the U cognates: Mordvin: 'to promise, take an oath, curse'; Cheremiss: 'to pray, worship, adore'; Ostyak: 'to curse, scold, chide; cast a spell'; Hungarian: 'to bless; to sacrifice; to curse, swear, cast a spell'. The origin of this word can be found both in Hebrew and Arabic. Cfr. the Hebrew verb אלה *ālāh* 'to curse; swear; take an oath; to worship, adore, deify' as well as Arabic الى آ *aʔla* 'to swear, take an oath'. Since the U cognates preserved many of the meanings of the Hebrew word, the borrowing took very likely place from Hebrew rather than Arabic.

- **F-U *äh3ć3** 'raspberry'. Cfr. Post-Biblical Hebrew אבש *ăveš* 'dry grape' (cognate of יבש *jěvēš,* the latter occurring in the Peshitta).

- **F-U *aŋke** 'narrow, pressed; difficulty; to become narrow' (UEW: 12). In this case the UEW mentions that the Balto-Finnic cognate could be a borrowing from Germanic (pply. Gothic) *aggwus* 'narrow'. The extensor of the UEW did not realize that the U base itself is a loan from I-E *$an\hat{g}h$– 'eng, einengen, schnüren' (IEW: 42), cfr. Greek ἄγχω, Latin *ango,* Old Norse *aŋge* 'annoyance, grievance, loss', Old Swedish *eŋgi,* etc.

- **F-U *ar3** 'to tear, rip off, tear up'. The UEW (17) reports here just a part of the meaning of the concerned F-U base, cfr. Votyak 'to tear off, to thin out, to weed'; Zyryen 'to tear off, break off, tear up; to become feeble, weak; to wear off; to dwindle, fade; to suffer damage, to be wounded'; Ostyak 'to grow thin; to become weak'; Vogulic 'to grow thin'; Ostyak 'to tear up; to wear out (clothing), to become thin through wearing'; Hungarian 'to go bad, perish (cooked food, bread, water, etc.); to damage, harm, be detrimental'. Compare Greek ἄραιω 'to wear out, wear thin; to make rarer, thin out; to make weak, feeble; to mellow; to make soft, friable, crumbly' and its adjec-

tive ἀραιός 'worn out, thin, soft, mellow, friable; moldy, rotten, decayed, crumbled'.

- **F-U *äne** 'voice, sound' (UEW: 25). Cfr. Hebrew ענה *ʿanaḥ* 'to answer, reply, respond', ענה *ʿanah* 'to sing'; Aramaic ענא *ʿănaʾ* and Syriac ענא *ʿĕnaʾ* 'to answer, to hear'; Syriac עני *ʿani* 'to sing'.[29]

- **F-U *äh3** 'fire; to burn' (UEW: 26). Cfr. Hebrew אח *aḥ* 'fire port, brazier, fireplace' (this word is on its turn a loan-word from ancient Egyptian *ah* of s.m.)

- **F-U *ćaδa ~ *śaδa** 'run, rush, dash; to run; rut of female animals, to rut' (UEW: 28). Cfr. Hebrew סלד *salad* 'to jump, spring' and Arabic صلد *ṣalada* 'to beat the ground in running'. The borrowing in the Finnic language group took probably place from Arabic. In fact, the emphatic *ṣ* of Arabic affects the pronunciation of the a vowel next to it, with the result of inducing what is thought to be a "sporadic change" *a > *ō > uo in Finnish *suota*. Moreover, an ectlypsis (loss) of the medial syllable took place in the Finnic cognates, while other cognate languages preserved the medial *l*. This is another case when Hungarian – with the word *szalad* 'to run' – preserved the original form of its lexifier.

- **F-U *ćara** 'hinge, pivot of the door; to hinge' (UEW: 30). Cfr. Hebrew ציר *ṣīr* 'hinge, pivot of the door', ציר 'to turn, revolve, hinge', Aramaic צירא *ṣīrâ* 'hinge, pivot of the door'. Hungarian (*csír*) and Zyryen (*źir*) are the F-U languages that better preserved the vowel quality of the lexifier.

  By comparing the Hungarian words *csír* 'hinge' and *csira* 'seedling, sprout' (Ugric *čX̌r3 'to germinate, sprout; keimen, sprossen', UEW 840) we can see the constraint *ṣ/š/z* > *ć at work in the alleged "Ugric layer" too: in fact, the Ugric form is a borrowing from Hebrew זרע *zeraʿ* 'seed; seedling; sowing; sowing season; offspring; semen', cfr. Aramaic and Syriac זרע *zĕra,* Arabic *zara- a,* Akkadic *zĕru* 'seed', and so on.

- **F–P, ?U *ćar3** 'hard, strong, firm' (UEW: 30). The U word goes back to Hebrew שרר *šārir* 'to be strong, firm', Syriac שר *šar* 'to be strong, firm', Aramaic שרר *šĕrir* 'to be strong, firm'. Cfr. also Sanskrit सार *sár-* 'pit or sap of a tree; vigour, firmness, strength'. The choice of a word that occurs contemporaneously in several lexifiers seems to be a most typical behaviour of *lingua franca*-s and pidgins.

- **F-U *ćapp3–** 'to hit with a banging or cracking sound; to slap, spank' (UEW: 29). Cfr. Hebrew ספק *sāfaq* 'to clap (the hands in sorrow)'[30] as well

---

[29] The Semitic verbs are possibly related to Ancient Egyptian *anni* 'to sing'.

[30] There is also a secondary form שפק *sāfaq* 'to clap, strike the hands'.

as Arabic صفق *ṣafaqa* 'to slap the face, to strike the hands'[31]. This word might also be an onomatopoeia .

- **F-U *ćähk3– ~ ćähk3–** 'to break' ('break off, smash') (UEW: 31). The reconstruction supplied by the UEW does not take into account the final *-p* of Ostyak and Vogulic, representative of an original *f* . Cfr. Aramaic שקף *śĕqaf* and Hebrew שקף *śāqaf* 'to beat, strike, knock'.

- **F-U *ćer3** 'grey' (UEW: 36). The UEW compares this base against Turkic Tob.-Tatar. *sur* 'grey-blue' and Kas.–Tat. *soro* 'grey' – the vowel quality of which is different. Cfr. instead Akkadic *šeru* 'dawn, daybreak, the darkness preceding dawn' (a meaning occurring also in Votyak and Vogulic) as well as *šūru* 'coal' (cfr. also Jewish Aramaic שחר *šĕḥar* 'to be black'; Syriac שחר *šĕḥar* 'to be or become black'; Hebrew שחר *šâḥar* 'to be black'). Cfr. also Russian *серый* 'grey, light grey', Bulgarian *сер,* etc. from a more ancient Slavic **sĕrъ* (Vasmer, III:610), as well as Latin *sera* 'evening' ⇐ 'to be late'. The choice of a word that occurs contemporaneously in several lexifiers seems to be a most typical behaviour of *lingua franca*-s and pidgins.

- **F-U *ć8pp3** 'drop; to drip, drop' (UEW: 49). Cfr. Hebrew שפך *šāfakh* 'to pour, pour out, spill; to pour from one vessel into another; to empty out', Jewish Aramaic and Syriac שפך *šĕfakh*, Mandaic and Christian Palestinian Aramaic שפך *šāfakh* of same meaning.

- **F-U, Ug *čaw3** 'sour, to turn sour' (UEW: 54). Cfr. Aramaic שב *šav* 'alum', Hebrew שב *šav* 'alum, vitriol', Arabic شا ب *šab* 'alum'. Alum (that is salt of potassium sulphate, aluminium sulphate or potassium-aluminium sulphate) has a tart, acrid taste and was first used by the Phoenicians to tan hides and skins as early as the 6th century before our era. It can be occasionally used to replace rennet.

- **F-U *čonče** 'string, cord, rope' (UEW: 61). Cfr. Hebrew שנץ *šānaṣ* 'to fasten with cords, lacing, tying', שנץ *šeneṣ* 'cord, rope'.

- **F-U *čukk3** 'thick, dense'. The UEW (62) compares this base against Turkic *čoq* 'much, many', Mongolian *čogča* 'a lot of, loads of, masses' and Tunguz *čokčo* 'hill, hillock'. Nevertheless, cfr. Akkadic *šūqu* 'abundance' and Hebrew שוק *šuq* 'to be abundant', where the Semitic word reflects better the vowel quality of the F-U base. The Altaic cognates are very likely Semitic loan-words.

- **F–P, ?F-U *δ'8kk3** 'to prick, sting, push, butt' (UEW 66)**.** Cfr. Hebrew דקר *dāqar* 'to pierce, stab, thrust', Jewish Aramaic דקר *dĕqar* of s.m.

- **F-U *iče ~ iśe** 'shadow, shade-soul' (UEW: 79). The UEW mentions as primary the unexpected meaning of 'shadow; shade soul', cfr. Ostyak 'shade

---

[31] There is also a secondary form سفق *safaqa* 'to slap the face, to struck the hands'.

soul; spirit; sg incorporeal in men; puppet made after the death of an adult'; Vogulic: 'form, shape, picture; the real shadow of men which always accompanies them and which becomes visible in the sunshine; the transparent, metamorphic »spirit« of a live or dead man', etc. whence 'spirit' > 'evil spirit'. The popular belief of the existence of a double-ganger can be found among many peoples. According to the Egyptians, men are composed by three elements: the body, the soul and the *ka*, i.e. something intermediate between soul and body, an incorporeal, spiritual double which was to be seed in men's shadow as well as in their mirrored image, and which takes care of many spiritual functions. The Greek *daimon*[32] and the Roman *genius* were shade-entities or metaphysical double-gangers which accompanied people and participated in their joys and sorrows. The myth survived in the Christian belief of the Guardian or Tutelary Angel. Something similar can be found in the popular belief of the Semitic peoples (cfr. Hebrew כפיל *kāfīl* 'double, alter ego, second self' and Arabic *kifl* 'double, double-ganger'), and in that of the Teutonic peoples (cfr. German *Doppelgänger* and English *double-ganger*). Nevertheless, the meaning of the Finnish, Mordvin, Lappish, Cheremissian and Votyak words is 'self'. And the latter was in fact the primary meaning of the FU base, which is such a perfect double of Latin *ipse, –a, –um* 'selbst' (where -ps- > U *ć, *č) that it needs no further comments.

- **F-U *jähe** 'ice' (UEW: 93). Cfr. Modern Persian يخ *jaḥ* 'ice' < I-E *jeg–* (IEW: 503). The word is a Middle or Late Middle Persian borrowing.

- **F-U *jer3** 'curse, oath; to curse, scold' (UEW: 97). Cfr. Greek ἱερεύω 'to consecrate to a god, to sacrifice'.

- **F-U *jip3** 'owl, eagle-owl'. This is perhaps a Slavic or Balto-Slavic borrowing, cfr. Lettish *upis* 'eagle-owl' as well as Old Russian въипль, Old Slavonic въипъ 'seagull' (IEW: 1103).

- **F-U, ?U *kača** 'hole, cavity; container, receptacle; (wooden) shelf, case' (UEW: 112). Cfr. Arabic قا شوه *qašwa* 'a basket made with palm leaves' (cognated with Hebrew קשוה *qaśwa* 'a vessel for libations') but see also Latin *capsa* '(wooden) box, holder, container; Behältnis, Kapsel, Kasten' (whence French *caisse,* Italian *cassa,* Spanish *caja,* Portuguese *caixa*, etc.) > Byzantine Greek κάψα > Post-Biblical Hebrew קפסה *qufsa.* The borrowing might have taken place either from Arabic *qašwa,* Greek or Latin *capsa*.

- **F-U *kač3** 'something curved, convex, swelling' (UEW: 144). Cfr. Hebrew קשט *qōšeṭ* 'bow, rainbow, arc'.

- **F-U *kač3** 'to understand, know, can' (UEW: 144). Cfr. Hebrew קשב *qāšav* 'to incline one's ear, to listen, hark, give attention'.

---

[32] Socrates and Plato identified the *daimon* with the inner voice of conscience.

- **F-U *kaja–** 'to throw' (UEW: 116). Cfr. Arabic قاه *qāa ~ qayya* 'to vomit, spit out', Hebrew קיא *qā* 'to vomit, spit out'.

- **F-U *kaľw3** 'cuticle, membrane, pellicle, scale' (UEW: 121). Cfr. Hebrew קלפה *qalipā* 'peel, bark, cortex, husk, shell' (< קלף *qālaf* 'to scrape, peel, pare); Akkadic *kalāpu* 'to peel', *qilpu* 'skin' and *quliftu* 'scale of a fish'; Arabic قلف *qalafa* 'to scrape, peel'. See also Latin *callum* 'hard and thick skin, callum'.

- **F-U *kanta** 'margin, border, edge; bank (of a river)' (UEW: 123). Cfr. Latin *cantus* 'corner' < Greek κανθόν 'corner of the eye; corner'; German *Kante,* Dutch *kant,* Anglo-Saxon *cant* 'edge, border'; Cymric *cant,* Irish *kantz* 'dam, circle, edge, border' (see IEW, 526). The loan might have taken place from any western I-E language.

- **F-U *kar3** 'to be afraid of something, to be frightened, scared' (UEW: 129). Possibly a loan word from Old Norse, cfr. Icelandic *skjarr* 'timid, shy', Middle English *skerren* 'to scare', *skerre* 'timid, shy', etc.

- **F-U *kawa-** 'to come to the surface, rise' (UEW: 131). The correct reconstruction should be **kapa-,* as Finnish *p*- forms like *kapua* seem to show (the *v*- or *w*-forms might be secondary). Cfr. Hebrew קפא *qāfâ* 'to float, to be on top; to curdle, to become condensed, to be congealed', Aramaic קפא *qēfâ* 'to float, to be on top; to curdle, to become condensed, to be congealed', Syriac קפא *qēfâ* 'to skim off'.

- **F-U *kawka-** 'long' (UEW: 132). Cfr. Hebrew קוקו *qawqāw* 'to measure (the length) with a line, to mark (the length) with lines' (a reduplicated root from קו *qaw* 'measuring line, cord, string').

- **F-U *käć3** 'trapping net, hunting snare'. Cfr. Hebrew קש *qāš* 'to lay snares'.

- **F-U *kälä–** 'to wade' (UEW: 133). Cfr. I-E *kel-, kelə* 'to rise, raise' > Lithuanian *keliù, kélti* 'to lift, raise, carry; to transport over water'; *kélna* 'boat, ferry'; *kélta, kéltas, kéltuvas* 'ferry boat'; Lettonian *celtava* 'little ferry boat'; Slavic **cln* 'boat' > Russian чолн 'boat, small ferry', Czech *člun* of same meaning, etc. The word is probably a borrowing from the Balto-Slavic languages.

- **F-U *käl3** 'cord, rope' (UEW: 135). Cp Post-Biblical Hebrew קלע *qelāʿ* 'plait, twisted cord'.

- **F-U *käme(–ne)** 'flat hand, the hollow of the hand, palm of the hand' (UEW: 137). Cfr. Aramaic קמע *qēmaʿ* 'to take a handful, to enclose with the hand' (> קמעא *qěmĕāʿ* 'a little, a handful), cognated of Hebrew קמצ *qāmaṣ* 'to enclose with the hand, grasp, take a handful' and Arabic قمـز *qamaza* 'to take with the fingertips'.

- **F-U *kär3** 'to bind, tie up, thread' (UEW: 139). Cfr. Greek καιρόω 'binde das Gewebe zusammen; to tie or bind the thread together'.

- **F-U *kät3** 'hand' (UEW: 140). Cfr. Aramaic and Syriac קתא *qatā,* Hebrew קת *qath* < Akkadic *qātu* 'hand'.

- **F-U *käč3** 'knife' (UEW: 142). Cfr. Hebrew קצע *qāṣaʿ* 'to cut; to cut off; to scrape, to scrape off', קצב *qāṣav* 'to cut off; to determine; to butcher', קצח *qāṣaḥ* 'to cut off; to destroy', קצה *qāṣah* 'to cut (trees); to decide', קצץ *qāṣaṣ* 'to cut off; to cut trees; to stipulate', קצר *qāṣar* 'to cut; to reap, harvest', *etc.*

- **F-U *keð'3** 'hide, skin, leather, peel' (UEW: 142). Cfr. Hebrew קטיפה *qěṭifā* 'cloak, mantle', Arabic قطفة *qaṭifā* 'garment, coat' (emphatic *ṭ* > **δ'* ?).

- **F-U *kepä** 'light; leicht' (UEW: 146). Cfr. Aramaic קפא *qēfāʾ* 'to float, be on top', Hebrew קפא *qāfāʾ* of s.m.

- **F-U *kerä** 'to ask; bitten' (UEW: 149). Cfr. Aramaic קרא *qēra* 'to call, invoke, summon, invite, proclaim', Syriac קרא *qēra* and Hebrew קרא *qāra* of s.m.

- **F-U *ker3–** 'to belch, to burp' (UEW: 151). Cfr. Hebrew כרסם *kāram* 'to chew, gnaw, devour'. The root is widespread in all Semitic languages, cfr. Akkadic *karašu, karšu* 'stomach, belly'; Aramaic כרא *karā* 'stomach, belly'; Hebrew *kare* 'belly'; *kareš* 'stomach, belly'; Arabic *kariš, kirš* 'stomach, belly'; Ethiopian *karš* 'stomach, belly'.

- **F-U *kič3** 'illness, disease' (UEW: 153). Cfr. Aramaic and Hebrew כשף *kešef* 'witchcraft, sorcery, magic' < Aramaic כשף *kēšaf* 'to practice sorcery' (cognate of Akkadic *kuššupu* of s.m.). A most typical adaptation to the popular belief that illness is the result of a spell cast by a sorcerer.

- **F-U *kiwe** 'stone' (UEW: 163). Cfr. Aramaic כיפא *kēfā* 'cliff, rock, stone', Syriac כאפא *kēfā* 'cliff, rock', Hebrew כף *kēf* 'cliff, rock, stone'. If the borrowing was from Aramaic, the expected F-U form would had been **kipe.* The borrowing took probably place from Hebrew, since the final *f* in Hebrew had a fricative character that could easily explain the presence of **w* in F-U.

- **F-U *koppa** 'something hollow, dent, convex, doomed' (UEW: 181). According to UEW, the Estonian word is possibly a L–W from Swedish *kopp* 'cup', the Ostyak word is assumed to be a L–W from a Turkic language (*i.e.* Čagatay *qapa* 'head'). Moreover, the UEW considers Hungarian *kupa* 'cup, goblet, beaker; pate, back of the head; pit(fall), hole, hollow, dip, dent' a Latin loan, yet words like *kuporodik* 'cower, crouch', *kuporgat* 'scrape, rake together', *kuporol* etc. are not taken into any account. Moreover, the UEW does not explain why the Magyars should have borrowed the meaning of 'pate, back of the skull' from Low Latin while, at the same time, the Ostyaks borrowed a similar meaning of 'skull' from Turkish. Cf. Aramaic קבב 'to be bent, crooked;

to hollow out, to vault' > Hebrew קב *kab* 'a measure of capacity', Arabic قاب *qab* of s.m; > Greek κάβος 'a measure of capacity for cereals'; Arabic قب *qubba* 'to be vaulted', قبة *qubba* 'anything vaulted; cupola, dome' (whence the Latin hypocoristic form *cupo-la* as well as *cubus* '[canopied, baldaquined] bed'). See also I-E **keu–p–* (IEW: 191) > Sanskrit कूप *kúpa* 'pit, pitfall, hole'; Greek κύπελλον 'goblet, beaker', κύπρος 'a measure of capacity for cereals'; Latin *cūpa* 'barrel, cask, vat, tun'

- **F-U *mete** 'honey, a drink made of honey' (UEW: 273). We often saw this word compared against Sanskrit मधु *madhu–* 'a sweet drink, honey, mead[33], Avestan *maδu* 'honey', Greek μέθυ 'wine', but the researchers often tend to forget the Celtic word for 'mead' (Old Irish *mid,* Cymric *medd,* Old Welsh *medu,* Breton *mez*), the Germanic word for 'mead, hydromel' (MHG *mete,* OHG *metu,* German *Met,* Dutch *mede,* Swedish *mjöd*), the Slavic word *medъ* 'mead, hydromel', etc. This means that this word could have been borrowed at any time in history from any one of a number of I-E languages and certainly not, as some researchers maintain, from the most ancient layer of Indo-European.

- **F-U *paľa–** 'ice crust, frost; to freeze' (UEW: 352). As a matter of fact, the origin of this word goes back to Arabic فج *faʤa* 'to be cold', cognate of Aramaic and Syriac פג *pag* 'to be cold'. Greek borrowed the Semitic word as παγετός 'frost, ice, hoar-frost' and παγερός 'frozen, icy, cold, chilly'). The geminate *dʒ* of the Arabic word caused a perturbation in the output, cfr. the Hungarian acrolect *fagy* 'to freeze; frost' : *fáz-ik* 'to chill, to feel cold'. The reconstruction of the F-U **ľ* phoneme is very likely wrong.

- **F-U *panč3 ~ *pač3, *ponč3 ~ *poč3** 'tail' (UEW: 353). Possibly an Iranic but more likely a Slavic loan-word, cfr. Sanskrit *púccha-, púccha-m* 'tail'; Russian and Polish *puch* (< **peukso-, *poukso-*), Sorbian *o-puš, opyš* 'tail' > I-E *pūk-, peuk-* approx. 'dick behaart, buschig (buschiger Schwanz), dichtvollig' (IEW: 849).

- **F-U, ?U *pola** 'berry' (UEW: 392). Cfr. Aramaic פולא *pōlā* 'bean' (cognate of Hebrew פול *fōl* 'bean' and Arabic فول *fūl* 'bean , broad-bean'). A typical example of a meaning that accommodated to the changed socio-climatic conditions.

- **F-U *tälwä** 'winter' (UEW: 516). Cfr. Aramaic תלג *tĕleg* 'snow, ice'; Arabic ثلج *theldʒ* 'snow, ice' (> ثلجت *thulidʒet* 'to snow, freeze, ice'); Syriac תלגא *tēlggā* 'snow, ice'.

- **F-U *wire** 'blood' (UEW: 576). Cfr. Arabic وريد *warīd* 'large blood vessel, vena, arteria', ورد *warda* 'rose', وردي *wardī* 'pink'; Hebrew וריד *vārīd* 'large

---

[33] I.e. a liquor consisting of a mixture of honey and water.

blood vessel, jugular vein', וריד *vered* 'rose', וריד *vērēd* 'to be pink, red'; Syriac ורידא *varīdā* 'large blood vessel, vein, jugular vein'; Akkadic *ur'udu* 'vein, blood vessel'. The Semitic words originate from Persian **wrda* 'rose', whence also Greek Ϝρόδον 'rose' and Latin *rosa* 'rose'. The Hungarian acrolect preserved the connection of this word with the meaning of 'blood', 'pink- or red-coloured' and 'flower': cfr. Hung. *virág* 'flower' originating from a former *vira* as it appears in the Murmellius Lexicon (1533): "Flosculus: Eyn blumlin: *viratzka* ", as well as the Hung. forms *virrad* 'the day is breaking' (pply. 'the blushing of dawn') and *virradat* 'dawn' (pply. 'the blush of dawn').

Finno-Ugric numerals

- **F–P, ?F-U *ikte (*ükte)** 'one' (UEW: 81). The numeral goes back to Hebrew אחד *eḥād* m., אחת *eḥāth* f. 'one'. The trailing *e* developed into *i* (in some languages *i > ü* according to the usual Turkic constraint, e.g. cfr. Mediaeval Greek τιμόνι > Turkish *dümen* 'rudder, pilot wheel'). The original form can be found in the Finnish genitive, cfr. *eχad > *iχ-∅-d > yhd-(en)*. On the other hand, the Hungarian form *egy* 'one' goes back to the Akkadic form *ēd(u)* 'one'.

- **F-U, ?U *kakta ~ *käktä** 'two' (UEW: 118). We already mentioned the origin of the F-U word for 'hand', *i.e.* Akkadian *qātu*, Aramaic and Syriac קתא *qatâ*, Hebrew קת *qath* 'hand'. The numeral "2" originates from the dual of this word, i.e. קתים *qattaim* 'two hands'. The final F-U form was conditioned by the geminated *tt* of the Hebrew dual.

- **F-U *kolme** 'three' (UEW: 174). Perhaps the ultimate origin of this word is Greek χρῶμα 'surface, skin, colour, dye', possibly the popular name of the Byzantine coin the Latin name of which was *tremissis* 'the third part of a shekel, one third of a shekel'. This assumption is made possible by the Hebrew loanword כרום *krōm, kĕrōm* 'yellow and green colour; a coin made of coloured, non-ferrous metal; a fabulous multicoloured bird' that seems to point in the same direction. The replacement of the liquids /r/ > /l/ that took place in some U languages is yet to be explained. As usual, Hungarian preserved the acrolectic form *három* '3'.

Section III: Drawing an inference

Extent of the basic vocabulary

From the second half of the XVIII century it became common knowledge that U and F-U include a noteworthy number of words of Indo-European and Altaic origins and, ever since, they were regarded as borrowings that went back to time immemorial. The prejudice that wanted Uralic to be "at least" as ancient as Indo-European hindered every progress in the field.

- A number of lexical concordances was left out of attention because they were much too recent to fit the theory that wants U to be "very ancient". Nevertheless, there were words such as Finnish *pappi* : Hungarian *pap* 'priest'; Finnish *risti* : Hungarian *kereszt* : Mokša *k'řest* 'cross, crucifix'; Hungarian *fátyol* 'veil' : Erzya and Mokša *paćala* 'kerchief, shawl' (< Byzantine Greek φακιώλιον < Latin *faciale* 'kerchief') and so on. These and many more do indeed belong to the shared wordstock of U/F-U languages.

- Some proposed concordances were arbitrarily rejected. For example, B. Munkácsi (Keleti Szemle 1: 208, 6: 206; ÁKE 408, 650 etc.) and H. Sköld (Die Ossetischen Lehnwörter im Ungarischen. Lund-Leipzig, 1925: 27) showed that the U word **keća* 'a species of fish' is a loan-word from Ossetic *käsag* 'der frische Flußfisch'. Yet, the concordance was rejected because it did not qualificate as "very ancient" and "from the Uralic neighbourhood". The F-U word **miń3* 'heaven' was related by Lindström (Suomi 1852: 58), Munkácsi (ÁKE, 462), Schmidt (Nyr. LV: 99) and Simonyi (Nyr. XLII: 435) to Iranic, namely Avestan *mainyu* 'spirit, heaven' and New Persian *mīnō* 'heaven, paradise'. The UEW considers the concordance not even worth mentioning, probably because it should also state that the match is closer to New Persian rather than Avestan.

- The number of lexical items that pertain to the base language is much larger than those appearing in the UEW. Some words were preserved by one daughter language only. The concordances of other words were not recognized because their meaning accommodated to the environment where the concerned U peoples live, cfr. for example the "Uralic" word **tewä* (UEW: 522) Finnish *teva* 'male elk', Samoyedic *tī, tia,* etc. 'reindeer', Hungarian *teve* 'camel' originating from the Turkic word for 'camel', see Ujghuric *täwä, täbä,* Čuvaš *təvε,* etc.

- Some of the loan-words underwent substantial modifications in their structure and/or meaning, thus it became very difficult to match them with the original words of the lexifier languages. Just as an example compare the Hungarian dialectal word *agyigó* ("Agyigó fassang", from Z. Kodály's *Villő*)

against its Italian origin *a dio, addio* 'good bye, adieu' to realize how complex a match can sometimes be.

- Some words were probably borrowed from unknown dialects and/or lost languages. Cfr. for example the U word **kuŋe* 'moon; month' (that fits well the I-E base *kand-* 'leuchten, glühen; hell' – see UEW, 526 – and cfr. Middle Bretonian *cann* 'full moon', Sanskrit *candrá-* 'moon'; Albanian *hânë, hënë* 'moon'), where a borrowing from the unknown Celtic language spoken by the Galatians or an Iranic dialect might come into account. See also the U word **śilmä* 'eye' – cfr. Albanian *sy (sü)* 'eye', *symath* 'big-eyed', Greek ὄσσε 'eye' (< I-E **okṷ–* 'eye', UEW: 775), where a borrowing from the unknown language spoken by the Albanians of the Caucasus could possibly account for the origin of this word.

In the former sections we gave a selection of a few words only. It is impossible, within the limits of the present paper, to go into details and expand the glossary any further. However, these borrowings enabled us to catch a glimpse of the structure of the base language.

We accounted for a number of borrowings that took place from historically attested languages: Germanic, Early Slavic, Greek, Latin, Late Middle Persian or New Persian, Hebrew, Aramaic, Arabic. The number of lexifying languages is much greater, though.

This confirms what the researchers always suspected: the base language is made up of loan-words. Once we understood the mechanism that rules borrowing, I believe I am not far from the truth if I say that more than 50 to 60% of the total stock can be easily recognized and reconducted to the languages from which the borrowing took place.

Many of the loan-words we examined before do enable us to *date* the presumable epoch in which they were borrowed. Since the borrowings took place in what is considered to be the "most ancient" layer of Proto-Uralic, and owing to the fact that the datable borrowings came about in the lexifiers during the first centuries of the common era, the "Uralic" proto-language goes back *at the very earliest* to the seventh century of our era.

The basic language

Evidence shown above reinforces the idea that the wordstock of the so-called Proto-Uralic language is probably composed of loan-words only and that it lacks the autonomy that would be expected from an independent language family. As a matter of fact, any independent language family may include a certain number of borrowings from other languages, yet the bulk of its vocabulary should be shared with no other language family. When this happens, though, two possibilities only can come into account: a *genetic* relationship or a *loan*-relationship. In the case of

Uralic, we have seen that several distinct languages belonging to different families contributed in building the wordstock of "Proto-Uralic", thus making the recognition of a genetic relationship impossible.

It was only in recent years that a relatively small number of linguists begun to realize that the so-called U languages might be connected by something other than a "regular" genetic relationship. For example, the Finnish scholar W. Tauli wrote: *"The origin of all common traits of the Uralic languages is so far not known. The existence of a certain grammatical category or morpheme in separate languages need not indicate that this originates from a common proto-language or a parallel development of different languages – which is the classical point of view on the argument – but may be due to the spreading of the phenomenon from one language to other contiguous languages. Owing to such circumstances it is often difficult to establish whether the absence of a certain phenomenon in a language signifies that it has been lost, or that it never existed in that language. In addition, we have to take into account that the Uralic languages, as most languages, are structurally more or less mixed»* (Tauli, 1966). Similarly, the Estonian scholar Ago Künnap claims that *"From our point of view, any F-U language has always been and still is a mixture of languages and so it is not possible to reduce it only to an assumed affined Proto-Finno-Ugric"* (Künnap, 1997).

Not only words, but also grammatical and/or morphological structures were borrowed. Nevertheless, the borrowings took place independently in the individual U language that so were able to build their own grammar (process of grammaticalization), as in the following examples:

1. Cfr. the personal endings that are added to the verb stem in Hungarian and New Persian:

	Hungarian (definite conjug. and possessive suffixes)	*New Persian* (past tense)
1PsSg	–Vm	–m (–am مـ)
2PsSg	–Vd	–d (–īd ـيـد)
3PsSg	–ja, –i (poss.: –a, –e, –ja, –je)	–∅ (see note below)

N.B. Concerning the 3PsSg, compare the Hungarian form against the New Persian particle expressing possession, apposition and qualification, known as *ēδafé* or *ēzafé* اضا ف. The *ēδafé* is a short syllable, pronounced *é* after consonants and *yé* after vowels. Hungarian preserved it in its original conformation in constructs like *ez a ló az apámé* "this horse belongs to my father"[34].

---

[34] This sentence is made up of the following loan-words and morphemes: *ez* 'this' < Hebrew accusative mark את *eth*; *ló* 'horse' < Byzantine & Mediaeval Greek ἄλογον 'horse', a word that was possibly borrowed also in Turkic (*ulaγ* 'relay horse') and shows a phonological resemblance with Chinese *luò* 'a white horse with a black mane, mentioned in ancient Chinese books'; *apa* 'father' <

2. Cfr. the Finnish personal pronouns against their Greek counterparts:

	Finnish	Greek[35]
1PsSg	*minä* 'I'	μένα *ména* 'me'
2PsSg	*sinä* 'thou'	σένα *séna* 'thou'
3PsSg	*hän* 'he'	ἕ *hé* 'him'

A closer examination of the peculiarities of "Proto-Uralic" puts into evidence a loan situation reaching its utmost limits: the entire language is the result of borrowings from several different sources. Yet, these borrowings were considerably modified in the process. This might mean that the so-called Uralic languages are very likely the offspring of a *lingua franca* or *pidgin* language originally used for basic communication needs only.

This idea is not new. Kalevi Wiik (Turku), János Pusztay (Szombathely), and Ago Künnap (Tartu) assume that the common features of Uralic languages developed thanks to various contacts among different languages and that a language of the *"lingua franca"* type could have operated as an intermediary. Nevertheless, the main problem of their hypotheses is that they posited that this *lingua franca* went back to time immemorial, when the languages we know of were still *in statu nascendi*. In the present study we show that the *lingua franca* goes back to the VII–VIII century of our era and that it originates from several languages, many of which are well-known to us.

What a pidgin is

I deem necessary here to make clear to my Uralist colleagues what a pidgin language is. We can say that a pidgin is a *new language* that comes into being when groups of people speaking different languages come into contact for the first time. When this happens, they sometimes bring into existence a restricted language system in order to cater for essential common needs. A pidgin is frequently described as a "marginal" language, used by people who need to communicate for certain restricted purposes. For this reason, pidgins tend to arise along *trade routes*. This is the case, for example, of pidgins spoken along the costs of West Africa, in the Caribbean, and on Pacific Islands. A pidgin takes one or more existing languages as its point of origin. Many Pacific and West African pidgins are based on

---

Byzantine Greek ἄππα 'father'; *-m* possessive suffix 'my' < Persian *–m* possessive suffix 'my'; *-é* possessive suffix 'of his' < Persian *–é* "ezafé".

[35] The apheretic forms *ména, séna* are employed only after prepositions and adverbs ending in *a, o, e* (like ἀπό, κατά, γιά).

English, while a number of those found in the Caribbean are French based. Yet there are pidgins based on other languages, like Spanish, Portuguese and Arabic.

Peter Mühlhäusler had the great merit of shedding considerable light on the origins of Tok Pisin, spoken in Papua New Guinea, by showing that it is a product of the particular socio-economic conditions prevalent in Pacific in the last century. He quoted Governor Solf's diary for 1895 that included a number of relevant comments on the birth of this language: "*It is a well-known fact that almost every one of the various native islands of the blacks in the South Seas possesses not only one but a whole number of different languages… Thus, in what way do the workers from such different places and islands communicate, when thrown together in Samoa? They use that Volapuk of the South Sea, which has become international among whites and coloureds: pidgeon English… The words »belong« and »fellow« are especially important. The former used with nouns and pronouns indicates property, »house belong me, horse belong me« 'my house, my horse'… The latter is added to all numbers, without regard to the gender of the following noun «three fellow woman» 'three women', »two fellow horse« 'two horses'. It is incredible how quickly all blacks learn this lingua franca…*" (Mühlhäusler, 1978, 72).

A genuine pidgin must not be confused with a broken language, and it has consistent rules. Just as the rules of chess cannot be predicted from looking at the old Indian game from which it was adapted, so the rules of a pidgin cannot be deduced from the standard version of the lexifying language. A pidgin is a separate system, with an identity of its own. Bickerton (1981) as well as Thomason & Kaufman (1988), among others, analyzed pidgin and creole languages as the result of an abrupt break of grammar transmission of the lexifier language. Today there is a more or less general consensus on the fact that creole languages defy the *Stammbaum* (family tree) model, since they originate genealogically neither from their lexical base languages nor from the languages of their substrates.

A pidgin is not made up exclusively from elements of the base language, and vocabulary items are incorporated from native languages spoken in the area and from other areas further a field as well. A pidgin is, however, relatively easy to learn. It is simpler than a real language in two ways. First, it has a smaller number of elements. There are fewer sounds, fewer words, fewer constructions. There are relatively few vocabulary items, so the same word can mean a number of different things depending on the context. The time of an action is not normally specified, since verbs do not distinguish between tenses. In a true pidgin there is little or no embedding – that is the combination of two potential sentences by inserting one into the other does not normally occur. The second way in which a pidgin is simpler than a "real" language is that it is more *transparent,* in that it is nearer to the ideal of one form per unit of meaning, with systematic and easily detectible rules governing the alternation.

A pidgin is, as it were, a language in embryo, a foetus with the potential to become a full language, but not yet capable of fulfilling the entire communication needs of a human being. Some pidgins exist for a limited amount of time, and then

die out. Other get progressively more complex as the purposes for which they are used expand. Eventually there may come a time when the pidgin is learnt by someone as a first language. At this point it has become a creole.

Around the time of its "birth" as a creole, it grows rapidly and extensively. Some changes seem to occur before it is acquired as a first language, others are initiated by the new native speakers. There are at least four different types of alteration and expansion. The first involves the speed of speech, the second lexical expansion, the third the development of tenses and, finally, the development of relative clauses.

A creole is a "real" language in the sense that it is often the only language of those who learn it as their mother tongue. It therefore has to be capable of dealing with a greater range of communication needs than a pidgin. At first, it will be in a relatively immature state, and the language is likely to develop fast during the first two generations of creole speakers. Later, its rate of growth will slow down, as it becomes a fully mature language. In time, it will be a "normal" language, which takes its place among the thousands of others spoken in the world. Some Creolists believe that in the long run, there is no way to distinguish one-time pidgins and creoles from any other language. The Uralic languages show that this assumption might not be completely true.

"Universal" characteristics of a pidgin language

Pidginisation and creolisation are worldwide phenomena, yet pidgin and creole languages share some basic characteristics that seem to be totally independent of the language on which they were based. Certainly all of these features can be found individually in "real" languages. *The point is that where they cluster, it is symptomatic of the extreme reduction of one or more lexifier language.* This is especially clear when the lexifiers themselves have few if any of these traits. Of course, a pidgin is not required to have all of this or any similar list of features. However, we would expect most pidgins to display all or most of these features.

Morphology. The first and foremost distinctive characteristic of a pidgin is the lack of a morphology to such an extent that anything that the lexifier does morphologically is either lost or indicated periphrastically. There are no combinative rules for the formation of words, no flexion and no derivation. When a pidgin develops into a creole language, the little morphology that it has is typically recreated (grammaticalisation), rather than inherited from the lexifier. As far as the "Uralic protolanguage" is concerned, the linguistic system reconstructed by the researchers had no morphology at all.

Grammatical gender. A common characteristic of the pidgin and creole languages is that they lack grammatical gender. As a matter of fact, none of the U languages

developed grammatical gender (a characteristic that they share with the Altaic languages).

Lack of copula. The omission of the copula is a direct consequence of pidginisation and it is often used as a diagnostic trait (Ferguson 1971; Foley 1988:165), although it does not follow from this that every language lacking an expressed copula is a pidgin or the offspring of a pidgin language. It should be noted, though, that the U languages share the lack of copula both with their Altaic and Semitic lexifiers, as well as with Russian.

Congruence. In pidgin languages there is usually no congruence between the attribute and the denoted word. All the cognate languages bear witness to the fact that congruence in the Uralic languages is unknown. Full congruence can be found in Finnish only, while partial congruence exists in Estonian and Sami. Congruence at a very limited degree can be encountered in Mordvin and Jurak. In these languages, concordance in number and case is due – according to Hajdú and others – to a foreign influence and/or to an autonomous development.

Singular after a numeral. As often happens in pidgin languages, the name, after a numeral, is in the singular. This peculiarity is shared with the Altaic languages and, partly, with the Semitic languages, where numerals higher than 10 require the singular.

Conjunctions. Pidgins usually lack conjunctions and the sentences are structured with parataxis. According to Gy. Décsy, "*In der* [uralische] *Grundsprache gab es nur den einfachen Satz; keine der Konjunktionen der Einzelsprachen kann in die grundsprachliche Zeit zurückgeführt werden. Wie allgemein in »primitiven« Sprachen zu beobachten ist, dürfte die konjunktionslose bloße Aneinanderreihung von Aussagen als asyndetische Satzverbindung gegolten haben (»nebengeordneter zusammengesetzter Satz ohne Konjunktion«)*" (Décsy, 1965: 160). The U languages developed conjunctions independently one from the other (Hajdú, 1987: 1992). In some cases this process did not even take place. For example in the Samoyedic languages there are at most paratactic sentences, while there is no way to create hypotactic and subordinate phrases. Subordinate phrases are replaced by participial constructions grouped around a central *verbum finitum*, a solution that is well known in other U languages too, cfr. Finnish *hän teki itsemurhan viiltämällä* 'he committed suicide by cutting his throat'. It often happens that co-ordinate parts of speech in the U languages do have a predicate each, cfr. Samoyedic Yur. *ńiśāw χāś, ńeb'āw χāś, ńāw χāś* 'my father, my mother and my brother are dead' (pply. 'my father is dead, my mother is dead, my brother is dead'), Ostyak *taŋkə wetəs, ńoχəs wetəs* 'he killed squirrels and sables' (pply. 'he killed squirrels, he killed sables')[36].

Relative clauses. Pidgins usually lack relative clauses, which are implemented only at a much later stage of language development and/or after creolisation. The early stage of Hungarian, as shown by the *Halotti Beszéd,* indicates that use of the

---

[36] These examples were taken from Hajdú-Gheno, 1992: 252.

relative conjunction was still *in statu nascendi*, cfr. the Old Hungarian sentence: *latiatuc feleym zumtuchel mic vogmuc* 'you see brethren with your eyes what we are' (HB, XIII century), where the relative is lacking. The relative conjunction in modern Hungarian is *hogy,* originating from Greek ὅτι *hoti* of s.m. Yet, some U languages still do not use any relative conjunction at all (this trait is shared with the Altaic languages).

Palatalisation. Most of the well documented pidgin and creole languages tend to palatalise, although some researchers maintain that palatalisation was already present in the lexifier or in one of the substrate languages. If we compare the wordstock of the U languages and the corresponding words of their lexifiers, we realize the strong palatalisation it underwent.

Syntactic ambiguity, syntactic underspecification, cross-categorization, alias **nomina-verba.** In the past century, a striking peculiarity of the Uralic languages attracted the attention of the linguists. They noted that there are words that can be used both as nouns and verbs, for example Hungarian: *fagy* 'ice ‖ to freeze', *les* 'stalking, ambush ‖ to watch, spy', *nyom* 'trail, track, trace, spoor, footprint ‖ to press', *zavar* 'confusion, disorder ‖ to disturb, trouble'; Finnish: *neuvo* 'advice, counsel ‖ to advise', *kutsu* 'invitation ‖ to invite', *onki* 'hook ‖ to angle, fish', *toivo* 'hope ‖ to hope'; Zyrien *ger* 'plough ‖ to plough'; Vogul *kas* 'competition ‖ to compete' etc. This phenomenon was immediately classified under the name *nomen-verbum,* that is it was considered the relict of an "extremely ancient" linguistic layer that was still unable to distinguish between nouns and verbs, and which used a *neutral* category anticipating the nominal and verbal classes. By comparing the U daughter-languages, they discovered many bases that in some languages are used as verbs and in some others as nouns, as in the following example:

NOUNS	VERBS
Hung. *fej*, Finn. *pää* 'head'	Ńeńets *pa–* 'to begin'
Finn. *pala* 'mouthful, bite'	Hung. *fal–* 'to devour'
Hung. *íz* 'taste, flavour'	Sami *hâkse–* 'to smell, sniff'
Finn. *pura*, Ostyak *pәr* 'drill'	Hung. *fúr* 'to bore, drill'
Hung. *fagy* 'frost', Ostyak *poj* 'piece of ice'	Hung. *fagy–* 'to freeze
Hung. *zaj*, Vogul *soj* 'noise, sound'	Finn. *soi–* 'to sound'
Finn. *sula* 'fluid, liquid'	Vogul. *tol–,* Cherem. *šule–* 'to melt'
Vogul. *low*, Cherem. *lu* 'ten'	Finn. *luke–*, Mordvin *lovo–* 'to count, read'

Incidentally, most of these *nomina-verba* can be easily recognized as loan-words from various languages: Finn. *pala* 'mouthful, bite' and its cognates relate to the Aramaic/Hebrew root פלח *pālaḥ* 'to cleave open, to split, to slice (fruit), to break (bread)' (whence פלח *pĕlaḥ* 'part, slice', פלג *pĕlag, peleg* 'part, half'); Finn. *pura* 'drill' relates to the Akkadic root *puru* 'well; to drill, bore (a well, etc.)' – ori-

ginally a word of Sumeric stock that became an international loanword (cfr. also Turkic *bur-* 'to bore, drill', Tataric *borau* 'driller', Old Icelandic *bora,* German *bohren,* English *bore,* Latin *forō -āre* 'to drill, bore', etc.); Hung. *fagy* 'frost, to freeze' and its cognates originate from Arabic فجّ *fajja* and/or Aramaic פג *pag* 'to be or become cold', Finn. *luke-* 'to count, read' and the like originate from Greek λόγος 'count (=λογισμός), reckoning, account, etc.'

Nomina-verba, rather than an "archaic" and exclusive peculiarity of the Uralic languages, are the same phenomenon acknowledged by creolists as *syntactic ambiguity, underspecification,* or *word multifunctionality,* that is a typical characteristic of pidgin and creole languages. The phenomenon of multifunctionality, *i.e.* the same phonetical sequence may belong to more than a lexical category only, has been considered by Mühlhäusler (1997: 159–60) as one of the most typical traits of rudimentary pidgins. To a great extent, this is due to lack of morphology. Pidgin languages are based on parataxis, that is the juxtaposition of words conveying some sort of meaning but lacking a precise grammatical function or, better said, words that assume a grammatical function that is contingent to the sentence they have to express. The phenomenon of syntactic ambiguity in U is much more extended than expected. For example, in Samoyedic and Mordvin it is possible to use any part of the speech (noun, adjective, numeral, pronoun) as if it was a verb, without any need to add a deverbative suffix. The word in question may take any verbal personal and tense suffix, *e.g.:* Mordvin E *lomań* 'man': *lomań-an* 'I am a man', *lomańe–ľi–n* 'I was a man'; Eńets *ese* 'father': *ese-do* 'you are father', *ese-do-ś* 'you were father', etc. In the most recent Hungarian literature this phenomenon is called *"két- és többszófajúság"* or *"szófajváltás"* [approx. "switching of grammatical gender"] owing to the fact that not only verbs and nouns come into account but adjectives too, cfr. Hung. *agg* 'old man' and *agg(-ik)* 'to grow old'[37].

Verbal forms. Every pidgin or creole language enucleates a meaning-conveying root and isolates from it a zero form that can be used in a way structurally fit to the new language conventions. French, Portuguese and Spanish-based creoles for example generally use the 3PsSg or the infinitive form as a verbal base. The verb is then "conjugated" with time markers or by means of auxiliary morphemes. This is due to the fact that the pidgin languages behave like isolating languages, each word being a single morpheme. At this stage there cannot be any fundamental difference between isolating and agglutinating languages, a distinction that can only be made on the basis of written languages[38]. The borrowing of verbal forms in "Proto-Uralic" took place in the same way. Greek, Slavic, Persian and Germanic verbal forms were probably borrowed from the 3PsSg of the present mood of the indicative tense, while

---

[37] A. Molnár Ferenc: "A két- és többfajúság", in: Benkő, L. (ed.-in-chief): *A magyar nyelv történeti nyelvtana*. I. kötet. Akadémiai Kiadó. Budapest, 1991, 553 ff.)

[38] Cf. the orthographic problems encountered in Bantu languages: if "disjunctive" conventions are followed in writing, i.e. if clitic morphemes are not attached to the stem, these languages will appear very isolating, whereas if they are written "conjunctively" they will have an agglutinating outlook.

Hebrew, Aramaic and Arabic verbal forms from the 3PsSg of the simple past mood (Qal forms).

Reanalysis of word boundaries. It often happens in pidgins and creoles that the word boundaries of some words of the lexifier language are reanalyzed and "agglutinated" together. Take French creoles, for instance, where /lamɛ/ (< French "la mer") means 'sea' and not, as you might expect, *'the sea'. If you order a small beer in the Seychelles, you say something like /ɛn piti labiɛr, suple/ (< French *"une petite la-bière, s'il vous plaît"). In Haitian, 'aunt' translates with /matãt/ (< French "ma tante", 'my aunt'), and 'my aunt' is /matãt mwê/ (< French *"ma-tante moi"). Pidginisation and creolisation often involves morpheme boundary reanalysis, so that grammatical morphemes get stuck to lexical morphemes of the lexifier. Yet, the creole word is prototypically and almost by definition monomorphemic from the synchronic point of view. In fact, /lamɛ/, /labiɛr/ and /matãt/ are synchronically *one morpheme* only. The same phenomenon can be observed in the Uralic languages. In Finnish the word for 'sky' is *ilma* (yet it assumed secondary meanings too, like 'air', cfr. *ilma-aine,* the substance of which the sky consists; 'weather, bad weather; world', cfr. *maailma*, literally land-sky. Cognate words can be found in Lydian, Votic, Estonian, Livonian, several Sami languages, Votyak, Zyrien, Vogulic and Ostyak. The UEW (81) reconstructs the F-U base as **ilma* 'sky' While in Hungarian they use a word originating from Turkic *tängri* 'sky' for 'sea' (*tenger*), Finnish and the F-U cognates mentioned above use the *Arabic* word الماء *ăl-maʔ* 'the sea' to indicate the sky. This is confirmed by the fact that – in literary and elevate Arabic – الماء الكبير *ăl-maʔ ăl-kebir* 'the big water' and الماء الأزرق *ăl-maʔ ăl-aʔzraq* 'the blue water' means 'sky' indeed. As we see, the word boundaries of the Arabic word have been reanalysed to include the Arabic article, exactly as has happened in the French creoles cited above. This is not the only example, though. One of the key characters of the Kalevala is *Ilmari* (~ *Ilmarinen*), the mythical blacksmith. According to the SKES (105), the name of the god of the sky ("ilman jumalan nimi") originates from the word *ilma* 'sky, air'. This is false. In Finnish there are no –*ri* suffixes or words that could form a compound word having the meaning of 'lord of the sky' (*ilman jumala*). The only possible connection between *ilma* and *Ilmari* is their phonologic resemblance. As a matter of fact, the Finnish theonyms *Ilmari* and *Ilmarinen,* the Sami anthroponyms *Ilmaratshe* and *Ilmaris,* the Votyak word *ińmar, inmar, immar* and *ilmer* 'lord, god' go back to the Semitic root *măr* 'lord, master' (cfr. Aramaic and Syriac מר, מרא *măr, mără,* Hebrew מר *măr*) through Arabic المرء *ăl-maruʔ* 'the lord, the master'. Incidentally, إمرء *īmrŭ* is another form of Classic Arabic of the same word for 'lord, master', cfr. the Hungarian anthroponym *Imre.*

The base languages (lexifiers)

We have seen that the U languages possesses all the distinguishing marks of the offspring of a real pidgin language (based on more than one lexifier). Their word-stock is made up of loan-words from various languages. Some of these loan-words enable us to date the birth of the "Uralic" pidgin around the seventh or eighth century of the common era. There is an unavoidable question, then. Where did the "Uralic" language come about?

Since the very beginning of Uralic studies, the researchers realized that the so-called Uralic and Altaic languages are related by a number of isoglosses to such an extent that they thought that the two linguistic families went back to a common ancestor, which they called Uralo-Altaic. Many "Uralic" words have striking counterparts in Turkish, Mongolic or Tunguz. Rasmus Kristian Rask (1787–1832) maintained that the Uralic and Altaic languages are genetically related, an opinion shared in our days by the well-known Turcologist Karl H. Menges. M. A. Castrén (1813–1852), Wilhelm Schott (1802–1889) and Heinrich Winkler (1848–1930) gathered abundant lexical evidence. D. R. Fokos-Fuchs, M. Räsänen, A. Sauvageot, B. Collinder consider possible an ancient genetic relationship, while N. Setälä, H. Paasonen, G.I. Ramstedt, Y.H. Toivonen, L. Ligeti, P. Ravila, E. Itkonen and D. Sinor expressed the thought that the question is still open to discussion. In fact, about 70 words in each group – such as Finnish *käly* 'sister-in-law' and Ujghur *kalin* 'bride; daughter-in-law' – appear to be cognates. But the main problem was that the lexical correspondences between the two groups of languages were asystematic and they could have been the result of borrowing or chance. This is why the concept of possible affinity fell slowly into oblivion, until it was finally dissolved by Otto Donner.

Incidentally, since we mentioned the Turkic word *kalin* 'bride; daughter-in-law', it is easy enough to realize how striking is the resemblance with Hebrew כלה *kallā*, Ugaritic *klt,* Aramaic *kaltā,* Akkadic *kallatu* and *kallātu* 'bride, daughter-in-law'. And if we take a closer look at the Finnish and Estonian word *paha* 'bad'; Samoyedic *waevo, awoj, awaj;* Ostyak *paγaj* 'bad, poor, unclean' (a concordance not mentioned by the UEW) we could easily realize that it could be related to the following Turkic words: Tuva *paγaj, baγai;* Karaγas *bahaj, bak;* Ojrot *paγaj* 'bad, poor, low, unclean' (Menges, 1968: 179). Yet, these words might also be related to Hebrew פגל *pāgal* 'to foul, spoil, render unclean' and Aramaic פגל *pagel* 'to spoil, render unclean, render defective, make a sacrifice rejectable'. It goes much beyond the scope of the present study to deal with the origins of the Turkic word-stock. However it is quite strange that, despite the presence of a number of Aramaic inscriptions on archaeological finds discovered in the ancient kurgans of Kazakhstan (that is near the Altai mountains), no one ever thought of the possibility that some Turkic words might be of Semitic origin.

Besides lexical concordances, there are a number of structural concordances tying the Uralic and the Altaic languages together. Some of these were summarized

in 1838 by F. I. Wiedemann in 14 points, namely: 1. Presence of vowel harmony; 2. Lack of grammatical gender; 3. Lack of the article (with the only exception of modern Hungarian); 4. Agglutination; 5. Possessive flexion; 6. Abundance of suffixes for the formation of verbs; 7. Use of postpositions; 8. The attribute precedes the referenced word; 9. After a numeral attribute the referenced word is in the singular; 10. The ablative is used in the comparison; 11. Lack of a verb meaning *habeo* 'to have' (instead the *mihi est* form is used); 12. Presence of a negative verb; 13. The interrogative phrase is marked with a particle; 14. Verbal nouns and adverbs are generally preferred. To the isoglosses mentioned afore it is possible to add the following: 15. Absence of consonantic clusters in the initial position of the words; 16. SOV word order; 17. Use of nominalization to express subordinate phrases; 18. The possessive suffix is used to conjugate both names and verbs.

The Altaic-Uralic isoglosses enable us to posit that the so-called Proto-Uralic language was in origin a Turkic-based pidgin, or – taking into account the large number of Hebrew-Aramaic borrowings – that it developed out of Jewish Aramaic in an environment where the Turkic substrate was predominant. Its base word-stock is made up of elements originating from several Turkic dialects, Mongolian, Tunguz as well as Hebrew, Jewish Aramaic (Aramit), Arabic, Low Latin, Byzantine Greek, Middle Persian, early Slavic, Germanic (Gothic and Old Norse), as well as a number of Caucasian languages.

The origins of the "Uralic" pidgin

There has been just one epoch and one place in the world where so many languages could have met to give birth to a new language. And that place was the Xazarian Qaɣanate. Its territory spread from the Crimea (where Gothic was spoken), to the ancient Greek colony-states on the Black Sea, to the areas inhabited by Slavs that later were incorporated in the Viking-ruled state of Kiev (Rūs'), to several regions where Turkic and Hunnic populations lived – as well as areas where Persian was spoken. It included the one-time territory of the multi-ethnic state of Ermanarik and we may assume that the Xazarian pidgin developed out of a more ancient *lingua franca*

In the VII century, tribal groups known in the Greek sources by the name of Χάζαροι seem to have organized into a political unit. The Xazars are believed to be an originally nomadic Turkic group that reached the Russian steppes region from further east at some time not easily determinable. They may have belonged to the West Turkish Empire (from 555 OE) and, with the annihilation of the West Turkic Qaɣanate by the Tʿaŋ Chinese in 657–9, the Xazars became independent. At the time of Procopius Rhetor (VI century) the region immediately north of the Caucasus was held by the Sabirs who are referred to by Jordanes as one of the two great branches

of the Huns[39]. According to Theophanes, the Xazars, described as a *"great nation… from the interior of Berzilia in the First Sarmatia",* took possession of the territory as far as the Black Sea. and established a powerful military kingdom, that existed from the mid-7th to early-11th century. By 680, the Xazars had organized their own state, keeping to such ancient Turkic traditions as the double-rule[40], exerted there – according to Constantinos Porphyrogennetos – by the χαγάνος (= *qaγan*) and the πέχ (= *beg, beḡ*). According to Arab travelogues, the Xazarian name of the *qaγan* was اشا *iša* (يشان *išan* in Persian)[41] while the *beg* was called يلك *ilak*[42]. Arabic accounts of the Arab-Xazar wars afford no precise evidence of the dual kingship, yet the Arabic geographers regularly mention it. It is to be noted that the early Hungarian monarchies were also divided up into two kingships.

The three hundred years 650–950 OE. mark the epoch of the greatness of the Xazarian state. Surrounded by the Islamic Eastern Caliphate of Persia and the Christian Byzantine Empire, the Xazars chose Judaism as their state religion to avoid being religiously (and hence politically) dominated by either Empire. Their home was in the spurs of the Caucasus and along the shores of the Caspian – the "Xazar Sea" – and their cities, all of them populous and civilized commercial centres, were *Itil* (*Atil, Ätil*), the capital, in the delta of the Volga (the "river of the Xazars", whose name was also *Itil, Atil or Ätil*), *Semender* (Tarkhu), the older capital, *Khamlidje* or *Khalendzh*, *Belendsher*, the outpost towards Armenia, and *Sarkel* on the lower Don.

The great capital city of modern Ukraine, Kiev[43], at the Dnepr river, had been founded – according to O. Pritsak – by the Xazars around the beginning of the 8th century as a trading and administrative centre in the western part of the Xaza-

---

[39] "*Hunni quasi fortissimarum gentium foecondissimus cespes, in bifariam populorum rabiem pullulantur. Nam alii Aulziagri, alii Aviri nuncupantur, qui tamen sedes habent diversas. Juxta Chersonem Aulziagri, quo Asiae bona avidus mercater importat, qui aestate campos pervagantur effusos, sedes habentes, prout armentorum invitaverint pabula; hieme supra mare Ponticum se referentes. Hunungari autem hinc sunt noti, quia ab ipsis pellium murinarum venit commercium: quos tantorum virorum formidavit audacia*" (Jordanes, *Getica,* caput V). The *Aviri*, alias *Hunungari*, are generally acknowledged to be the *Sabirs* mentioned by Procopius (*Goth.* IV, 3) and by other Byzantine writers. According to Procopius (*Goth.* IV, 5) the Huns were composed by three branches: Cimmerians, Ujghurs and Kutrighurs.

[40] The double kingship was a phenomenon found among other Turkic peoples, for example the Qara-Khanids, and not unknown elsewhere: compare the double kingship at Sparta in antiquity, and the Shogun and Mikado of mediaeval Japan.

[41] As mentioned elsewhere in the present study, *iša* originates from Soγdian *ixšiđ* 'a high dignity' (Menges, 1968:168).

[42] *Ilak (~ ilik, ilek)* was also the appellation of the Ujghur king.

[43] There are different views concerning the origins of the place-name of Kiev, which has different names in the coeval sources: Arabic كويا بة *Kūjābā,* Hebrew קיוב *?Kiyōb,* VIII. c. anonymous *Kieu, etc.* Julius Prutzkus maintains it originates from Turkic *kui* = riverbank + *ev* = settlement (Pritsak, MNy. 80:16), while many etymologists explain it from Slavic **kyjь* 'large stick, mallet, double-hand hammer' or Polish *kujawa* 'place on an unfertile field; sand dune in the wood', Carpatian Ukrainian *куяба* 'mountain chasm or cleft; inaccessible place; ravine, gorge'.

rian empire. The Scandinavians accordingly called it *Changard* or *Könugård* (< Old Swedish "stronghold of the Kha[ga]n"). A community of Jewish Xazars lived in Kiev. Other towns of the Xazars, which also had important Jewish communities, included *Cherson, Chufut-Kale, Feodosia, Tmutorokan (Phanagoria), Olbia,* and *Sarkel*. The local governors (*tutun* 'civilian governor'[44]) of these cities and districts were usually Jewish. A major brick fortress was built in 834 in *Šarkhel (Sarkel)*, along the Don River.

Arabs and Xazars had already been in conflict on the line of the Caucasus (first Arab-Xazar war: 642–652, second Arab-Xazar war: 722–737). The Xazars were forcefully holding the Caucasian approaches at Därbänd and the Darjal-Gorge in the upper Terek valley against numerous Arabic assaults, thereby keeping the Arabs and their Islam out of Eastern Europe for a long time. The emerging Rūs' were also defeated, and tribute was exacted to allow Viking and Rūs' ships to pass through Xazar dominated waterways to raid for treasure in the Persian cities on the Caspian Sea coast.

Throughout the 6th century Xazaria was a mere highway for the wild hordes to whom the Huns had opened the passage into Europe, and the Xazars took refuge amongst the seventy mouths of the Volga. The rise of the first Turk empire in Asia (554) precipitated the Avars on the West. The conquering Turks followed in their footsteps (560–580). They beat down all opposition, wrested even the Crimean Bosphorus from the empire. The empires of the Turks and Avars, however, ran their course swiftly, and the Xazars arose out of the chaos, extending their rule over the Bulgarian hordes left masterless by the Turks and compelling the more stubborn to migrate to the Danube (641). The agricultural Slavs of the Dnieper and the Oka were reduced to tribute, and before the end of the 7th century the Xazars had annexed also the Crimea, won complete command of the Sea of Azov and, seizing the narrow neck which separates the Volga from the Don, organized the portage that has continued since as an important link in the traffic between Asia and Europe. The alliance with Byzantium was revived. Simultaneously and in concert with the Byzantine campaign against Persia (589), the Xazars had reappeared in Armenia, though it was not till 625 that this people take their place as Xazars in the Byzantine annals. In 627 Theophanes in his *Chronographia* mentions that "*the Turks from the East whom they call Xazars*" under their *beg* Ziebel passed the Caspian Gates (Derbend) and joined the emperor Heraclius at the siege of Tiflis. They are then described as a powerful nation which held the coasts of the Caspian and the Euxine, and took tribute of the Viatitsh, the Severians, and the Polyane. The qaɣan furnished Heraclius with 40,000 men for his Persian war, who shared in the victory over Chosroes at Nineveh. In the interval between the decline of the Mohammedan empire and the rise of Russia the Xazars reached the zenith of their power. Merchants of Byzantium, Armenia and Baghdad met in the markets of Itil (to where the capital had been transferred from Semender,

---

[44] As mentioned elsewhere in the present study, common Turkic *tutuŋ* originates from Chinese *ḏu–t'uŋ* 'civilian governor' (Menges, 1968:169)

since the raids of the Mohammedans), and traded for wax, furs, leather, and honey that came down the Volga. In Constantinople this traffic was held as so important that, when the Don portage was endangered by the irruption of a fresh horde of Turks (the Petchenegs), the emperor Theophilus himself dispatched the materials and the workmen to build for the Xazars a fortress impregnable to their forays.

Graffiti on potware found in Sarkel

In 833, the Byzantines helped the Xazars in the erection of a stronghold against the Rūs', the Russo-Normans and the dangerous nomads. Petronas Kamateros, a Greek, served as chief engineer during the construction. Famous as the only stone structure in that stoneless region, the post became known far and wide amongst the hordes of the steppe as *Šarkhel* or the "White Abode" (the **Бѣла Вежа** "White Tower" of the Old-Russian chronicles, and the مدينة البيضا *Medina-t-al-Baiḍa* "White City" of the Arabs). Merchants from every nation found protection in the Xazar cities. The Jews and the Pagans who refused to become converts to Christian faith, expelled from Constantinople, sought a home amongst them and developed the Xazar trade. Although the dynasty accepted Judaism there was equal tolerance for all, and each man was held amenable to the authorized code and to the official judges of the faith which he professed. At the Byzantine court the Qaγan was held in high honor. The emperor Justinian Rhinotmetus took refuge with him during his exile and married his daughter (702). Justinian's rival Bardanes in turn sought asylum in Xazaria, and in Leo IV. (775) the grandson of a Xazar sovereign ascended the Byzantine throne. Xazar troops were amongst the bodyguard of the imperial court; they fought for Leo VI. against Simon of Bulgaria (888); and the Qaγan was honoured in diplomatic inter-

course with the seal of three *solid* [45], which marked him as a potentate of the first rank, above even the pope and the Carolingian monarchs. Indeed his dominion became an object of uneasiness to the jealous statecraft of Byzantium, and Constantine Porphyrogenitus, writing for his son's instruction in the government, carefully enumerates the Alans, the Petchenegs, the Uzes, and the Bulgarians as the forces he must rely on to restrain it.

The extent of the territory ruled by the Xazars has been variously estimated, but the normal Xazaria may be taken as the territory included between the Caucasus, the Volga, and the Don, with the outlying province of the Crimea or "Little Xazaria". At one time Xazar rule extended westward a long way beyond the Crimea-Caucasus-Volga region. East of the Volga, in the direction of Xwārazm, the situation is obscure. The southern boundary never greatly altered; it did at times reach the Cyrus and the Araxes, but on that side the Xazars were confronted by the great power of Byzantium and Persia, and were for the most part restrained within the passes of the Caucasus by the fortifications of Dariel. Amongst the nomadic Turkic tribes and agricultural Slavs of the north their frontier fluctuated widely, and in its zenith Xazaria extended from the Dnieper to Bulghar upon the Middle Volga, and along the eastern shore of the Caspian to Asterabad.

As the date when Xazars converted to Judaism the year 740 is suggested by converging considerations, which might or might not be true. A certain number of Jewish communities in the Crimea and on the Taman Peninsula with the city of Kerč, which were now included in the Xazar realm, were instrumental in bringing this about. Another possible source of Judaism were the Jews of Xwārazm (Persia). However, the majority of Xazars did not profess Judaism, but were divided into Christian, Muslims, and pagans. The soldiers in the Xazar army were mainly Muslims, and the Slavs, Bulgars, and other ethnic groups within the Xazar empire were also not Jewish, but in the 10th century the Xazar tribe is described by most sources as fully practicing Judaism. It is believed that Judaism gained a stronghold among the common Xazar people starting in the late 9th century.

Arab travelogues provide useful contemporary details about the life of the Xazars. Armenian, Slavic, and also Hebrew sources form the core of our knowledge about the Xazar people. But there is more left to discover. Within the past few decades, archaeological excavations in Russia and Ukraine have unearthed Xazar jewelry, pottery, gravesites, and tombstones containing engraved menorahs and Turkic tribal symbols. One of the most famous sites was Šarkhel, which was later flooded for a dam, and is not available for further research[46]. Currently, efforts are underway to locate the precise site of the Xazar capital of Itil; some believe the wall which surrounded Itil has been found underwater, while others associate Itil with a hill in Daghestan. Primary sources are: *The Kievan Letter*, written by the

---

[45] The *solidus* was a Byzantine golden coin.

[46] See Flyorova (Флёрова, 1997) for the graffiti material found in Šarkhel.

Xazar Jews of Kiev in early 10th century[47], found in the Cairo Genizah; the *Xazar Correspondence* between Qaɣan Joseph and Ḥasdai ibn Shaprut of Spain[48] as well as the Schechter Letter, found in the Cairo Genizah[49], a semi-historical account of the conversion of Xazars to Judaism and of the migration of Jews to Xazaria.

It was, however, from a power that Constantinos Porphyrogennetos did not consider that the overthrow of the Xazars came. Long before, when a band of Slav prisoners was brought into the Xazar camp, a sage had prophesied: "*These men's swords have two edges; ours have but one. We conquer now; but some day they will conquer us.*" The arrival of the Varangians amidst the scattered Slavs (862) had now united them into a nation and launched them upon that career of conquest which within a hundred years carried Russian arms to the Balkans and the Caucasus. The advance of the Petchenegs from the East gave the Russians their opportunity. Before the onset of those fierce invaders the precarious suzerainty of the qaɣan broke up. By calling in the Uzes, the Xazars did indeed dislodge the Petchenegs from the position they had seized in the heart of the kingdom between the Volga and the Don, but only to drive them inwards to the Dniepr. The Hungarians severed from their kindred and their rulers, migrated to the Carpathians, whilst Oleg, the Rūs' prince of Kiev, passed through the Slav tribes of the Dniepr basin with the cry "Pay nothing to the Xazars" (884). On several occasions, notably c. 913 and again in 943, the Russians made raids down the Volga, passing through Itil. Apparently in 965, Xazaria was the object of a great Russian attack, which was aimed at the Xazar capital Itil and reached as far as Samandar, as we know from Ibn Ḥawqal. The Xazars appear to have recovered only partially from this disaster. The kingdom dwindled rapidly to its ancient limits between the Caucasus, the Volga, and the Don, whilst the Russian traders of Novgorod and Kiev supplanted the Xazars as the carriers between Constantinople and the north. When Ibn Fadhlān visited Xazaria forty years later, Itil was still a great city, with baths and market-places and thirty mosques. But there was no domestic product nor manufacture. At the assault of Sviatoslav of Kiev, whose troops were equally at home on land and water, the rotten fabric crumbled into dust. Šarkhel, Itil, and Semender surrendered to him (965–969). He pushed his conquests to the Caucasus and established Russian colonies upon the Sea of Azov. The principality of Tmoutorakan, founded by his grandson Mstislav (988), replaced the kingdom of Xazaria, the last trace of which was extinguished by a joint expedition of Russians and Byzantines (1016). The last of the qaɣans, Georgios Tzula, was taken prisoner. A remnant of the nation took refuge on an island in the Caspian (Siahcouyé); others retired to the Caucasus; part emigrated to the district of Kasakhi in Georgia, and appear for the last time joining with Georgia in her successful effort to throw off the yoke of the Seljuk Turks (1089).

---

[47] Голб-Прицак (1982 and 1997).

[48] Marcus (1938).

[49] Голб-Прицак (1982 and 1997).

After 965, the Xazars are still mentioned occasionally, but scarcely for long as an independent people. Oleg, head of the Rūs', is called in a seal of the XI–XII century "archon of all Xazaria", whatever this may mean. An important Xazar community remained in Kiev. In addition, western Jewish traders joined the Xazars by settling in Kiev by 1018. In the 11th century, Xazar Jews were Slavicized. They adopted the Cyrillic script in place of Hebrew and spoke east Slavic (probably Ukrainian, sometimes called "Canaanic" since Benjamin of Tudela called Kievan Rūs' the "Land of Canaan"). A messianic letter from 1096, written by Rabbi Nissim, says that during the earliest Crusades seventeen communities of Xazars left their native land for "the Wilderness of the Nations". The Cairo *Genizah* document published by J. Mann tells of a messianic movement supposedly in Xazaria but more likely in Kurdistan at the time of al-Afḍal, the great Fatimid vizier who ruled 1094–1121. Sephardi (Spanish) Jews met Xazar Jews in Toledo in the 12th century. The Xazar state probably subsisted until the second half of the tenth century, or the eleventh century at most. By the XII century the Qipčaqs or the Cumans appeared in the steppes once ruled by the Xazars. At the time of the Mongol invasions in the thirteenth century, it was they, not the Xazars, who were in possession. Family traditions indicate the likely persistence of Xazar Jewish settlements in Hungary and Transylvania. Some descendants of the Xazars may still live in the north Caucasus. We also know that Xazar settlements existed in Turkey, Egypt, and on the borders of Azerbaijan.

The Xazars are usually called Turks or are classified with the Turkic peoples. In this form this is certainly incorrect, for the Xazars included groups of different national and linguistic origin (as did all the nomadic groups of the area). Coeval sources mention the fact that two anthropologically different groups constituted the national body: one fair-skinned and light in type, and the other very dark. The latter group is possibly related to the ἡ μαύρη λεγομένη Βουλγαρία "the so-called Black Bulgaria" mentioned by Constantinos Porphyrogennetos or the *Uΐιηȷqhȷ* Sevordik – "Black Sons" – of the Armenian chronicles. Recently, an archaeological find in Xwārezm brought to light the remains of a garnison composed by negroids (Boulnois, 1992: 209) that were very likely brought there by slave traders. This means that we have to possibly reckon with a few words of African and/or Dravidian origin too[50].

The language of the Xazars

The question of the language spoken by the Xazars was raised by many researchers. The main historical sources are Arab travelogues, but – since the original works were lost – we know their contents only because later geographers quoted the

---

[50] Cfr. for example the words *tate* 'father', *naine* 'mother', appearing in many Bantu dialects, against U *ćečä* 'uncle, grandfather' (UEW: 34) and F-U *naje* 'woman, wife' (UEW: 297).

original sources in their own works. One of the most important sources in this connection is the Al-Balhi tradition. Al-Balhi was born approx. 850 C.E. in the city of Balk (Xorāsān, Xwārezm) and studied in Iraq. After returning to his birth-place, he wrote his geographical work, that was partly preserved in the works of Al-Istakhrī[51], who published a new version of Al-Balhi's work under the supervision of Al-Balhi himself in 930–933 and 950, and by Ibn Ḥaukal[52], who complemented Al-Istakhrī's work in the X century. Al-Istakhrī mentioned the fact that:

<div dir="rtl">

ولسان بلغار مثل لسان الخزر ولبرطاس

لسان آخر وكذلك لسان الروس غير لسان

الخزر ولبرطاس

</div>

"The language of the Bulghars is similar to Xazarian, while the Burṭāssians have another language. At the same time, the language of the Rūs' is different from both Xazarian and Burṭāssian".

On the other hand, Ibn-Ḥaukal wrote: ولسان بلغار كلسان الخزر *"the language of the Bulghars is like that of the Xazars"*, but he added that غير لسان الترك وا لخزر ا لخلص لسا نهر *"the language of the Xazars is different from that of the Persians and Turks"*.

Naturally, since the Xazars are thought to be a Turkic stock, among modern researchers there is a more or less general consensus on the fact that they "necessarily" spoke a Turkic language, and the sentences mentioned above were considered a proof thereof. These researchers simply forgot to consider the fact that, to an Arabic ear, the stress falling on the first syllable, the affricates and the rounded vowels of two different languages – as Turkic and Xazarian possibly were – might have sounded very similar.

Yet, another Arabic source reported a different version. Ahmed ibn Fadhlān (ibn Abbas ibn Rāšid ibn Hammad) was a member of the embassy sent by Munkhtedir, caliph of Baghad, to the Volga Bulghars in the years 921–922[53]. The members of the embassy communicated through interpreters who knew the languages concerned. Among his fellow-travellers there were also a Rūs' named *Susan the Rūs'* (الروسى ar-Rūsī) and a Xazar named *Ibn Baštu the Xazar* (الخزري al-Ḥazarī). Thus, Ibn Fadhlān was perfectly aware of what he wrote and his report should have

---

[51] Critical edition: M. J. de Goeje (ed.): *Bibliotheca geographorum arabicorum. Vol. I.* Lugduni Batavorum, 1870–1892. Partially reported in Pauler, Gy. – Szilágyi, S. (eds.): *A magyar honfoglalás kútfői.*. MTA. Budapest, 1900. Reprint: Nap Kiadó. Budapest, 1995.

[52] Critical edition: M. J. de Goeje (ed.): *Bibliotheca geographorum arabicorum. Vol. II.* Lugduni Batavorum, 1870–1892. New edition: J. H. Kramers (ed.): *Bibliotheca geographorum arabicorum. Vol. II.* 2nd edition. Lugduni Batavorum, 1939. French translation: J. H. Kramers and G. Wiet: *Ibn Hauqal. Configuration de la terre I–II.* Beyrouth–Paris, 1964.

[53] Frähn, C. M.: *Ibn-Foszlan's und anderer Araber Berichte über die Russen älterer Zeit.* Sankt-Peterburg, 1823.

enjoyed much more credit than what it actually had. Ibn Fadhlān reported the following:

<div dir="rtl">

ولسـان الخـزر غـير لسـان الترك والفـارسـية

ولا يشـاركـه لسـان فريق من الامم

</div>

"*the language of the Xazars is similar to neither Turkish, nor Persian nor any other language in the world*"

About the year 740 (?), many of the Xazars became converts to Judaism. The report of the existence of a Jewish kingdom aroused the curiosity of בן-יצחק יבן-שפרות חסדאי Ḥasdai (or Ḥisdai) ben-Itsḥaq ibn-Šaprūt (about 915–970). Ḥasdai was not only the personal physician of the Spanish caliphs Abd-al-Raḥmān III (912–961) and his son Ḥakam II (961–976) but was also inspector-general of customs and an adviser in foreign affairs. To satisfy his curiosity he wrote to the ruler of the Xazars about 960 and some time later received an answer from Joseph, the reigning king. The letters of Ḥasdai and Joseph were both originally written in Hebrew. Among the many questions he posed in his letter, Ḥasdai asked the Xazarian king: "*in which language do you express yourselves?*". The letter of King Joseph contained many details, yet he seems to leave this question unanswered. There is no direct reply to the question of the language the Xazars spoke, but the answer *is* indeed in the letter, its meaning being committed to the understanding of the reader. We shall examine the contents of the letter at the end of the present study.

In the coeval sources we find just a few embedded words of Xazaric. The letter of King Joseph contains some Hebrew words used in their Arabic meaning, thus showing a strong Arabic influence, like the words דין *dīn* (Hebrew = 'judgement, verdict') used in the same meaning of the Persian and Arabic word دين *dīn* 'religion'; מדינה *medīnā* (Hebrew = 'country') used in the meaning of Arabic مدينة *medīnā* = 'city'). Besides, the Arabic word القاضى *al-qāḍī* 'judge; a civil judge among the Turks, Arabs, Persians, etc.; usually the judge of a town or village' was transliterated in Hebrew as אלקאצי (*alqaci, alqaṣi* or *alqači*). The letter צ *ṣadhe* lacks the diacritic point on top which would be required in order to modify its sound. This might show that the interdental consonants underwent a peculiar treatment, but could also be due to an error of the later copyist.

Compare nevertheless the following ethnonyms and place-names appearing in King Joseph's letter (long version), where the צ *ṣadhe* replaces consistently the *č* affricate as in סמקרץ *Samkerč,* קרץ *Kerč* (Kerch, Crimea, Ukraine)*,* but also the sibilant *z* as in גרוצין *Gručin* (=Gruzinov, Rostov province, Russia). This seems to comply with the U constraint concerning the sibilants we mentioned above.

A peculiarity of the place-names mentioned in the Letter is that, whenever a name ends in -*a,* an Aramaic ending is used (i.e. א *aleph* instead of the expected Hebrew ה *hē*). In connection with the name of river Danube, *i.e.* דונא *Dūnâ,* many authors noted that it coincides with its present-day name in Hungarian.

A list of "Xazarian" words was drawn up by Z. Gombocz (1960: 22) and a more complete one by P. Golden (1980). The origin of some of these words is unknown, while some other words are attested in the coeval Turkic sources or can be found in other Turkic languages, like the words *tutun* 'civilian governor', cfr. Common Turkic *tutuŋ* < Chinese *ḍu–tʿuŋ* 'civilian governor' (Menges 1968: 169) and *iša, išaḍ* 'the Xazarian viceroy' which probably originates from Soγdian *ixšiđ* 'a high dignity', whence common Turkic *šad* of s.m. (Menges, 1968:168). Marquart[54] supplied us what seems to be a proof of the mixed character of the Xazarian language in connection with the other name by which the capital city of Xazaria, *Itil*, was known. The place-name concerned appears as سا رغشن *Saryšan* in the source works of both Ibn Rusteh and Gardīzī, a lection emendated into سا رغشر *Saryšar* by Á. Vámbery. Marquart showed that the place name is composed by the Turkic word *šarïγ* 'white' and the Persian word شهر *šahr* 'city'.

The name of the game

The name of the Xazars is frequently written and pronounced with an *a* vowel, as in Greek Χάζαροι and Arabic ا لخزر *al-Xazar,* but there are traces of different pronunciations. Hebrew has it as כוזר *Kozar, Kuzar.* Besides Χάζαροι, in Greek we find also the form Χάτζιροι, where the Greek digraph τζ is the representative of **ć* or **č.* To find the origin of the ethnonym, researchers scrutinized every possible vocabulary entry, from Turkic *qaz-* 'to wander' ('nomadize'?), *quz-* 'side of mountain exposed to the north' (the Caucasus?), *qač-* 'to run', *qadir* 'wicked, aggressive, violent', *aγač* 'woodman', to Ossetic *qazar* 'dear', to Latin *Caesar* (through Tibetan *Gesar*), and so on. In two fine articles, A. Róna-Tas summarized the hypotheses presented so far to explain the etymology of this name (Róna-Tas, 1981 and 1985). It is odd enough that in the case of a people that opted for the Jewish creed, no researcher ever thought of a *Hebrew* ethnonym.

In Hebrew the verb חזר *ḥāzar* 'to go round, to return' is endowed with a particular meaning. It occurs for the first time in the *Sanhedrin* treatise of the *Talmud Bavli* (Babylonic Talmud, ca. 350 OE), where (בתשובה) חזר *ḥāzar (bitěšūvâ)* means 'he repented, returned to the faith'. Here too, a verb in the 3SgPs of the simple past (Qal) is used, exactly as in the case of U forms. Since the ethnonym appears before the alleged date of conversion of the Xazar Qaγan to Judaism, it is possible that it originally referred to a group of Radanite Jews that became converts to rabbinical Judaism and influenced with their conversion the following conversion of the Qaγan. The Xazars became converted approx. in 740 but the ethnonym appears earlier in the sources. This might mean that the ethnonym originally referred to a group of semi-judaized people or professing a syncretic Mosaic creed that converted to a

---

[54] WZKM XII:194.

more orthodox form of Judaism (rabbinical Judaism?), as in the case of the Rādhā-nites or the Calisians (Χαλίσιοι) that we shall mention later.

In his letter to the king of the Xazars, Ḥasdai calls them אלכזר *al-Khazar,* that is he refers to the Arabic lection الخزر *al-Xazar.* Yet, he uses a כ *khaph* instead of a ח *ḥēth,* possibly to avoid that the ethnonym of his co-religionists might suggest the Hebrew word חזיר *ḥazīr* 'pig, swine'. Later, Hebrew writers added a ו *wāw,* so as to avoid any possible misunderstandings, thus the Hebrew name of this tribe became כוזר *Kozar, Kuzar* (Kohn, 1881: 17). The introduction of a *wāw,* representative of an *ō* or *ū* was probably a choice due to the fact that some Hebrew dialects pronounced the long a-vowel (*qāmāṣ gādhōl*) as if it were a short o-vowel [ɔ] (*qāmāṣ qāṭān*) – and the Massoretic symbol for both vowels is one and the same, *i.e.* (ָ). Thus, the verb concerned can also be read as *ḥozar.* This may account for the Hebrew *Kozari ~ Kuzari* form, yet, it does not explain why Xazar settlements in Hungary were called *Kozár* (cfr. *Kozárd, Kozármajor, Kozármisleny, Kozárvár, Nagykozár, Kozár, Egy-házaskozár*) instead of **Hazár,* unless we assume that either the new Hebrew lection כוזר *Kozar ~ Kuzar* got fashionable and became implanted among the Xazars who settled in Pannonia together with the Magyars or the Magyars used the Slavic name *Kozar* as it occurs in Nestor's chronicle. The Slavic syllables do not allow /x/ as a trailing consonant and /a/ is changed to /o/ in the first syllable (e.g. Latin *paganus* 'heathen' > Old Slavonic **поганъ, поганинъ**, Russian *погань,* etc.)

If Ḥasdai intended to avoid using the ethnonym חזר *Xazar* because it was too closely reminiscent of the word חזיר *ḥazīr* 'pig, swine'*,* he was right. In fact, one of the verbal forms of חזר *ḥāzar* is חזיר־ *ḥazīr-,* cfr. the noun חזירה *ḥăzīrā* 'a coming back, returning' formed from the same root (with the ה– substantive suffix). This verbal form accounts for the lection Χάτζιροι *Xáčiroi* that we find in some Greek authors. It is important to note that this variant seems to obey the *same linguistic constraint* we hypothesized for the *z* sibilant in the U languages!

How did the Xazarian pidgin spread

The Magyars. As far as the Magyars are concerned, they represent an exception among the U peoples. The language they spoke was an acrolect, that is to say that the forms concerned were – from a phonological point of view – the nearest to the lexifier languages. The Magyars, that were called "Turks" by most contemporary sources, learned the Xazarian pidgin directly from the very source, as testified by the Byzantine emperor Constantinos Porphyrogennetos in his work *De Adminis-trando Imperio* (Moravcsik, 1988, 46):

§ 39. Περί τοῦ ἔ νους τῶν Καβάρων. Ἰστέον, ὅτι οἱ λεγόμενοι Κάβαροι ἀπό τῆς τῶν Χαζάρων γενεᾶς ὑπῆρχον. Καί δέ συνβάν τινα παρά αὐτῶν ἀποστασίαν γενέσθαι πρός τήν ἀρχήν αὐτῶν, καί πολέμου ἐμφυλίου κα ἱστάντος, ἡ πρώτη ἀρχή αὐτῶν ὑπερίσχυσεν, καί οἱ μέν ἐξ αὐτῶν ἀπεσ-

φάγησαν, οἱ δέ ἐξέφυγον, καί ἦλ αν καί κατεσκὴνωσαν μετά τῶν Τούρκων εἰς τήν τῶν Πατζινακιτῶν γῆν, καί ἀλλήλοις συνεφιλιὼ ησαν, καί Κάβαροι τινες ὠναμάσθησαν. Ὅ εν καί τήν τῶν Χαζάρων γλῶσσαν αὐτοῖς τοῖς Τούρκοις ἐδίδαξαν, καί μέχρι τοῦ νῦν τήν αὐτήν διάλεκτον ἔχουσιν· ἔχουσιν δέ καί τήν τῶν Τούρκων ἑτέραν γλῶσσαν.

"*[39] About the nation of the Kavars. You ought to know that the so-called Kavars were of the race of the Xazars. Now, it fell out that a secession was made by them to their government, and when a civil war broke out their first government prevailed, and some of them were slain, but others escaped and came and settled with the Turks in the land of the Pechenegs, and they made friend with one another, and were called Kavars*[55]*. And they taught to these Turks the tongue of the Xazars and to this day they have the same language, but they have also the other tongue of the Turks.*"

The core of the "Turks" (Magyars) originally spoke a Turkic dialect indeed, just as Ármin Vámbery maintained in vain all through his life. We know for sure that a few years after settlement in Pannonia Hungarian legates were sent to Byzantium and the interpreters' office of the Byzantine court reported that the members of that mission spoke *Turkic*[56]. Yet, the Magyars learned easily and quickly the Xazarian pidgin that granted greater ease of communication within their community – composed by speakers of many different languages – as well as with the outside world. By the time they reached Pannonia a few years later, the Xazarian pidgin had already grown into a regular creole language among the Magyars. Besides its Turkic substrate, it was probably during the pre-creolization period that it picked up a relevant number of Persian and Greek words. The so-called "Proto-Hungarian age" goes back to the first half of the IX century only.

It is a well-known characteristic of pidgin languages that around the time of their 'birth' as a creole, they grow rapidly and extensively. Some of the changes seem to occur before they are acquired as a first language, others are initiated by the new native speakers. Although realizing that – according to the evidence supplied by the linguistic monuments – several phenomena took place simultaneously in the Hungarian language immediately before or shortly after the settlement of the Magyars in Pannonia, Hungarian philologists were unable to piece them together into a general view. For example, they realized that the Hungarian words embedd-

---

[55] The ethnonym Kabar means "to be great, be much, be many", cp. Hebrew רבכ *kābar* and Syriac רבכ *kĕbar* of s.m., Arabic كبر *kabbara* 'to be great, increased', Akkadic *kabāru* 'to be great, huge, mighty'.

[56] "*Sinor Dénes magyar származású amerikai egyetemi tanár igen találóan mutatott rá a Körösi Csoma-év akadémiai megnyító ünnepségén arra, hogy krónikás, császár tévedhet, de tolmácsiroda nem tévedhet abban, hogy milyen nyelven tolmácsolt. Amikor tehát a honfoglalás után magyar követség járt a bizánci császári udvarnál és törökül beszélt, ezt tényként kell elfogadni.*" (László, Gy. — Szombathy, V.: *Magyarrá lett keleti népek.* Panoráma. Budapest, 1988, 16).

ed in the Latin text of the TA[57], written approx. one hundred years after settlement in Pannonia, show a paratactic sentence structure with a remarkable presence of *free morphemes* (cfr. phrases like *feheruuaru rea meneh hodu utu rea,* pply. 'Fehérvár toward going troops road toward', i.e. 'in the direction of the military road heading to Fehérvár'; *monarau bukurea* pply. 'hazelnut shrub-toward', i.e. 'in the direction of the hazelnut shrub'). In spite of this, the Hungarian linguistic school speaks of a generic "nominal origin of suffixes and derivations". In the linguistic monument mentioned afore, the hypocoristic *-di* suffix (cfr. *holmo*di, *fotu*di) and the 3PsSg personal suffix (cfr. *baluuan ~ baluuan*a, *kut ~ kut*a, *cut*a) are well represented and seem to show stable crystallization. Nevertheless, the first text written in Hungarian (namely the *Sermo super sepulchrum,* Halotti Beszéd) appears only in the XIII century. *The Historical Grammar of the Hungarian Language* (TörtNyt. 1991), although heavily biased by the usual, traditional prejudices, mentions a few undeniable facts: "*the early layer of the verbs*[58] *was conjugated without affixes… the suffix was simply apposed to the plain verbal root…", "roughly around the time of the settlement in Pannonia it became a linguistic need for the loan-verb forms to fit the Hungarian system of verbal roots…", "[after the settlement in Pannonia] the number of borrowings grew remarkably …*" (p. 56), "*…at the end of the Proto-Hungarian age, our language possessed only monoelementary, archaic affixes…*" (p. 60), etc.

Around the time it was growing into a creole language, the one-time Xazarian pidgin was lexicalized, morphologized and grammaticalized by using Turkic, Persian[59], Arabic[60] and Greek forms. After settlement in Pannonia, the Hungarian language was improved by expanding the basic vocabulary and grammar to include a great number of Slavic, Latin and German lexical and grammatical items.

Other "Uralic" languages. It seems that no one ever realized that many of the so-called "Uralic" languages are geographically situated along the edges of the ancient Silk Road and on the main trade roads of Mediaeval Eurasia. It is enough to lay a

---

[57] Tihanyi apatság alapítólevele, ca. 1055.

[58] Here the extensors of the TörtNyT refer to alleged "onomatopoeic" verbs. Unfortunately, if we believed the traditional views of the Hungarian school, most of the roots for which no etymologies could be found are classified in the category of the words "descriptive of an atmosphere" or, at best, in the category of onomatopoeses (sic).

[59] Besides the Persian possessive and verbal forms mentioned before, cfr. *e.g.* words like Hungarian *kutya* 'dog' (which, according to the TESz, is an "onomatopoeic" word!) against Prakrit *kutta-, kuttī-;* Sogdian *'kwt-;* Xwārezmi *'kt;* Ossetic D. *kui,* I. *kwdz;* Yagh. *kut;* Shughni *kud;* Yazgh. *kwod;* Sanglichi *kud* < Proto-Iranic **kuta-, kuti-* 'dog'.

[60] Cfr. for example words like Hung. *hasáb* 'billet (of firewood), wooden log' < Arabic خشب *ḥašab* 'wood', Hung. *burok* 'caul, amnion; cover, wrapper' < Arabic بورکه *burka* 'caul, amnion; a dress that covers the entire body', as well as the 2PsSg verbal forms in *-lak, -lek,* like *visz-lek* 'I take *you*', *szeret-lek* 'I love *you*' where the 'you'-suffix results from the agglutination of the Arabic pronoun لك *lek* 'you' (acc. and dat.) (Agostini, 1996).

map of the present-day locations of the U languages over a map of the Mediaeval routes and we see how these overlap almost perfectly.

Map of the main routes of mediaeval trade

For many centuries, the Silk Road represented the main commercial route leading from Szechwan, China, to the Roman Empire. Actually, there were two distinct itineraries. The best known part of it was the Southern Road, that started from Ch'aŋ-ŋan in Szechwan through Li-Hien, Ansi, Niya and, after crossing the Kushana Empire, passed through Merv, Shahrud and Hamadan and reached the coast of the Mediterranean Sea at Palmyra and Antiochia. The northern route branched off at Ansi and headed to Turfan. It skirted the southern shores of Lake Balkhash first, then the northern shore of Lake Aral and entered Xazaria on the northern coast of the Caspian Sea (the "Xazarian Sea"). After reaching Itil, the capital city of Xazaria, it bifurcated. One branch passed through Kiev, the other followed the course of the rivers Volga and Kama to reach Bulghar, the capital of the Volga Bulgarians, then it continued northeastward to the mysterious country of *Wisu,* men-

tioned in the Scandinavian reports, a journey of three months distance from Bulghar. The two main branches of the Road joined again in Novgorod, heading then to the Baltic Sea and beyond.

The Xazars' language seems to have grown out of a former *lingua franca* spoken along the trade routes. It is perhaps possible to distinguish two layers in its word-stock, an older one and a more recent one, on the basis of the following phenomena:

- We already mentioned before that the *š* in some instances maintains its original sound, while in other instances it is turned into *ć* or *č*.

- Some Greek words were preserved by Uralic in a form where the upsilon is pronounced as *u,* while other words seem to have been affected by Greek itacism, thus upsilon turned into an *i*[61].

- The Uralic word-stock contains a number of borrowings from Akkadic, a language that was already dead in the seventh century of our era[62].

This implies the possibility that Xazaric is the continuator and/or offspring of an older *lingua franca* used *for centuries* along the roads of Eurasia for commercial purposes. In the older layer, represented by the supposed *lingua franca*, the languages spoken by the Jews (Akkadic, Aramic, Hebrew, Greek and Persian) played a very important role. In the newer layer, though, we find a number of borrowings from languages spoken in the Roman Empire and along the shores of Baltic. This is a sure mark of the fact that the Xazars extended their trade routes to the very heart of Europe.

The Xazarian Qaɣanate succeeded in gaining full control over trade in the Caspian and Black Sea regions. People and merchants from almost everywhere met in Xazaria. There were Persians, Muslims, Jews, Christian Greeks and merchants coming from the Caspian Sea or from the Black Sea with their spices, perfumes and oriental silkware. The first Rūs' carried their products in primitive boats to the Xazarian market – namely wax, honey, mead, sable and marten furs from the forest and amber from the Baltic sea. There were Norsemen, and peoples coming from the basins of Volga, Dnepr, Don and Kama, as well as Slavic tribal fur-hunters, red-haired giants that astonished the Persian merchants with their rough manners. In his letter addressed to Ḥasdai, King Joseph supplied details about the population living in the Qaɣanate:

> "*You have also asked me about the affairs of my country and the extent of my empire. You ought to know that I dwell by the banks of the river known as Itil*

---

[61] This might be due to the presence of different Greek dialects, though, where dialects having more archaic traits lived or survived together with dialects where the itacism had already spread.

[62] It is possible, however, that Babylonian Jewry was instrumental in bringing this about. In fact, the Babylonian Talmud, which goes back approx. to 350 of C.E., preserved a number of Akkadic words (cfr. Akkadic *istēnu* > Talmudic Hebrew אשתן *išten* 'one, single, unique, the only one' > Hungarian *Isten* 'God').

[=Volga]. *At the mouth of the river lies the* גורגן ים *Gorgan Sea* [= The Caspian, also called جرجان dʒurdʒan "Georgian" by the Arabic sources] *and it extends eastward, a journey of four-months distance. Many and innumerable peoples live along this river in villages, towns and cities. These are their names:* בורטס *Burtas,* בולגר *Bulgar,* סואר *Savar* [cfr. the former name of the Magyars according to Konstantinos Porphyrogennetos, i.e. Σάβαρτοι ἄσφαλοι], אריסו *Arisu,* צרמיס *Čeremis* [cfr. the U people named Cheremiss], וננתר *Venenter,* סוור *Suvar,* צלויון *Člaviun* [~ Ṣlaviun, possibly Slavonians?], *and they all pay me tribute. From there the border turns to Buarezm* [Xwārezm] *up to the Gurgan Sea and every inhabitant of the seaside pays me tribute. To the south:* סמנדר *Semender* [=Tarkhu], בק תדלו *Bak Thadlu, up to* באב אלאבואב *Bab-el-Abwab* [= the "door of the doors" of the Arabs, today: Derbent]. *From there the border turns toward the mountains* [= the Caucasus]: אזור *Azur,* בק בגדא *Baq Bagda,* סרירי *Seriri* [cfr. the Serirs of the Arabic chronicles], קיתון *Kithon,* ארקו *Arku,* שאולא *Šavala* [or: Šaula], סגסרט *Sagsart* [or: סנסרט Sanasert], אלבוסר *Albusar,* אוקוסר *Ukusur,* כיאדוסר *Khiadosar,* ציגלג *Čiglag,* זוניך *Zunikh* [or: אניך Anikh], *who live on very high mountains; then all the* אלניים *Alans up to the border of* אפכאן *Afkhan and* כאסא *Khasa* [or: באסא *Basa*], כלכיאל *Khalkhiel and* תכת *Thakhath* [or: תגת Thagath]. *Moreover, every inhabitant of the country, up to the sea of Constantinople* [=Black Sea], *to the extent of a two-months journey* [pays tribute]. *To the west:* שרכיל *Šarkhel* ["White Abode": Σάρκελ in the Greek chronicles, **Бѣла Вежа** "White Tower" in the Old Russian chronicles, and مدينة البيضا Medina-t-al-Baiḍa "White City" in the Arabic travelogues], סמקרץ *Samkerč,* קרץ *Kerč* [=the old Greek city of Pantikapaion, later called Xerson, Crimea], סוגדאי *Sugdai,* אלוס *Alus* [=Alušta], למבט *Lambat,* ברתנית *Barthenith* [=Partenit], אלוביכא *Alubikha* [=Alupka], כות *Khuth,* מאנקוף *Mankuf* [or: מאנקות Mankuth], בודק *Budak* [or: בורק Burak], אלמא *Alma,* גרוצין *Gručin* [=Gruzinov]. *Then the border turns to the north up to the great river, the name of which is* יוזק *Yuzak* [or: יוזג Yuzag] *and these* [N.B. the names of the peoples concerned are missing] *live in open places that are not fenced and they wander in the steppes up to the border of the* היגראים *Higra people. They are as many as the sand on the seaside and they all pay me tribute. The extension of their country is a journey of four months distance. Bear in mind that I dwell at the delta of the Itil and, by God's help, I guard the mouth of the river and do not permit the* רוסיים *Rūs' who come in ships to enter into the Caspian so as to get at the Moslems. Nor do I allow any of their* [= the Moslems'] *enemies who come by land to penetrate as far as Derbend. I have to wage war with them, for if I would give them any chance at all they would lay waste the whole land of the Moslems as far as Baghdad. You have also asked me about the place where I live. I wish to inform you that, by the grace of God, I dwell alongside this river on which are situated three capital cities. The queen dwells in one of them; it is my birthplace. It is quite large, built round as a circle, the diameter of which is fifty parasangs. Jews and Moslems live*

in it, but there are people speaking many different languages living there. Jews, Christians and Moslems live in the second city. Besides there are many slaves of all nations in it. It is of medium size, eight square parasangs in length and breadth. In the third I reside with my princes, officers, servants, cupbearers, and those who are close to me […]"[63]

As we said before, King Joseph did not answer straightforwardly Ḥasdai's question concerning the language he spoke. Here we find the reason why he did not. In his kingdom – and even within the same city – the inhabitants spoke *many different languages*. This is a most typical situation that brings about the need for a restricted language system, *i.e.* a *lingua franca* or a *pidgin*, in order to cater to essential common needs. King Joseph was simply unable to answer Ḥasdai's question: in the IX century the notion of *lingua franca* had not yet been invented and the concept of *mixed language* came about only in the fifties of our century.

The so-called "Uralic" languages probably developed along the part of the Silk Road controlled by the Xazar Qaɣanate, very likely where the Xazars had their trading posts. The idea of a trading post as the core of linguistic diffusion is not so far-fetched as it might seem at first sight. Archaeological evidence of a Xazarian presence along the Silk Road has been found in the Talas Valley, Kazakhstan, and at Birka in Sweden.

A tally stick from the Talas Valley, Kazakhstan. The indeciphered Turkic-type runic script is thought to be Xazarian.
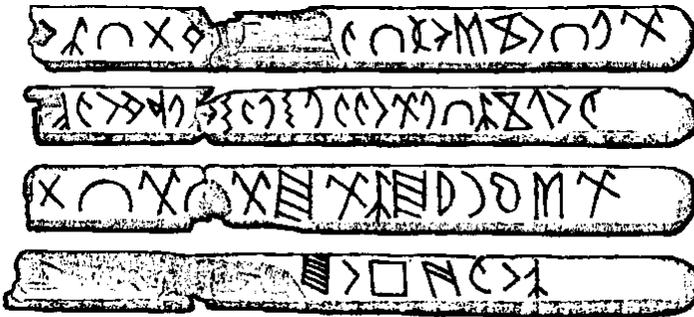
One of the most important mixed Xazaro-Swedish trade settlements in Northern Europe was possibly Birka, in the Swedish Mālaren, the very Byrca mentioned in the mediaeval European chronicles. It was founded, along with other similar trading places in eastern and northern Europe, during the late VIII century. Archaeological findings support the thesis of a Xazarian presence in the site. Another Xazarian trade

---

[63] My translation of King Joseph's letter (long version) is based on the work of S. Kohn (1881, 24 ff.)

place was the present Staraja Ladoga, called *Aldeigjuburg* or *Ostroburg* in the chronicles, situated at the eastern connexion of the eastern river systems to the Baltic. Another such place was present-day Hedeby, called *Haithabu* (but also *Tängrilbyr*[64] in the Nordic sagas), and located in the southwestern corner of the Baltic. Birka was for a very long period one of the most prosperous trade posts in Northern Europe (Mats, 1997). The place-name has been often explained by popular etymologies based on sound-alike words, e.g. *birk* 'birch', *bjarn* 'bear', *bjur* 'beaver' and *bjark* 'hill'. Nevertheless, none of these etymologies stands serious criticism.

Coins found in Burka, Sweden

In this connection we shall note that in the Arabic-speaking countries (or where Arabic had a strong impact) بركة *Birka* is a very common place-name (cfr. Wadi el-Birka, Egypt; Birka, Egypt; Birka, Afghanistan, etc.) The Persians also borrowed this word in the VII century. It means 'shallow waters', 'place where the water stands still'. In present days it is mainly used in the meaning 'pond, pool'. It is possible that Arabic, Persian *and* Xazarian merchants might have used this word to name a port of trade in a secluded and secure bay where waters were shallow. We can find the same word in Hungarian, cfr. *berëk* 'fen, moor, swamp, marsh, marshy posture; grove'. Many Hungarian compound place-names like *Beregszó, Beregszeg* etc. are always in strict relationship with shallow waters. The word perhaps survived also in Votyak J. *ber-gop* 'moor puddle' (*gop* = 'gorge, ravine, canyon, mine') and Votyak S. *pera* 'black mud'.

The presence of Xazars in Northern Europe is possibly connected with a group of people professing a "Mosaic creed", the so-called Χαλίσιοι, *Xalisioi*. This ethnonym appears in the Syriac sources as *ḥwalis,* in the old Russian name of the Caspian Sea – **Хвалисское море** [Xvalisskoe more] [65], in Polish as *Kalisz* (cfr. the place-name

---

[64] The word seems to be composed by the Ťuvash-Turkic word *täŋir* 'sky, heaven, god' – yet cfr. Hungarian *tenger* 'sea' – and the Turkic word for "commercial post". Its most likely meaning is "sea trade post".

[65] The coeval sources call "Sea of the Xazars" the Caspian.

Kalisz)*,* in Hungarian as *Káliz* (cfr. the ancient Hungarian place-name *Kaluzdij* "place of the Calisians" going back to 1082)*,* in Greek as Χαλίσιοι*.* They are thought to be originally a Persian or a Turkic[66] group. Some of the Calisians settled in Pannonia possibly together with the Magyars. Another group of Calisians, that lived in Danubian Bulgaria, joined the Hungarians during the rule of Géza II in 1152. In the XII and XIII century, the Calisians played a central role in the financial affairs of Hungary as customs officers and tax-collectors. They minted the coins of the Árpád dynasty with Hebrew characters[67]. The same happened in Poland, where the coins of king Bolesław V Wstydliwy ("The Pious", 1221–1279) were minted by Jews. By his *Kalisz Edict* (from the name of the city of Kalisz, named after its inhabitants) he allowed the Jews to settle in Poland (or rather acknowledged a *de facto* situation). Hebrew coinage continued under king Kazimierz III Wielki ("The Great", 1310–1370). We know also the name of his minter, who was a Jew named Lewko. In the XII century, the Byzantine historian Ioannes Kinnamos (Ἰωάννης Κίνναμος) in his Epitome[68] mentioned the Χαλίσιοι as follows:

[...] οὐχ ὅσον μόνον ἐγχώριον, ἀλλά καί μύριόν τι φῦλον συμμαχικόν ἔκ τε Οὕννων ἱππέων καί δέ καί τῶν παρ' αὐτοῖς ἑτεροδοξούντων Χαλισίων. Οὕννων γάρ τά Χριστιανῶν πρεσβευόντων, οὕτοι Μωσαϊκοῖς καί τούτοις οὐ πάντη ἀκραιφνέσιν εἰσέτι καί νῦν διεξάγονται νόμοις.

"*...not only the inhabitants of the country, but also a part of the Hun cavalry and some of the Calisians that live among them, but are of another religion. While the Huns profess the Christian faith, up to present these* [=the Calisians] *use the laws of Moses, but not in their pure form*" (Cap. 7)

[...] Ὡς δέ βασιλεύς [=Στέφανος] ἐκεῖθεν μεταβάς ἕτερόν τι φρούριον ἐνεούργει, ἐν ᾧ πολλούς τῶν ἐν Σιρμίῳ φέρων ᾠκίσατο Οὕννων, οὕς παρ' αὐτοῖς Χαλισίους ἔ ος καλεῖν ἐστι (καί εἰσιν ἑτερόδοξοι, καθάπερ ἤδη ἔφην, Πέρσαις ταυτοφρονοῦντες) [...]

"*...The king* [=István] *when the emperor departed to besiege another city where he settled many Huns of Sirmion, that are called Calisians by them and have, as I said before, another creed since they profess the same religion of the Persians*[69]*...*" (Cap. 25)

We do not know what Kinnamos meant with these words. Possibly the Calizians practiced a sort of syncretic Judaism that reconciliated diverse religious practices. Many of them became convert to rabbinical Judaism, as the frequent Jewish family name *Kalisch, Kaliz* and *Kalisz* seems to show[70].

---

[66] M. Gyóni, Magyar Nyelv 34 (1938: 86–96; 159–168).

[67] Scheiber, Sándor: *Ősi érmek.* Beszéd. Karcag, 1948 (részlet belőle: Kunsági Hírlap. II.1948. 17. sz.); *A héber betűjeles Árpád-házi pénzekhez.* Numizmatikai Közlöny LXXII–LXXIII. 1973/74, 91.

[68] Ioannis Cinnami *Epitome rerum ab Ioanne et Alexio Comnenis gestarum,* rec. A. Meineke. Bonnae, 1836.

[69] According to Gy. Moravcsik, the "religion of the Persians" is Islam.

[70] The family name *Kalisch* or *Kalisz* might be also related to the Polish place-name *Kalisz.*

Birka and Hedeby represented the starting point from which the Vikings became aware of the economical importance of trade routes. As a matter of fact, the expansion of the Vikings in Eastern Europe was due to economical factors. At first they traded with the Slavs and, much alike the Xazars, they established their trade posts, named *garðar,* on sensible spots of river navigation – or, much likely, they used the Xazarian trade posts. Essentially, these were fenced and fortified towns that were connected one to another via the enormous river system that branched out up to the Caspian Sea. At first, the Norse *felagaR* 'companions; partners' kept trade relationships with *Garðarikr* 'the country of fenced places' (i.e. trading posts) – as they called Eastern Europe. From Scandinavia to the city of Bulghar, situated on the great bend of Volga, to the mysterious country of *Wisu,* located at a distance of three months journey north of Bulghar, the Vikings traded furs, silk, cotton and slaves, but also wheat, fish, wood, hides, salt, wine, glue, horses, honey, wax, wool, amber and silver, etc. Although lacking any information on the Xazarian trading routes, it is possible to assume that the Scandinavian merchants very likely followed the traditional routes and water ways of the Xazar, Persian, Arab and Jewish merchants. This fact enables us to get a general idea of the possible location of the most important trade posts.

Norsemen were quite numerous in Xazaria. Abu'l Qāsim 'Ubayd Allāh bin 'Abd Allāh Ibn Khurdādbeh (fl. 840–890), the caliph's chief of intelligence (*murīd ab-barīd* 'postmaster general') initiated the genre of the Arab descriptive geography. In his classical work *Kitāb al-masālik wa'l-mamālik* (Book of the routes and kingdoms) he has a chapter dealing with the international *negociatores* trading companies. The first of these consisted of Jewish merchants (*al-tuğğar al-yahūd ar-Rādhāniyya*), and the others of Rūs merchants (*tuğğar ar-Rūs*), who were a kind of *Ṣaqaliba*[71]. They had a great assortment of merchandise, but above all the "eunuchs (*al-khadam*), female slaves (*al-ğawārī*) and boys (*al-ghilmān* or *aṣ-ṣabī*)"[72]. Another coeval writer, Abu'l-Ḥasan 'Ali bin al-Ḥusayn al-Mas'ūdi (d. 956) was certainly the most prolific Arab polymath and traveller. In his work *Murūğ adh-dhahab* he names only one company of international *negociatores,* namely those of the Rūs'. He writes: *"The Rūs' are a colluvies gentium* (أمم – umam, pl. of *umma* 'people, nation, generation') *of diverse kinds* (ذات أنواع شتى – dhāt anwāʿ šattā). *Among them there is a kind called al-Lo(r)domāna* [cfr. Spanish Latin *Lordoman < Nordoman*]. *They are the most numerous. They frequent with their wares the country al-Andalūs* [=Spain], *Rūm* [=Rome, the Roman Empire], *Constantinople and that of the Xazars"* (Pritsak, 1970). The French mediaevalist Georges Duby writes: "*All that can be said is that eighth- and ninth-century sources, when referring to negociatores* (=‘traders, merchants’), *frequently allude to two ethnic groups, whose colonies were dispersed along the main routes and streached far beyond the frontiers of the Empire: Jews and 'Frisians'*"[73]. As we have seen above, the ninth-cen-

---

[71] *Ṣaqalab,* pl. *Ṣaqāliba* is the Arabic name of the Slavs.

[72] Ibn Khordādbeh, *Kitāb masālik wa'l-mamālik.* Ed. M. J. De Goeje. Leiden, 1889, pp. 153–155.

[73] Duby, Georges: *The early growth,* p. 101.

tury author Ibn Khurdādhbeh made a very similar statement; however, he used the name "Rūs" instead of "Frisians". One has the right to assume that this is the way a number of eastern Germanic words entered the word-stock of Xazaric and hence Samoyedic.

In 884, economic factors enticed the Rūs' to get rid of the tribute laid on them by the Xazars. Oleg (Helge), the Rūs' prince of Kiev, passed through the Slav tribes of the Dniepr basin with the cry "Pay nothing to the Xazars". In 965–969 Sviatoslav, prince of Kiev, conquered the Xazarian cities of Itil, Šarkhel and Semender. It is quite interesting to note that the presence of Kufic coins in Scandinavia, particularly significant in eastern Sweden and in the Island of Gotland, came to an end just before 970 (Jones, 1977, 14), thus the fall of the Xazarian Qaɣanate decreed the decline of the Xazarian trade posts in Scandinavia.

The so-called "Uralic" peoples arose where the Xazarians had their trade posts, with the only exception of the Hungarians, who migrated from the interior of Xazaria. However, following a chronicle the original of which is lost, Mahmud Terdžüman in 1543 wrote the *Tarihi Üngürüsz* [History of Hungary]. In this chronicle, when describing the Hungarian conquest of Pannonia, he wrote: "*When they arrived in that region they saw the copiousness of rivers, abundance of fruits and profusion of crops, and they spoke the same language*" (Blaskovics, 1996). In the one-time Hungarian town of Dunacséb next to the Danube (today: Čelarevo, near Novi Sad in Vojvodina, Serbia) the archaeologists unearthed 450 burials containing elements of the Jewish creed, namely symbols of the *menorah*[74]*, ethrog*[75]*,* and *shofar*[76] (Scheiber, 1983). On some of the bricks there Hebrew words were found, such as *Jehuda, Jerushalaim* and *Israel.* This would not be surprising: in Pannonia, since Roman times, there had been trading centres and in all of them ancient Jewish tomb stones were found (Scheiber, 1976). Yet, the bones found in the tombs of Dunacséb-Čelarevo belong to the *Mongolic phenotype.* The burials date back as from the late VIII century to the XI century. Before the find, no Jewish presence was known in that area.

Even today, the speakers of the U languages are scattered along rivers and seashores and the map of their settlements overlaps very closely the map of the ancient trade routes. The so-called "Uralic" languages developed out of the original pidgin spoken by the Xazarian merchants. For example, the Sami languages are spoken around a place named Birka, exactly as the one on Lake Mälaren, but located 400 km north of it on Lake Storsjön. Finnish and Estonian developed on the Baltic Sea. Ingrian and Votian are spoken in the Luzhskaya Bay, from where ships sailed to Birka and Hedeby. Finnish developed on the Baltic coast around the city of Turku (< Old Finnish *turku* < Old Russian тържъ 'market'). Livonian is spoken

---

[74] *Menorah*: a holy candelabrum having seven branches used in the ancient temple of Jerusalem.

[75] *Ethrog*: the bitter fruit used for the celebration of Passover.

[76] *Shofar*: an ancient Hebrew musical instrument usually made of a curved ram's horn, still used in Jewish religious services.

in a few towns on the Latvian coast situated almost opposite the Island of Gotland. The Vogulians live on the left side, the Ostyaks on the right side of the Ob river. The other peoples belonging to this language family are scattered along the main rivers of the great waterway net that represented the main trade route of Late Antiquity and the Early Middle Ages. Unfortunately, the trade routes of north-east Siberia are still unknown.

The pidgin evolved into a set of creoles that were influenced to a greater or lesser extent by the linguistic substrate of one or more languages spoken by the natives. The importance of the substrate is clearly shown by the single "Uralic" languages. Some of the languages concerned, such as the Balto-Finnic languages, were heavily influenced by Balto-Slavic and Germanic. Other languages, as for example Votyak, show evidence of an Iranic substrate. Permian languages were influenced by Slavic, while the "Uralic" languages spoken in Siberia depended heavily on Turkic languages for the development of their vocabulary and syntax. There is even the example of a so-called Paleo-Siberian language, viz.Yukagir, that seems to have been influenced by the Xazarian word-stock.

The similarities that some of the newborn creole languages show, as in the case of the Finnic languages, are due to the very fact that the contacts among speakers of different creoles persisted along the trade routes controlled by the Norsemen up to the late XIII century. At the same time, the Samoyedic languages – which were cut off from any contact with the "mother country" after the fall of the Xazarian Qaγanate – were left free to develop independently from the other languages of this group. This, and not the great antiquity of the branching off, is the real reason for many of the differences but also of the convergences existing between Samoyedic and the other languages of this group. We are using the word "group" because the "Uralic" languages cannot be defined a "family" in the traditional, linguistic sense of the term, since the relationship that ties them together is not the traditional genetic relationship, but rather a generic "loan"-relationship for which no name has yet been coined.

The evolution of some of these languages, like Sami and Samoyedic, which are divided into a subset of languages in their own right, might have followed closely the development hypothesized by Chaudenson (1992) for "plantation creoles" (*créoles de plantation*). During the first decennia, the trading posts were not very important. They might have been made up by small villages where the Xazarian traders were often more numerous than the natives. Under such conditions, the natives spoke an approximate variety of the Xazarian pidgin. As the importance of such trade posts grew greater and a growing number of natives settled in the neighbourhood of the trade posts, the newcomers begun to learn the pidgin not directly from the Xazarian traders but rather from each other. As a consequence of this, the approximation of the first natives who learned the Xazarian language became the principal model for subsequent approximation, that is to say that, by developing approximations of the former approximation, the newcomers in their turn supplied the model for the natives that followed. Therefore, some of the so-called Uralic languages could be the final result of a set of "approximations of approximations".

Final considerations

The present study is a pioneer work and, as such, it is likely to contain some inaccuracies. The problems it raises, however, are more than those that it solves.

One problem in postulating the genesis of Proto-Uralic as late as the VII–VIII century of our era is that this would give the daughter languages just a few hundred years to develop their complex system of cases and other elaborate morphological details. Of the languages that we know to be creoles, some have existed about three hundred years, the oldest (São Tomense) almost five hundred years. There is still not a single case among these languages and very little evidence of the emergence of other morphology. However, we must take into account the ambiental factors which influence the growth of a creole. The first and foremost factor is that most of the creoles we know developed in a linguistic milieu that was tied to a certain geographic reality and therefore is relatively poor in terms of outer stimuli. The Xazarian pidgin, hypothesized in the present work, developed along the trade routes and underwent a coacervation of stimuli originating from a number of different languages.

Usually, when a *modern* pidgin develops, the bulk of its vocabulary originates from one or two lexifier languages only. The language of the Xazars, though, took its vocabulary items from many different languages. It seems to have been multi-layered, thus composed of an older *lingua franca*-type layer (possibly a *lingua franca* formerly spoken along the trade routes) with a number of more recent additions. These layers should be analysed separately, and this will involve a more refined method for sieving out data.

We have been using the word "pidgin" throughout the present study, but it might not be the right term. We simply do not know whether the language spoken by the Xazars was a pidgin, a partially creolised pidgin or a real creole. The difference is substantial. If it was a creole or at least a partially creolised pidgin, we could admit the presence of a limited number of affixes and tied morphemes – that is a proper basic structure from which the daughter-creoles might have developed further. If it was a pidgin, free morphemes only can come into account. If the language spoken by the Xazars was a creole, though, it was possibly re-pidginised and re-creolised a number of times following the waves of different migrations.

As we see, the language of the Xazars poses a number of unsolved problems. The unveiling of the many secrets it hides will be the task of the future.

Acknowledgements

BIBLIOGRAPHY

AGOSTINI, Paolo (1993). "Per una nuova teoria sulla genesi delle lingue uraliche", in: *Rivista di Studi Ungheresi*. Vol. 8. Roma.

AGOSTINI, Paolo (1996). "Concordanze lessicali arabo-magiare. Arabo-Hungarian lexical matches", in: *Rivista di Studi Ungheresi*, Vol. 11. Roma.

ALLEN, W. S. (1953). "Relationship in comparative linguistics"*,* in: *Transactions of the Philological Society*. London.

ANDERSON. James M. (1973): *Structural Aspects of Language Change*. London: Longman.

ΆΝΔΡΙΩΤΗΣ, Ν. Π. (1983). Ετυμολογικό λεξικό της κοινίς Νεοελληνική. Θεσσαλονίκη: Τριτι εκλοση.

BAALBAKI, Rohi (1993). *Al-Mawrid* المورد . *A modern Arabic-English dictionary*. 5th ed. Beirut.

BENFEY, Theodore (1991). *A Sanskrit-English dictionary*. New Delhi–Madras: Asian Educational Services.

BERECZKI, Gábor (1984). "La misura e il carattere dell'influsso turco sulle lingue ugro-finniche", in: *Studi miscellanei uralici e altaici*. Vol. 20. A. Csillaghy (ed.). Venezia: Libreria editrice Cafoscarina.

BICKERTON, Derek (1981). *Roots of Language*. New York: Karoma Publishers.

BLASKOVICS, József (1996). *Mahmud Terdzsüman – Tarihi Üngürüsz*. Erdem. Léva.

BOULNOIS, Luce (1992). *La route de la soie.* Genève: Olizane.

BROOK, Kevin (1998). *The Khazaria Info-Center*. Online. Available: http://www.khazaria.com/

BRÜCKNER, Aleksander (1927). *Słownik etymologiczny języka polskiego*. Kraków.

CHAUDENSON, Robert (1992). *Des îles, des hommes, des langues. Essai sur la créolisation linguistique et culturelle.* Paris: L'Harmattan.

COLLINDER, Björn (1960). *Comparative Grammar of the Uralic Languages*. Stockholm.

COLLINDER, Björn (1977). *Fenno-Ugric vocabulary. An etymological dictionary of the Uralic languages*. Second revised edition. Hamburg.

COMRIE, B. (1981). *Language Universals and Linguistic Typology*. Chicago.

CORENCHY, Éva (1972). *Iranische Lehnwörter in den obugrischen Sprachen*. Budapest: Akadémiai Kiadó.

CRYSTAL, D. (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.

DÉCSY, Gyula (1965). *Einführung in die finnisch-ugrische Sprachwissenschaft.* Wiesbaden: Harrassowitz.

DÉCSY, Gyula (1990). *The Uralic proto-language: a comprehensive reconstruction*. Bloomington, Indiana.

DUNLOP, D.M. (1967). *The History of the Jewish Khazars*. New York: Schocken Books.

FERGUSON, Charles (1971). "Absence of Copula and the Notion of Simplicity: A Study of Normal Speech, Baby Talk, Foreigner Talk and Pidgins", in: Hymes (ed.), Cambridge: Cambridge University Press, pp. 141–150.

FERNANDEZ BELL, María Rosa and Gilbert GLENN (1998). *Glossary of Terms related to the Study of Pidgin and Creole Languages*. John Benjamins Publishing Company. Online. Available: http://www.siu.edu/departments/cola/ling/glsintro.htm

ФЛЁРОВА, В. Е. (1997). *Граффити Хазарии. Граффити Хазарии.* Москва: Эди-ториал УРСС.

FOLEY, William A. (1988). "Language birth: the process of pidginization and creolization", in: Frederick Newmeyer (ed.). *Linguistics: The Cambridge Survey 4: Language: The Socio-cultural Context*. Cambridge: Cambrige University Press, pp 162–183.

FUF = *Finnisch-ugrische Forschungen*. I-13, Helsingfors–Leipzig, 1901–1913; 14–, Helsingfors (later:) Helsinki, 1914–.

GOLB, Norman and Omeljan PRITSAK (1982). *Xazarian Hebrew Documents of the Tenth Century*. Ithaca – London: Cornell University Press.

ГОЛЬ, Норман; Омельян ПРИЦАК (1997). *Хазарско-еврейские документы X века*. Научная редакция, послесловие и комментарии В. Я. Петрухина. Гешарим. Москва (1997), Иерусалим (5757).

GOLDEN, Peter (1980). *An Historico-Philological Inquiry into the Origins of the Xazars*. Budapest: Akadémiai Kiadó.

GOMBOCZ, Zoltán (1960). *Honfoglalás előtti bolgár-török jövevényszavaink.* Nyelvtudományi Értekezések. 24. sz.

GREENBERG, J. H. (1957). *Essays in Linguistics.* Chicago.

GREENBERG, J. H. (1993). *Indo-European and its closest relatives: the Eurasiatic Family*. Stanford University Press.

HAJDÚ, Péter & Péter DOMOKOS (1978). *Uráli nyelvrokonaink*. Budapest: Tankönyvkiadó.

HAJDÚ, Péter & Péter DOMOKOS (1987). *Die uralischen Sprachen und Literaturen*. Budapest: Akadémiai Kiadó.

HAJDÚ, Péter & Danilo GHENO (1992). *Introduzione alle lingue uraliche*. Torino: Rosenberg and Seller.

HAJDÚ, Péter (1968). *Chrestomathia Samoiedica*. Budapest: Tankönyvkiadó.

HAJDÚ, Péter (1988). *Bevezetés az uráli nyelvtudományba*. 4. kiadás. Budapest: Tankönyvkiadó.

HÄKKINEN, K. (1990). *Mistä sanat tulevat.* Helsinki: Suomalaisen Kirjallisuuden Seura.

HALÁSZ, Ignác (1893–4). "Az ugor-szamojéd nyelvrokonság kérdése", in: NyK. 23–24. Budapest.

HARMATTA, János (1977). "Irániak és finnugorok, irániak és magyarok", in: A. Bartha et alii (eds) *Magyar őstörténeti tanulmányok.* Budapest: Akadémiai Kiadó.

HEISE, J. and H.A.I. STADHOUDERS (1997). *Akkadisch vocabularium*. Online. Available: http://saturn.sron.ruu.nl/~jheise/akkadian/

HOLLADAY, William L. (1988). *A Concise Hebrew and Aramaic Lexicon of the Old Testament*. Based on the First, Second and Third Editions of the Koehler-Baumgartner Lexicon in Veteris Testamentis Libros. Leiden: E. J. Brill.

HONTI, László (1993). *Die Grundzahlwörter der uralischen Sprachen*. Budapest: Akadémiai Kiadó.

IEW = Julius POKORNY (1959). *Indo-germanisches etymologisches Wörterbuch*. Bern–München: Francke Verlag.

ITKONEN, Erkki (1946). "Zur Frage nach der Entwicklung des Vokalismus der ersten Silbe in den finnisch-ugrischen Sprachen, insbesondere im Mordwinischen", in: FUF, 1946: 29.

JANHUNEN, Juha (1981). "On the structure of Proto-Uralic", in: FUF. vol. 44, pp. 23–42.

JOKI, Aulis J. (1988). "Zur Geschichte der uralischen Sprachgemeinschaft unter besonderer Berücksichtigung des ostseefinnischen", in: D. Sinor (ed.). *The Uralic Languages: description, history and foreign influences*. Leiden: E. J. Brill. .

JONES, Gwynn (1973). *I Vikinghi.* Newton. 1995 (Italian translation of: *A History of the Vikings.* Oxford University Press).

JOÜON, Paul and T. MURAOKA (1993). *A Grammar of Biblical Hebrew*. Roma: Pontificio Istituto Biblico.

JSFOu = *Journal de la Société Finno-Ougrienne. Suomalais-ugrilaisen Seuran Aikakauskirja*. I.–. Helsinki, 1886 –.

KATIČIĆ, Radoslav (1970). *A Contribution to the General Theory of Comparative Linguistics*. The Hague–Paris: Mouton.

KLEIN, Ernest (1987). *A comprehensive etymological dictionary of the Hebrew language*. מילון אטימולוגי עברי-אנגלי מקיף. Jerusalem: Carta.

KOHN, Sámuel (1881). *Héber kutforrások és adatok Magyarország történetéhez*. Budapest.

KORENCHY, Éva (1988). "Iranischer Einfluss in den finnisch-ugrischen Sprachen", in: D. Sinor (ed.). *The Uralic Languages: description, history and foreign influences*. Leiden: E. J. Brill.

KÜNNAP, Ago (1997). *The Uralic Languages*. Online. Available: http://www.ut.ee/Ural/TYE.html.

LAKÓ, György (1974). *A magyar nyelv finnugor alapjai*. Budapest: Tankönyvkiadó.

LAMBERT, M. (1972). *Traité de la grammaire hébraïque*. (Paris, 1931–38). Reprint: Hildesheim, 1972.

LÁSZLÓ, Gyula (1987). *Őstörténetünk.* Egy régész gondolatai néppé válásunkról. 3. kiadás. Budapest: Tankönyvkiadó.

LOKOTSCH, Karl (1927). *Wörterbuch der europäischen (germanischen, romanischen und slawischen) Wörter orientalischen Ursprungs*. Heidelberg.

MAKKAY, János (1991). "Az uráli-finnugor őstörténet néhány kérdése az indoeurópai őstörténet szemszögéből", in: Századok. *A magyar Történelmi Társulat folyóirata*. Budapest, 1991, 1–2, pp. 3–34. .

MARCUS, Jacob (1938). *The Jew in the Medieval World: A Sourcebook. 315–1791*. New York: JPS.

MEID, W. (1989). *Archäologie und Sprachwissenschaft. Kritisches zu neueren Hypothesen der Ausbreitung der Indogermanen*. Innsbruck.

MENGES, Karl H. (1968). *The Turkic languages and peoples*. Wiesbaden: Harrassowitz.

MORAVCSIK, Gyula (ed.) (1988). *Az Árpád-kori magyar történet bizánci forrásai. Fontes byzantini historiae hungaricae aevo ducum et regum ex stirpe Árpád descendentium*. Budapest: Akadémiai Kiadó.

MszFgrE = György LAKÓ (1972–78). *A magyar szókészlet finnugor elemei. Etimológiai szótár*. Voll. I–III. Budapest: Akadémiai Kiadó.

MÜHLHÄUSLER, P. (1978). "Samoan Plantation Pidgin English and the Origin of New Guinea Pidgin", in: *Papers in Pidgin and Creole Linguistics*, I. pp. 67–119.

MÜHLHÄUSLER, Peter (1997): *Pidgin and Creole Linguistics*. Expanded and revised edition. London: University of Westminster Press.

NyK = *Nyelvtudományi Közlemények*. I–. Pest (later:) Budapest, 1862–.

PAASONEN, Heikki (1917). "Beiträge zur finnischugrisch-samojedischen Lautgeschichte", in: KSz 13–17. Budapest.

PEDERSEN, H. (1910). *Linguistic Science in the Nineteenth Century*. Cambridge, Mass.

PHILIP, Mats (1997). *Birka at the Silkroad*. Online. Available: http://home1.swipnet.se /~w-14723/birka/

[POLAK, Av. N.] (1949). = תולדות ממלכה יודית באירופה. הוצאת מוסד ביאליק ע״י «מסדה». תל־אביב. אב. נ. פולק: כזריה.

PRITSAK, Omeljan (1970). "An Arabic Text on the Trade Route of the Corporation of ar-Rūs in the Second Half of the Ninth Century", in: *Folia Orientalia*, 12, 1970, pp. 244–259.

RAUN, Alo (1988). "Proto-Uralic: comparative-historical morpho-syntax", in: D. Sinor (ed.). *The Uralic Languages: description, history and foreign influences*. Leiden: E. J. Brill.

ROMAINE, S. (1988). *Pidgin and Creole Languages*. New York: Longman.

RÓNA-TAS, András (1981). "A kazár népnévről", in: NyK 1981/2, pp. 349–380.

RÓNA-TAS, András (1985). "Újabb adatok a kazár népnév történetéhez", in: NyK. 1985/1, pp. 126–133.

SALMINEN, Tapani (1980). "Uralilaisten kielten käsikirja vihdoin ilmestynyt", in: *Virittäjä* 94, pp. 493–502.

SALMINEN, Tapani (1997). "Facts and myths about Uralic studies: a review article of Jazyki mira: Ural'skie jazyki". Online. Available: http://helsinki.fi/~tasalmin.

SAMMALLAHTI, P. (1977). "Suomalaisen esihistorian kysymyksiä", in: *Virittäjä*, 1977/2.

SAMMALLAHTI, P. (1979). "Über die Laut- und Morphemstruktur der uralischen Grundsprache", in: FUF. Vol. 43, pp. 22–66.

SCHEIBER, Sándor (1976). *Pinkas Hakehillot*. Jerusalem.

SCHEIBER, Sándor (1983). *Menoroth from Čelarevo*. Beograd.

SEGAL, M. H. (1927). *A Grammar of Mishnaic Hebrew*. Oxford: Clarendon Press.

SETÄLÄ, E. N. (1899). *Yhteisuomalainen äänehistoria*. Vol. I–II. Helsinki.

SINOR, Denis (1988). "The problem of Ural-Altaic relationship", in: D. Sinor (ed.). *The Uralic Languages: description, history and foreign influences*. Leiden: E. J. Brill.

SINOR, Denis (ed.) (1988). *The Uralic Languages: description, history and foreign influences*. Handbuch der Orientalistik and Handbook of Uralic Studies 1. Leiden: E. J. Brill.

SKES = Toivonen, Y. H. et alii (1955–1978). *Suomen kielen etymologinen sanakirja*. Lexica Societatis Fenno-Ugricae XII. 1–6. Helsinki: Suomalais-Ugrilainen Seura.

STASZEWSKI, Józef (1968). *Maly slownik. Pochodzenie i znaczenie nazw geograficznych*. Warszawa.

STEINITZ, Wolfgang (1944). *Geschichte des finnisch-ugrischen Vokalismus*. Acta Instituti Hungarici Universitatis Holmiensis. Series B. Linguistica 2. Stockholm.

STEINITZ, Wolfgang (1952). *Geschichte der finnisch-ugrischen Konsonantismus*. Stockholm.

SZINNYEI, József (1922). *Finnisch-ugrische Sprachwissenschaft*. 2. Aufl. Berlin-Leipzig.

TAULI, Walter (1966). *Structural tendences in Uralic languages*. The Hague: Mouton.

TESz (1967–1976) = Loránd BENKŐ (ed.-in-chief). *A magyar nyelv történeti-etimológiai szótára*. Budapest: Akadémiai Kiadó.

THOMASON, Sarah Grey and Terrence KAUFMAN (1988). *Language contact, creolization and genetic relationship*. University of California Press.

TörtNyt. (1991) = Loránd BENKŐ (ed.-in-chief). *A magyar nyelv történeti nyelvtana*. Budapest: Akadémiai kiadó.

UAJb = *Ural-Altaische Jahrbücher*. 24–, Wiesbaden 1952–.

UESSON, A. (1970). *On linguistic affinity: the Indo-Uralic problem*. Malmoe.

UEW = RÉDEI, Károly (1988). *Uralisches Etymologisches Wörterbuch*. Vol. I–II. Budapest: Akadémiai Kiadó.

UNTERMANN, J. (1985). "Ursprache und historische Realität", in: *Studien zur Ethnogenese*. Abh. der Rheinisch-Westfälischen Akademie der Wissenschaften. Vol. 72 (1985). Opladen.

VASMER, M. (1950–8). *Russisches etymologisches Wörterbuch*. Heidelberg. Reprint: Фасмер, Макс: *Этимологический словарь русского языка*. Санкт Петербург: Издательство Азбука, 1996.

VINCENZINI, Maurizio (1996). *Storia della moneta*. Roma: Newton & Compton.

von SODEN, Wolfram (1995). *Grundriss der Akkadischen Grammatik*. Roma: Pontificio Istituto Biblico.

WEINREICH, Uriel. "On the compatibility of genetic relationship and convergent development", in: *Word*, vol. 14, 374–379.

WICKMAN, Bo (1988). "The history of Uralic linguistics", in: D. Sinor (ed.). *The Uralic Languages: description, history and foreign influences*. Leiden: E. J. Brill.

WINTERS, Werner (1970). "Basic Principles of the Comparative Method", in: *Method and Theory in Linguistics*. Ed.: Paul L. Garvin. Janua Linguarum series major 40. The Hague.

Paolo Agostini

JEZIČNA REKONSTRUKCIJA – PRIMJER URALSKIH JEZIKA

SAŽETAK

Pošto je upozorio na poteškoće i metodološke slabosti u općoj teoriji jezične rekonstrukcije, autor iznosi svoje kritike o drevnosti uralskih jezika. Protouralski, kako su ga znanstvenici rekonstruirali, navodno ima skup obilježja nekoliko različitih jezičnih porodica. U članku se razmatraju primjeri leksičkih podudarnosti s povijesno utvrđenim jezicima po čemu autor zaključuje da je protouralski rječnik nastao pozajmljivanjem iz vrlo raznolikih jezika: baltoslavenskih govora, starošvedskoga, više turskih narječja, mongolskoga, tunguškoga, aramejskoga, hebrejskoga, arapskoga, srednjoperzijskih narječja, latinskoga i grčkoga. I drugi se jezici moraju uzeti u obzir, primjerice kineski, kavkaški jezici, ali i danas nepoznati jezici. Uvažavajući nekoliko fonoloških pravila može se lako identificirati velik broj osnovnih oblika u uralskom rječniku. Moguće je pokazati da su jezična obilježja pojedinih uralskih grana potekla od pidžina koji se rabio duž trgovinskih putova između Puta svile i istočnoevropskih trgovinskih pravaca. Dobro je poznata činjenica da ljudi različitih jezika, kad se prvi put susreću, za bitne zajedničke potrebe stvaraju nov reduciran jezični sustav (*lingua franca* ili pidžin). Zato pidžini često nastaju uz trgovinske putove. Uzevši u obzir značajke izvornoga rječnika, pa i izvješće bizantskoga cara Konstantina Porfirogeneta da su Mađari naučili svoj jezik od Hazara, Hazarski kaganat može se identificirati kao mjesto podrijetla uralskoga pidžina. Hazarski je kaganat uspio uspostaviti vlast nad trgovinom u području Kaspije i Crnoga mora tijekom tri stotine godina – od 650. do 950. n.e. Poznato je da su Hazari uspostavili trgovačke postaje sve od doline Talasa u Kazahstanu do unutrašnjosti Švedske. Hazaria je u ranom srednjem vijeku označavala jedan od glavnih trgovinskih putova, i tu je najvjerojatnije nastao novi jezik. Uralski su jezici najvjerojatnije potekli od drevnoga hazarskog pidžina koji se razvio oko hazarskih trgovačkih postaja i uz hazarske trgovinske putove.

KLJUČNE RIJEČI: jezična rekonstrukcija, usporedbena lingvistika, uralski jezici, ugrofinski jezici, pidžinski jezici, kreolski jezici, Hazari, trgovina u srednjem vijeku, Put svile

Paolo Agostini

AZ URÁLI NYELVEKRE ALKALMAZOTT NYELVÉSZETI REKONSTRUKCIÓ

ÖSSZEFOGLALÁS

A szerző, miután kimutatta az általános nyelvészeti rekonstrució teóriájának hiányosságait és metodológiai gyengéit, vitatja az ősuráli nyelv állítólagos ősiségét. A tudósok által rekonstruált uráli nyelv sok jellemző vonással rendelkezik, amelyek más nyelvcsaládokból való átvételeknek tűnnek. A tanulmány feldolgoz egy sor szókincsbeli konkordánciát történelmileg attesztált nyelvekkel és arra a következtetésre jut, hogy az uráli szókincs nem áll másból, mint kölcsönszavakból, amelyek a legkülönbözőbb nyelvekből valóak. A kérdéses nyelvek néhanya a balto-szláv, az ósvéd, több török nyelvjárás, a mongol, a tunguz, az arámi, a héber, az arab, több kései középperzsa nyelvjárás, bizánckori görög és latin, ám sok más nyelv is számításba jöhet, mint pl.a kínai, a kaukázusi nyelvek valamint olyan nyelvek, amely azóta már nyomtalanul eltűntek. Az uráli nyelv szókincsének számottevő része könnyedén felismerhető néhány fonológiai szabály segítségével. Az uráli leány-nyelvek jellemzői azt mutatják, hogy egy olyan *lingua franca* leszármazottjai, amely az észak- és kelet-Európát a Selyemúttal összekötő kereskedelmi útvonalakon jött létre. Ismert jelenség az, hogy néha, amikor különböző nyelveket beszélő emberi csoportok érintkeznek először egymással, egy új, csökkentett nyelvi rendszer jön létre ahhoz, hogy a közös, alapvető közlekedési szükségleteket ellássa. Ezek a csökkentett nyelvi rendszerek, amelyek általában a kereskedelmi útvonalakon jönnek létre, a *lingua-frankák* vagy *pidginek.* Ha figyelembe vesszük az eredeti szókincs jellemző vonásait, valamint azt, hogy Bíborbanszületett Konstantinos bizánci császár azt tanusította, hogy a magyarok a kazároktól tanulták a ma is beszélt nyelvüket, feltételezhető, hogy a kérdéses „uráli" pidgin a kazár kaganátusban jött létre. Háromszáz éven keresztül, az i.u. 650 és 950 között, a kazároknak sikerült kiterjeszteniök hatalmukat a Kászpi- valamint a Fekete-Tenger tartományain keresztül lebonyolodó kereskedelemre. A kazárok kereskedelmi állomásai a kazah Talaszvölgytől végig a Svédország szivében levő Birka városáig terjedtek. Kazária az egyik legfontosabb kereskedelmi útvonalon helyezkedett el a korai középkorban, így a legmegfelelőbb hely volt, ahol egy új pidgin nyelv kifejlődhetett. Az uráli nyelvek minden valoszínűség szerint a régi pidgin nyelv kreolizált utódai, amelyek a kazár kereskedelmi állomások körül és azok útvonalai mentén fejlődtek tovább.

KULCSSZAVAK: nyelvészeti rekonstrukció, összehasonlító nyelvtudomány, uráli nyelvek, finnugor nyelvek, pidgin nyelvek, kreol nyelvek, kazárok, középkori kereskedelem, Selyemút