REIN TAAGEPERA (Tartu—Irvine, California), *AGO KÜNNAP* (Tartu)

DISTANCES AMONG URALIC
AND OTHER NORTHERN EURASIAN LANGUAGES

Abstract. The present occurrence or non-occurrence of 46 structural features is analyzed in language groups ranging from Finnic to Eskimo-Aleut. Normalized measures of commonalities and distances between two languages are developed and used for graphical representations. The conventional Uralic and Altaic groupings emerge. So does a Paleo-Siberian concatenation (Yukaghir/Eskimo-Aleut/Chukotka-Kamchatkan/Yeniseian), more diffuse than even Uralic-Altaic jointly, but with some similarities to Samoyedic and Mordvin in the Uralic group. Within Uralic, four clusters emerge: Samoyedic, Mordvin/Saamic/Finnic, Mari/Permic/Ob-Ugric, and most distant from the rest, Hungarian. Surprisingly, far from supplying a bridge to Altaic, Hungarian also appears as the Uralic language the most remote from Turkic and other Altaic. Turkic and Tungusic present almost as many commonalities with Uralic languages (except Hungarian) as with each other. Common origins cannot account for all these crisscrossing patterns; a continuum of contacts, selective migrations and participation in *lingua franca* areas must come into play. Extension of the existing set of structural features is urged, to include the distinguishing features of Basque, Indo-European, Semitic-Hamitic and Dravidian, as well as further Altaic and Paleo-Siberian features.

Keywords: Uralic, Altaic and Paleo-Siberian languages, structural linguistic features, commonalities between languages.

1. Introduction

Languages in the northern belt of Eurasia are traditionally grouped as Indo-European, Uralic, Altaic and a loose residual collection of Paleo-Siberian languages that extend to Eskimo-Aleut in circumpolar North America and Greenland. Angela Marcantonio (2002) has recently questioned the validity of the Uralic grouping that joins Samoyedic and Finno-Ugric, arguing that Samoyedic and even Ob-Ugric may have as much in common with Northern Tungusic as with the rest of the Finno-Ugric languages. She also argues that Hungarian may be as close to Turkic as to the conventional Uralic grouping. Instead of clearly distinct groups, there may be a continuum of languages in contact, ranging from Finnic, Saamic, Mordvin, Mari, Permic, Hungarian, Ob-Ugric and Samoyedic to Altaic and Paleo-Siberian languages.

These challenging ideas resonated with a group of linguists, motivating them to gather a collection of 60 structural features and determine their essentially present occurrence or non-occurrence in languages ranging from Indo-European all the way to Eskimo-Aleut. Their results (Klesment, Künnap, Soosaar, Taagepera 2003) confirmed the existence of a distinct Uralic bloc but also gave some credence to the notion of a continuum. The emphasis of P. Klesment, A. Künnap, S.-E. Soosaar, R. Taagepera (2003) was on establishing and documenting a database for occurrence of structural features, while the resulting analysis remained cursory. The present study builds on this database but excludes the Indo-European languages and 14 of the 60 features, for reasons explained in Appendix, which lists the remaining 46 features.

Two related restrictions characterize the database established by P. Klesment, A. Künnap, S.-E. Soosaar, R. Taagepera (2003). It tries to avoid any preconceived notions about the origin of languages and, therefore, deals only with c o m m o n a l i t i e s p r e s e n t l y o b s e r v e d.[1] This endeavor makes it bypass lexical similarities, which inevitably lead to reconstruction toward hypothetical proto-languages in what risks becoming circular reasoning (cf. Östman, Raukko 1995). Some lexical similarities may result from sheer chance (see Ringe 1999; Marcantonio 2002: 136—153). Along with Klesment, Künnap, Soosaar, Taagepera 2003, we avoid placing trust in the existing etymological sets.[2]

The questionable nature of the existing etymological set of Uralic vocabulary is reinforced by more universal concerns. Historical linguistic statistics specialist Sheila Embleton observes with good reason that "Linguistically, words are perhaps easier to work with: 'universals' are most easily avoided; words tend to be longer than affixes or other functional elements (short forms are more prone to chance resemblances ...). [———] It is often assumed that morphosyntactic data should be considered primary when assessing genetic relationship, largely because it is assumed to be immune from borrowing." (Embleton 1986 : 165—166; but cf. e.g., McMahon, McMahon 2003). If some researchers feel comfortable with using lexical features, let them do so, but it should not be a demand imposed on others.

We do not exclude the possibility that morphosyntactic features might also be borrowed by one language from another. Alexandra Aikhenvald, in particular, has lately pointed out the extensive travel of morphosyntactic features among the Amazonian languages in South America (Aikhenvald 2003 : 1—2): "In the linguistic area of the Vaupés in northwest Amazonia, several different mechanisms help create new contact-induced morphology. [———] Language contact in the multilingual Vaupés linguistic area has resulted in the development of similar — though far from identical — grammatical structures. [———] Languages in contact [———] gradually become more like each other. Language contact may bring about gradual convergence resulting in structural isomorphism, whereby the grammar and semantics of one language are almost fully replicated in another ..."

---

[1] Given that people tend to live a good part of a century and that young people and old people may speak somewhat differently, a reasonable operational dividing line between synchronic and diachronic approaches might be to count as synchronic those features that have been used during the last 100 years.

[2] The present-oriented criterion differentiates this database from the set of 72 morphological features used by J. Greenberg (2000) for his "Eurasiatic" macrogroup.

Still, in the final account, it may be argued that vocabulary is liable to change faster and more thoroughly than language structure. Since our purpose is not to establish genetic commonalities but only to gauge the present similarities of languages, we are not disturbed by conceivable exchange of morphosyntactic features among languages.[3]

Some of the commonalities we observe may indeed hark back to a common ancestral language, but all of them cannot, given the crosscutting nature of patterns (see especially Thomason, Kaufman 1988). Some features may be due to contacts (see e.g., Dixon 1999; Künnap 2000; Renfrew 2000; Wiik 2002). Some commonalities occur between languages so far-flung geographically (e.g., Finnic and Eskimo-Aleut) that one would have to invoke either extremely ancient contacts or selective migrations (such as all-male hunting parties) or separate coincidental invention. After all, the number of phonemes available for case endings etc. is limited. A further possibility is participation in a common *lingua franca* (Taagepera 2000). We do not take a stand on the origin of commonalities. Accordingly, we talk of groupings and subgroups, rather than using loaded terms like "families" and "branches", which imply a family tree model.

The choice of languages included in this study is based on the following considerations. The possibility of diffusion of features makes us focus on a relatively compact geographical area, the northern tundra, forest and steppe belt in northern Europe and Asia. This includes the conventional Uralic and Altaic groups, the Paleo-Siberian languages that Uwe Seefloth (2000) has connected to Uralic (Yukaghir, Chukotka-Kamchatkan and Eskimo-Aleut), and Yeniseian.[4]

Within this range, our objective is two-fold. F i r s t, we develop a general method for visualizing the distances among categories (such as languages) on the basis of any set of features (such as structural traits) they could share. S e c o n d, we apply this method to an existing set of features of northern Eurasian languages, collected by P. Klesment, A. Künnap, S.-E. Soosaar, R. Taagepera (2003).

---

[3] In which way are single similar phonemes more significant in morphosyntactic features than in lexical ones? Compared to a single phoneme functional ending, a multi-phoneme word root has a higher probability that one of the phonemes happens to be the same by random coincidence. This possibility is enhanced when reconstructions allow for correspondences between quite distinct phonemes. In contrast Klesment. Künnap, Soosaar, Taagepera 2003 allows for only minor variants such as *u/w* (Feature 17 in Appendix), *i/j* (Features 21 and 24), and at most, *m/b* (Feature 32). But of course, the quality of features used here also varies.

[4] The limitations of linguistic expertise excluded other languages such as Korean and Japanese that S. Georg and A. Vovin (2003) include in a "Macro-Tungusic" or "Manchuric" group, as well as Ainu and Gilyak. Compared to the Eurasiatic macrogroup proposed by J. Greenberg (2000), the database excludes Indo-European, Korean, Japanese, Ainu and, of course, Etruscan, while adding Yeniseian. Compared to earlier Nostratic approaches, the database also excludes Kartvelian, Dravidian and Afro-Asiatic, while adding Paleo-Siberian languages. This choice of languages by no means implies denial of any wider connections. It would be easier to criticize the scope chosen than to propose another set of comparable extent that would itself be immune to critique. Widening it to everything from Japanese to "Standard Average European"-Sprachbund (Haspelmath 1998) and beyond would of course be the eventual goal (see Conclusions) — but first the novel methodology proposed here should be tested in a more restricted range.

This database has its limitations but still remains the broadest such attempt that we are aware of, when limiting oneself to an ahistorical approach.

This database concerns those phonetic and grammatical features that occur in at least some languages in the Uralic grouping and also occur in some other languages of the northern belt of Eurasia. This approach, also espoused in the present study, offers similarities and differences when compared to the approaches used in previous well-known studies that also compare the structures of languages in various groupings.

The most prominent of such studies is the book by Johanna Nichols, "Linguistic Diversity in Space and Time" (1992). Her impressive world-wide sample consists of 174 languages, including several extinct ones. Three of them are Uralic: Hungarian, Permic Zyrian Komi and Samoyedic Nenets. Structural features rather vocabulary are her central concern, a focus that we share. However, the features considered overlap only in part (e.g., possessive affixes). Nonetheless, it would make sense in the future to compare our data and those of J. Nichols (and also of M. Fortescue (1998)) regarding northern Eurasia. The present study may be considered as complementary to the much more extensive work by J. Nichols, using an appreciably different approach.[5]

Among the more recent works the extensive article by D. Ringe, T. Warnow and A. Taylor (2002) stands out. It deals with the Indo-European language grouping and considers 22 phonological, 15 morphological and 333 lexical characteristics. In contrast to ours, this study focuses on a single language grouping. It also stresses the genetic relationships among its members, and takes into account lexical commonalities.

The most surprising result emerging from our present-oriented approach is that Hungarian looks like the most isolated among the Uralic subgroups and also the most remote from Turkic and other Altaic subgroups. We further observe unexpected subgroups within the Uralic and a loose but still recognizable Paleo-Siberian concatenation (Yukaghir/Eskimo-Aleut/ Chukotka-Kamchatkan/Yeniseian). Turkic and Tungusic appear as having as many structural commonalities with Uralic (except Hungarian) as with each other. We do not claim that this delineation of groupings should replace the customary one, given the limitations of the database used. We can only say: To the extent this database is significant, it would lead to such conclusions.

---

[5] Major differences between our approach and J. Nichols' (1992) include the following. J. Nichols focuses on the diversity of populations of languages, while we focus on their distances (in terms of lack of commonalities). J. Nichols considers sample individual languages (such as 3 out of the many Uralic languages) over a wide range, while we consider all languages conventionally labeled Uralic, Altaic and Paleo-Siberian, lumping languages so similar that no controversy about related-ness has arisen. Our analysis applies a binary YES/NO criterion to the existence or non-existence of a feature in the given subgroup of languages; J. Nichols has room for more nuanced specifications. As a technical difference, J. Nichols' book-size study includes the details of features studied and the sources, while our article refers to Klesment, Künnap, Soosaar, Taagepera 2003 in this respect.

2. A general method for visualizing inter-category distances

The purpose here is to convert into visual distances a table of numbers of commonalities of features (such as structural traits of languages) shared by categories (such as languages). The goal is a graphical representation where two identical languages coincide (zero visual distance), while two languages with nothing in common are at a maximum distance from each other. Our approach differs from Douglas Biber (1986) and may be considered complementary to it.

Consider the simple Table 1 with four hypothetical languages, A, B, C and D. Among a much larger number of features considered, the figures shown in bold along the diagonal indicate how many occur in each language: 10 in A and D, and 20 in B and C. The other numbers indicate how many of the features are common to two of the languages. A and C have no commonalities, while A and D have a perfect overlap and thus are identical, as far as the features considered are concerned. B and C have one-half of their 20 features in common. A and B also have 10 features in common, but here a contrast emerges. These 10 represents all of the features occurring in A, but only one-half of those occurring in B.

Table 1

Hypothetical commonalities of four languages

	A	B	C	D
A	**10**	10	0	10
B		**20**	10	10
C			**20**	0
D				**10**

A reasonable measure of commonality (c_{ij}) of features in languages i and j should satisfy the following conditions:

When there are no common features, $c_{ij} = 0$.

When all features are in common, $c_{ij} = 1$ unit.

When one-half of the features are in common, $c_{ij} = 0.50$. The simplest measure that satisfies these conditions is

$$c_{ij} = a_{ij}/(a_i a_j)^{0.5},$$

where a_i and a_j are the total numbers of features occurring in languages i and j, respectively, and a_{ij} is the number they have in common. Thus for languages A and B, $c_{ij} = 10/(10 \times 20)^{0.5} = 0.707$. The index is 1 for A and D, and 0.50 for B and C, as required.

If we want to show visually the relative distances (d_{ij}) among languages, all we have to do is to define d_{ij} as $1 - c_{ij}$:

$$d_{ij} = 1 - a_{ij}/(a_i a_j)^{0.5}.$$

Identical languages (such as A and D in Table 1) are at zero distance from each other, and those with no commonalities whatsoever (A and C) are at a distance of 1 unit. This 1 unit is the utter maximum for d_{ij}, the end of

the world, so to say.[6] Languages having one-half of the features in common (B and C) are at a distance of 0.50 units. When one language has everything in common with another, yet the latter has only one-half in common with the first (A and B, as well as B and D), the distance turns out to be 0.293. Table 2 shows the distances corresponding to the raw counts of commonalities in Table 1. Languages are, of course, at zero distance from themselves.

Table 2

Relative distances among the hypothetical languages in Table 1

	A	B	C	D
A	**0**	0.29	1.00	0.00
B		**0**	0.50	0.29
C			**0**	1.00
D				**0**

Now suppose three languages, K, L, and M, sport distances $d_{KL} = 0.50$, $d_{KM} = 0.40$ and $d_{LM} = 0.20$. L and M are much closer to each other than either is to K. We can show it visually by constructing a triangle with 0.50, 0.40 and 0.20 as sides — see Figure 1a. However, Table 1 is devised so as to expose a potential problem. From distances in Table 2, we cannot construct a triangle ABC, because 0.293 + 0.50 falls short of 1. In actual practice, however, it is rare to encounter languages far from each other, yet close to a third language. In such cases we can still visualize the distances as shown in Figure 1b. Therefore, we stick to the definition of distance as given above. In the Northern Eurasian data set used, cases like the one in Figure 1b do not occur.[7]
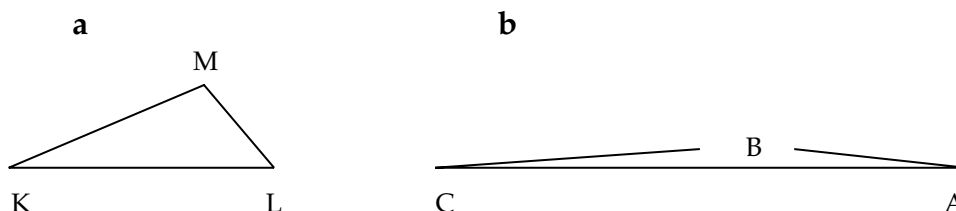

Figure 1. **Distances among hypothetical languages.**

---

[6] It could be well argued that languages with no commonalities are at an infinite distance from each other. This is easily done — just use $d'_{ij} = d_{ij}/(1 - d_{ij}) = (a_i a_j)^{0.5}/a_{ij} - 1$. Finite distances, however, are handier for visual representation.

[7] There are ways to define distance so as to avoid such a paradox, but they introduce further inconsistencies. A more general formula to tie a measure of distance to $c_{ij} = a_{ij}(a_i a_j)^{0.5}$ is $d''_{ij} = [1 - a_{ij}{}^m]^{1/n}$. All such measures have a range from 0 to 1. Our measure d_{ij} corresponds to $m = 1$, $n = 1$. If we set $m = 2$ and $n = 1$, we always can construct a triangle for three languages. However, languages that have one-half the features in common now appear at a distance not 0.5 but 0.707 units from each other — awkwardly close to the utter maximum distance of 1. Such an outcome blurs the distinction between relatively close and extremely distant languages. No combination of m and n can satisfy at once both of these conditions (half-commonality at 0.5 and ability to always construct a triangle). A table of Northern Eurasian distances calculated with $m = 2$, $n = 1$ is given in Klesment, Künnap, Soosaar, Taagepera 2003.

3. Distances of Northern Eurasian languages

Among the conventional Uralic [U] grouping, our analysis considers separately the following subgroups: Finnic [FIN], Saamic [SAA], Mordvin [MOR], Mari [MAR], Permic [PER], Hungarian [HUN], Ob-Ugric [OU], and Samoyedic [SAM]. The conventional Altaic [A] group includes Turkic [TUR], Mongolic [MON] and Tungusic [TUN]. In Siberia and North America, Yeniseian [YEN], Yukaghir [YUK], Chukotka-Kamchatkan [CHU], and Eskimo-Aleut [ESK] are included, referred to as the Paleo-Siberian [PS] languages. The 46 features considered are listed in Appendix, which also discusses some strengths and weaknesses of this choice of features.

Why is it that we take an agnostic approach to the broader genetic relationships (Altaic, Uralic, even Ugric), yet build our analysis on groupings like Finnic, Samoyedic and Mongolic, rather than individual languages such as Finnish, Nenets and Khalka? The reason is that similarities within the subgroups listed are so overwhelming that, to our best knowledge, no serious controversy has arisen on whether a given language belongs or does not belong to the subgroup. Going down to the level of individual languages or even dialects would add little, as long we make sure that distances inside the subgroups are indeed small compared to distances among the subgroups. We'll soon carry out a critical test in this respect, comparing Finnish and Estonian, but let us first consider the broad picture.

Table 3 shows the number of commonalities (a_{ij}) between these subgroups and, on the diagonal, the number of features occurring in the given subgroup (a_i). This number ranges from 15 (YEN) to 41 (MOR and SAM). The number of commonalities for two subgroups ranges from 9 (TUR/YEN) to 37 (MOR/SAM). The order in which subgroups are presented is explained in connection with the next table.

Table 3

**Raw count of commonalities among Northern Eurasian languages,
based on 46 features**

	HUN	OU	PER	MAR	FIN	SAA	MOR	SAM	TUR	TUN	MON	YUK	ESK	CHU	YEN
HUN	31	30	28	28	28	27	29	29	23	20	16	15	19	17	10
OU		37	28	28	32	32	34	35	26	26	20	16	21	20	12
PER			33	32	32	30	32	31	25	24	20	15	18	16	10
MAR				36	32	32	33	33	26	27	22	15	19	17	10
FIN					36	35	36	33	26	28	23	17	20	19	10
SAA						36	35	34	26	28	23	17	20	19	10
MOR							41	37	28	29	25	20	24	22	14
SAM								41	28	29	25	15	24	23	14
TUR									30	25	23	16	17	17	9
TUN										32	25	18	18	17	10
MON											27	17	17	15	10
YUK												21	18	15	11
ESK													26	18	13
CHU														24	13
YEN															15

Table 4 shows the distances among these subgroups as obtained by plugging the values of a_{ij} and a_i in Table 3 into the formula $d_{ij} = 1 - a_{ij}/(a_i a_j)0.5$. These distances range from 0.10 or less (for MOR/SAA, MOR/FIN, SAA/FIN, FIN/PER, PER/MAR, PER/OU and MAR/OU) to 0.58 (for TUR/YEN).[8] Recall that our graphical representation is such that two identical languages coincide (distance 0.00), while two languages with nothing in common are at a unit distance (1.00) from each other. Thus MOR and SAA (0.10) appear distinct but fairly close, while TUR and YEN (0.58) appear much more distant, although still far from utter lack of similarities (1.00).

Table 4

Language distances in Northern Eurasia: $d_{ij} = 1 - a_{ij}/(a_i a_j)^{0.5}$**, with values of** a
from Table 3. Thick lines indicate "error" locations, as explained in text

	HUN	OU	PER	MAR	FIN	SAA	MOR	SAM	TUR	TUN	MON	YUK	ESK	CHU	YEN
HUN	0	0.11	0.12	0.16	0.16	0.19	0.19	0.19	0.25	0.36	0.45	0.41	0.33	0.38	0.54
OU		0	0.08	0.07	0.12	0.12	0.13	0.10	0.22	0.24	0.37	0.43	0.32	0.33	0.49
PER			0	0.07	0.07	0.13	0.13	0.16	0.21	0.26	0.33	0.45	0.39	0.43	0.55
MAR				0	0.11	0.11	0.14	0.14	0.21	0.20	0.29	0.43	0.38	0.45	0.57
FIN					0	0.03	0.06	0.14	0.21	0.18	0.26	0.38	0.35	0.35	0.57
SAA						0	0.09	0.12	0.21	0.18	0.26	0.38	0.35	0.35	0.57
MOR							0	0.10	0.20	0.20	0.25	0.32	0.26	0.3	0.44
SAM								0	0.20	0.20	0.25	0.49	0.26	0.27	0.44
TUR									0	0.19	0.19	0.36	0.39	0.37	0.58
TUN										0	0.15	0.31	0.38	0.39	0.49
MON											0	0.29	0.36	0.41	0.50
YUK												0	0.23	0.33	0.38
ESK													0	0.26	0.34
CHU														0	0.31
YEN															0

The subgroups are presented in such an order as to make the distances increase from left to right and from the bottom to the top. In other words, group distances increase starting from the diagonal and moving toward the top right corner. Such an arrangement is usually not possible without "errors", meaning instances where distances decrease when they "ought to" increase. We count as meaningful the errors that surpass 10% of the preceding distance, rounded to full 0.01. Thus, on the first line (HUN), the drop from MOR to SAM (0.19 − 0.17 = 0.02) does not surpass 10% of 0.19 (which is 0.02) and is not counted as an error. In contrast, the drop from YUK to ESK (0.41 − 0.33 = 0.08) does surpass 10% of 0.41 and is counted. In Table 4, the number of such e r r o r s (s h o w n a s t h i c k l i n e s)

8 How robust are these results against adding or subtracting features? Suppose we add the use of polysyllabic words, which is common to all languages considered but would distinguish them from Chinese. The impact would be the strongest in the case of languages with the fewest features noted, YEN and YUK. Features occurring in YUK would increase from 21 to 22, and those in YEN from 15 to 16. Their present overlap of 11 would become 12. As a result, d_{ij} would shrink from 0.38 to 0.36. The change would be even less extensive for all other language pairs. Thus the distances are fairly robust against minor changes in the choice of features.

is 9 out of possible maximum of 91 when going from left to right, and 10 out of 91 when going from the bottom to the top. Thus the error rate is a tolerable 10.4%. All other ways to order the subgroups seem to increase the number of errors.

Such minimization of "errors" places the conventional subgroups of U and A next to each other, while placing the conventional PS groups next to Altaic. FIN/SAA/MOR and OU/PER/MAR appear as the closest-knit clusters, with FIN and PER linking these two clusters. "Errors" bunch in the following locations. Four U subgroups (HUN and OU, MOR and SAM) are as close or closer to ESK and CHU than to MON. ESK is closer than YUK to all U subgroups. Contrary to traditional expectations, minimization of "errors" forced us to place HUN far away from the Altaic. HUN appears as the U language the most distant not only from MON and TUN but also from TUR! We'll come back to this finding.

How do intra-subgroup distances compare with the between-subgroup distances shown in Table 4? We carried out a sample test on two Finnic languages, standard literary Finnish and Estonian. This is a severe test, because dialectal occurrences of features are excluded and the much-shortened standard Estonian lacks a number of typically Finnic features, such as possessive suffixes and negative auxiliary verbs. Among the 36 features that occur in the FIN group, 35 occur in Finnish but only 30 do in Estonian. Their overlap is the same 30, resulting in a distance of 0.07 units. This distance surpasses the FIN/SAA distance (0.03), the shortest inter-group distance in Table 4. This is so because the inter-subgroup distance registers commonalities that may occur in only one language or dialect in each subgroup.

To repeat, restriction to literary Estonian is a severe test. Including Estonian dialects would reduce the apparent distance to FIN appreciably. Still, this test should make us cautious. Distance differentials of less than 0.05 may not be significant.

3.1. Distances among Uralic, Altaic, and Paleo-Siberian

The conventional U and A groups emerge with fair clarity in Table 4. All distances within U and within A are less than 0.20, while nearly all other distances equal or surpass 0.20 (exceptions: SAA/TUN and FIN/TUN, 0.18). Distances between individual subgroups of U and subgroups of A range from 0.20 to 0.36, except the aforementioned SAA/TUN and FIN/TUN on the low side and HUN/MON (0.45) on the high side. This range is comparable to that among the PS groups (0.23 to 0.38), suggesting that PS and U-A are comparably loose super-groups. The distances between individual subgroups of U-A and individual subgroups of PS range from 0.26 to 0.58, mostly exceeding the typical range within U-A or within PS.

Within the loose PS cluster, a concatenation YUK-[0.23]-ESK-[0.28]-CHU-[0.31]-YEN appears, with all other distances larger than 0.33. YEN stands out as the language most distant by far from U and A, and also from the other PS languages. The other PS languages are at a mean distance of 0.28 from each other, while their mean distance to YEN is 0.34. Hence it is advisable to keep YEN separate from the rest of the PS languages. This result

agrees with Seefloth (2000), who groups YUK/ESK/CHU. But even with YEN removed, PS remains a cluster more diffuse than Uralic-Altaic. The "Uralo-Eskimo" connection (Seefloth 2000) does not emerge here but will be discussed later on.

<div align="right">*Table 5*</div>

Average distances within and between major language groups

	U	A	YUK/ESK/CHU	YEN
Uralic	(0.12)	0.25	0.37	0.53
Altaic		(0.17)	0.36	0.52
YUK/ESK/CHU			(0.28)	0.34
YEN				(0)

Table 5 shows the mean distances of subgroups in these four group-ings (U, A, YUK/ESK/CHU, and YEN) to subgroups in the other groups. Also shown, in parentheses, are the mean distances within these groups. The mean distance between the subgroups of U and subgroups of A (0.25) is less than the mean intra-YUK/ESK/CHU distance (0.28). The latter, in turn, is shorter than the mean distances from YUK/ESK/CHU to U (0.37) or A (0.36). Thus grouping YUK/ESK/CHU together makes some faint analytic sense. This grouping is almost as distant from YEN (0.34) as from A and U.
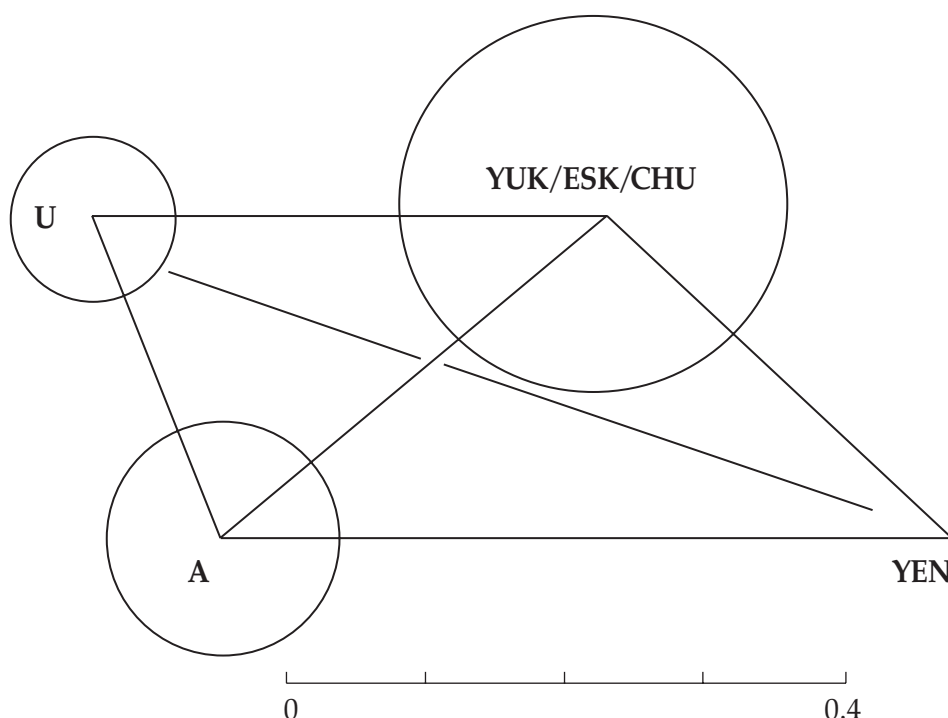

Figure 2. **Distances among major language groups in Northern Eurasia.**

We can easily represent any three of these distances graphically (Figure 2), when limited to two dimensions.[9] We can also join some edges of such triangles, so as to represent the distances to some further groups. In Figure 2, all distances are rendered exactly, except the longest, from U to YEN. The latter corresponds to the length of the interrupted line that does not reach the U and YEN corners. We can mentally raise the YEN corner out of the plane of the paper, until the distance U-YEN becomes 0.53, as given in Table 5. Then the four groups would form a tetrahedron where all four distances correspond exactly to those shown in Table 5.

The diameters of the circles shown correspond to the mean distances of subgroups within the major groupings, as indicated on the diagonal of Table 5. They remind us that within-group distances are sometimes comparable to those among the groups (especially regarding YUK/ESK/ CHU).

This visual presentation highlights the observation that, compared to YEN and YUK/ESK/CHU, Uralic and Altaic are relatively close, so that the old notion of a Uralic-Altaic super-group surfaces. YUK/ESK/CHU is equally distant from U, A, and YEN, which is far away from the rest.

3.2. Distances among Uralic and Altaic subgroups

It was noted that FIN/SAA/MOR and OU/PER/MAR appear as the closest-knit clusters, with all intra-cluster distances being less than 0.10. FIN and PER link these two clusters. Fusing the subgroups within these clusters helps to visualize their mean distances to other subgroups. We are left with 4 clusters within U. Their average distances among themselves and to subgroups of Altaic are shown in Table 6. Once again, the subgroups / clusters are presented in such an order as to make the distances increase from left to right and from the bottom to the top. No "errors" arise, by the criteria stated earlier.

Table 6

Average distances within and between the subgroups of Uralic and Altaic

	HUN	O/P/M	F/S/M	SAM	TUR	TUN	MON
HUN	(0)	0.13	0.18	0.19	0.25	0.36	0.45
OU/PER/MAR		(0.07)	0.12	0.13	0.21	0.23	0.36
FIN/SAA/MOR			(0.06)	0.12	0.21	0.19	0.26
SAM				(0)	0.20	0.20	0.25
TUR					(0)	0.19	0.19
TUN						(0)	0.15
MON							(0)

The distances among the clusters within U are visualized in Figure 3. All distances are exact, except the longest, from HUN to SAM. The latter distance corresponds to the length of the line that does not reach the HUN and SAM corners. The clusters OU/PER/MAR, FIN/SAA/MOR and SAM are at essentially equal distances from each other, while HUN stands apart,

---

[9] To represent fully the distances between N groups, it would in general take an (N − 1) dimensional space. Thus, even the use of 3-dimensional models would not allow us to go much further.
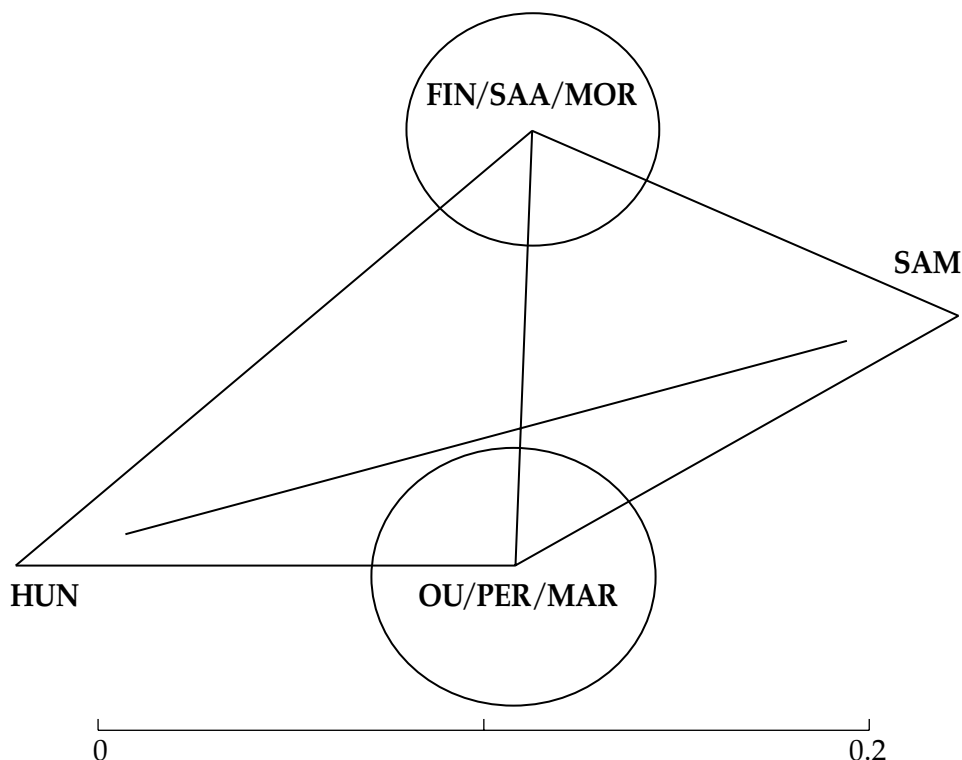
Figure 3. **Distances among four clusters within Uralic.**

being fairly close only to OU and PER in the OU/PER/MAR cluster. In other words, instead of the conventional division of U into Finno-Ugric and SAM, this set of features divides U into Finno-Samoyedic and Hungarian.

The "Ugric" commonalities of HUN and OU stand out only when viewed from the Hungarian vantage point — as was historically the case in Uralistics. From this vantage point, Ob-Ugric is indeed the closest group, but the reverse it not true. From the OU vantage point, not only MAR and PER but also SAM offer more commonalities than HUN. Moreover, the other three subgroups (FIN, SAA, MOR) are not much further than HUN.

Within this pattern of groups within groups, MOR and SAM offer a cross-cutting disturbance. MOR is close to FIN and SAA, yet also has features that make it close to SAM (0.10). Moreover, more than any other U or A subgroup, both MOR and SAM have also features in common with ESK (distance 0.26) and CHU (0.30 and 0.27, respectively). We'll return to this issue.[10]

---

[10] On grounds of current or recent geographical locations, some PS commonalities with SAM and even OU might be expected, but MOR is more puzzling. MOR does not stand out by its links to Altaic, yet has more commonalities with PS than does any other language presently spoken west of the Urals. Along with strong linkages to Finnic-Saamic in the west, Mordvin also contains a persistent eastern strain, as previously noted by J. Pusztay (1994), who joins Mordvin to Ugric and Samoyedic, to form the eastern trunk of Uralic.
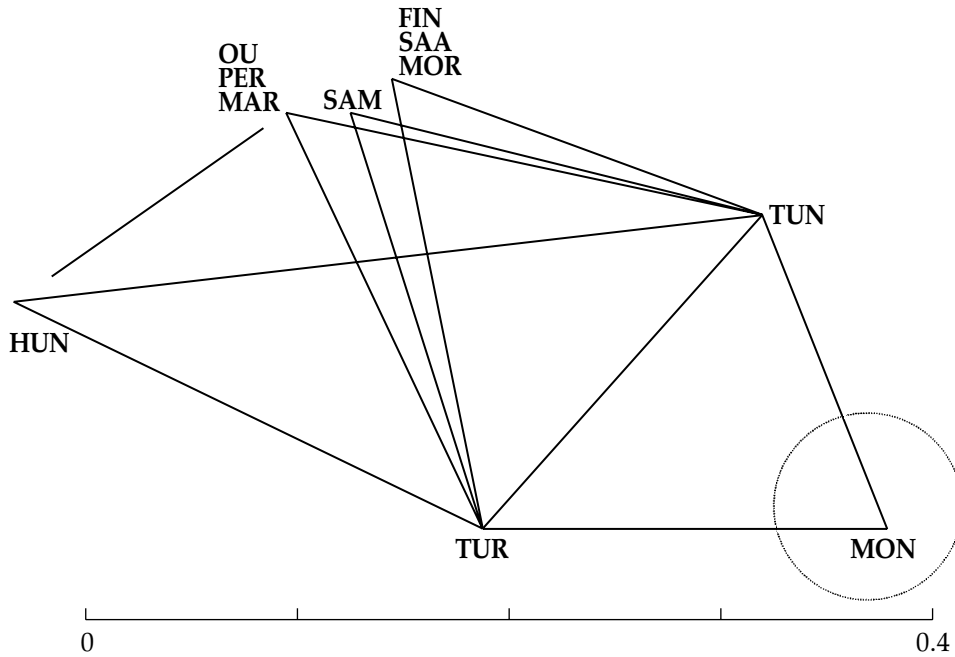
172

Figure 4. **Distances of Altaic and Uralic subgroups.**

Figure 4 shows the distances among the three subgroups of A. It also shows the distances of clusters within U to TUR and TUN. If distance lines were drawn from clusters within U to MON, they would all end within the dotted circle shown. The distance from HUN to OU/PER/MAR corresponds to the length of the line that does not reach the OU/PER/MAR and HUN corners.

Within A, MON and TUN appear closer to each other (0.15) than HUN is to most U subgroups. TUR appears equidistant from MON and TUN. In their distance to Uralic subgroups, TUR and TUN are pretty much on a par, except for HUN, which is very distant from TUN while also exceptionally far from TUR. From the vantage point of TUR, all U subgroups but HUN are almost as close (0.20 to 0.21) as TUN and MON (0.19). Within the U-A super-group, the TUR-TUN axis looks central, with MON on the one side and the U subgroups on the other — and not much more distant, except for HUN.

The remoteness of HUN and TUR (0.34) is notable in view of age-old claims of a tie between them, competently reviewed by A. Marcantonio (2002). In our analysis, however, what distinguishes HUN from the other Uralic subgroups is not Turkic or other Altaic affinities — to the contrary, Hungarian appears as the Uralic subgroup the most distant from Turkic (and other Altaic).

We should consider a possible artifact. The previous example of Estonian and Finnish reminds us that a subgroup with many languages and dialects is more likely to harbor a given feature in some of its corners, especially when history has split these languages or dialects politically, thus increasing their isolation. In contrast, a subgroup consisting of a single language is less likely to offer such variety, especially if this is a state language with a long literary tradition that has absorbed the dialects. This is of course the

173

case for HUN. Only 31 of the features considered occur in HUN, which does reduce the potential for commonalities with other languages.

Standard Estonian, which exhibits 30 features, would be in a comparable situation, if it were considered a separate subgroup. As a test, the distances of standard Estonian to the subgroups of U and A were calculated. The mean distance of Estonian to the subgroups of U apart from HUN is 0.16 — the same as for HUN. Thus the single-language factor may be among the causes of the apparent isolation of HUN. However, the distances of Estonian to TUR (0.18), TUN (0.23) and MON (0.33) are much shorter than is the case for HUN (0.25, 0.36 and 0.45, respectively — cf. Table 6). Thus the single-language factor cannot be the sole cause for the isolation of HUN. HUN does occupy a special position. In some ways, it may indeed be the "least Altaic" among the U languages.

The Samoyedic-Tungusic commonalities pointed out by A. Marcantonio (2002) turn out to be few. Compared to other subgroups of U (except HUN), SAM does not emerge as meaningfully closer to TUN, geographical contiguities notwithstanding. FIN/SAA/MOR, the Uralic cluster geographically furthest from TUN, has as many commonalities.

It should be recalled that relatively few of the 46 features in the database occur in any Altaic language group — the number ranges from 32 in Tungusic to only 27 in Mongolic. In contrast, the count for Uralic languages ranges from 41 to 31 features. It may well be that Altaicists could complement the present list with numerous features common to Altaic subgroups and possibly absent in the Uralic languages, changing the picture in Figure 4 appreciably. We hope to induce them to do so.

3.3. Distances within and around Paleo-Siberian

We do not consider PS a language group (like U or A) but designate by this term a residual category of languages that belong to neither U or A. Previous Table 4 showed that MOR and SAM have more commonalities with some PS languages than any other subgroup of U or A. Following up on this observation, Table 7 inserts the average distance of MOR/SAM among the PS languages is such a way as to minimize "error" as previously defined. Indeed, only one such error occurs (shown as a thick line) when MOR/SAM is inserted between CHU and YEN. The tightest cluster in this Table is not YUK/ESK/CHU but ESK/CHU/MOR/SAM. If no other Uralic languages existed, we would happily classify MOR and SAM as belonging to PS. In this limited sense the "Uralo-Eskimo" hypothesis (Seefloth 2000) receives support.

Table 7

**Distances among the Paleo-Siberian languages
and the average of Mordvin and Samoyedic**

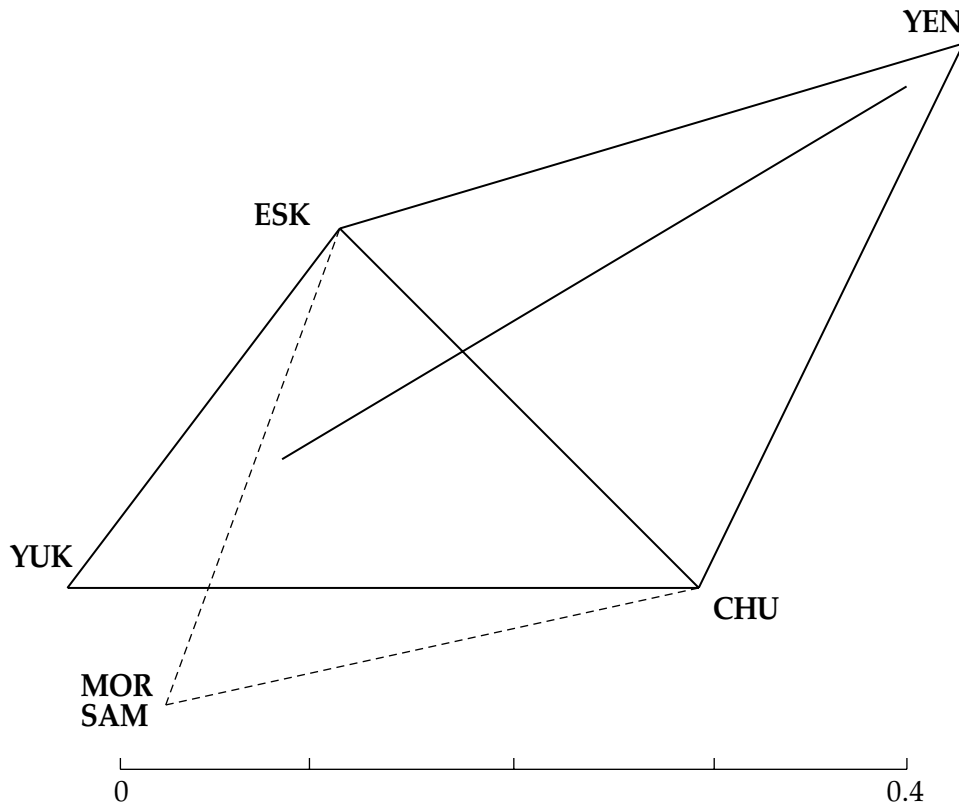|	YUK	ESK	CHU	MOR/SAM	YEN
YUK	0	0.23	0.33	0.40	0.38
ESK		0	0.26	0.26	0.34
CHU			0	0.28	0.31
MOR/SAM				(0.10)	0.44
YEN					0

Figure 5. **Distances among Paleo-Siberian languages and Mordvin/Samoyedic.**

Figure 5 highlights this relationship. ESK-CHU-YUK and ESK-CHU-MOR/ SAM form similar almost equilateral triangles. The distance between YUK and MOR/SAM is actually much larger than it appears in the flat scheme. It is as if the MOR/SAM corner were lifted out of the plane of the page. On the other hand, the distance between YUK and YEN is shorter, as indicated by the length of the line that does not reach the corners.

4. Conclusions

What are the major results, and how adequate are the methods and data used to obtain them? We will first discuss the general method, followed by results and their validity in view of the data set used.

The distance measure used is anchored at values 0, 1/2, and 1 in a logical way, and the simplest measure satisfying these conditions was adopted. No obvious dissonances have been observed that would suggest adoption of a more complex expression. Two-dimensional visualization limits us to the use of a concatenation of triangles, and there is some leeway in the order in which languages (or groups) are chosen to form such triangles. Some inter-group distances are not shown in the figures and, despite the lack of a joining line, the visual impression may distort the distance. The only way to avoid such risks fully would be to multiply the number of

graphs or resort to representations using more dimensions than two. We believe we have avoided wrong impressions.

As for the Uralic grouping, we observe that it does exist, A. Marcantonio's (2002) well-founded doubts notwithstanding. Practically all intra-Uralic distances are shorter than those to any languages outside Uralic. A. Marcantonio's reminder about Uralic-Altaic ties is well taken. Turkic, in particular, has almost as many commonalities with most Uralic subgroups as with Tungusic and Mongolic. The major surprise, however, is that Hungarian emerges as the Uralic language t h e m o s t r e m o t e from Turkic and from Altaic in general, instead of being the closest. The commonalities between Samoyedic and Tungusic, as pointed out by A. Marcantonio (2002), hardly exceed the Uralic average.

Within Uralic, the cluster of Finnic, Saamic and Mordvin is confirmed (despite the latter's curious Paleo-Siberian ties). In the geographic center of the Uralic area, Mari, Permic and Ob-Ugric are found to form another tight cluster. Most surprising, Hungarian rather than Samoyedic emerges as the most distant from all other Uralic languages.

The traditional subdivisions of Uralic have been subject to revision for some time already — and not surprisingly so. A language can slowly replace most of its vocabulary while preserving most of its grammar. Thus various lexical and structural considerations can lead to quite different ways to cut the Uralic pie. Table 8 shows some proposals for cuts into subgroups and sub-subgroups. Like the traditional view, T.-R. Viitso (2000) assumes a gradual branching off from a proto-language and implicitly places SAM apart. Within the rest, he allows for two possibilities: I. MOR/MAR grouped with FIN/SAA, or II. MOR/MAR grouped with PER/HUN/OU. Compared to traditional, PER is grouped with HUN/OU in both variants. J. Pusztay (1994) leaves the question of proto-language more open. He sees only a western and an eastern subgroup, and pushes Mordvin to the east.

Table 8

Subdividing the Uralic language group

	West	Central-West	Central-East	East
Traditional	FIN-SAA	MOR/MAR/PER	HUN/OU	SAM
Pusztay (1994)	FIN-SAA/MAR/PER		MOR/HUN-OU/SAM	
Viitso (2000) I	FIN/SAA/MOR/MAR		PER/HUN/OU	SAM
Viitso (2000) II	FIN/SAA	MOR/MAR/PER/HUN/OU		SAM
Present study	HUN	FIN-SAA/MOR	PER-MAR-OU	SAM

Our analysis yields four roughly equidistant clusters and takes no stand on how they came about. But these are not the traditional four groups, to which T.-R. Viitso's groupings (2000) are close. Along with J. Pusztay (1994), we note a Finnic-Permic connection but keep Mordvin tied to Finnic-Saamic, despite its Paleo-Siberian commonalities. Our analysis is the only one to set Hungarian apart from the rest. Yes, Ob-Ugric remains the subgroup closest to Hungarian, but the reverse is not the case: Ob-Ugric appears even closer to Mari and Permic. Instead of dividing Uralic into the traditional Finno-Ugric and Samoyedic branches, our results divide it into Finno-Samoyedic and Hungarian.

The set of features underlying our analysis is extensive and discriminating as far as Uralic is concerned. The 46 features considered were chosen such that every feature occurs in at least one of the 8 Uralic language groups considered. They include 25 that occur in all 8 Uralic language groups. Thus the resulting groupings are based on a fair number of features, and the outcome could not have been predicted ahead of time through biased selection of inputs. The four clusters within Uralic that emerge from our study deserve serious consideration, along with the other formats shown in Table 8.

Regarding Altaic, our most surprising finding is that Turkic appears almost as close to most Uralic subgroups as to Tungusic and Mongolic. The choice of features might be adequate, given that 37 features occur in at least one Altaic language and 21 occur in all of them. But probably further features common to Altaic and absent in Uralic subgroups could be found, thus reducing the distances within Altaic.

As for Paleo-Siberian non-group, 34 of the features considered occur in at least one language but only 6 occur in all of them. Yeniseian stands apart, but the other three are also as remote from each other as Uralic is from Altaic. There are hints of a deep layer common to Uralic-Altaic and the Paleo-Siberian languages. There are also hints of reciprocal borrowings among the Paleo-Siberian languages and two Uralic languages — Samoyedic and, most puzzling, Mordvin.

The broad picture is one of definite commonalities between Uralic and Altaic, and more distance from them to the hazy Paleo-Siberian grouping. All these findings are tentative and subject to cross-checking through other approaches. At the same time the method used looks of sufficient promise to recommend that analogous sets of structural features be collected for the languages around the Mediterranean to establish a comparable set of features so as to compare Semitic, Hamitic, Indo-European and Basque languages. Extension to Dravidian, Korean, Japanese and Ainu would also be of interest.

APPENDIX: Database

The 46 structural features considered

P. Klesment, A. Künnap, S.-E. Soosaar and R. Taagepera (2003) considered the occurrence of 60 features in U, A, PS, and also IE. We exclude from our analysis the Indo-European languages and 14 of the features, for the following reasons.

The database focuses on features that occur in northern Eurasia and does not include the numerous features common to Indo-European languages. Until one adds such features, one cannot draw any conclusions on the distances within IE or its distances to U, A and PS.

Once Indo-European is omitted, 14 of the features occur in only 1 or 2 of the 15 subgroups considered. (All others occur in at least 4.) The 10 "Finnic-Baltic" features connect FIN (and sometimes SAA) to Baltic and Slavic (and sometimes Germanic), while being completely absent in other Indo-European subgroups as well as in any other U, A or PS subgroup. By increasing the total count (a_i) for FIN, they artifactually increase the distances $d_{ij} = 1 - a_{ij}/(a_i a_j)^{0.5}$ from FIN to all other U groups. Similarly, the 4 "Samoyedic" features, which spill over to only one

PS or A subgroup, artifactually increase the distances from SAM to all other subgroups. Therefore, we omit these 14 features, leaving a total of 46.

These 46 features are listed below, indicating their occurrence in the subgroups of U, A, and PS, as well as in Germanic [GER], Slavic [SLA], Baltic [BAL]. The labels of groups of features are those in Klesment et al. (2003). IE indicates occurrence in Indo-European languages other than G-S-B = GER-SLA-BAL. For the same data in the form of a matrix for each feature in each language, see Klesment, Künnap, Soosaar, Taagepera 2003, which also lists the source materials for each feature.

["Nostratic"]

1. Vowel qualitative alternation — IE, G-S-B, Uralic, YEN, CHU
2. 'To be at' used in the sense of 'to have' — IE, SLA, BAL, Uralic (except HUN[11]), Altaic, YEN[12]
3. Use of postpositions — G-S-B, Uralic, Altaic, Paleo-Siberian (except YEN)[13]
4. Suffix of deverbal nouns/adjectives -*l* — G-S-B, Uralic, Altaic, Paleo-Siberian (except YEN)
5. Suffix of denominal and deverbal nouns -*m* — SLA, BAL, Uralic, Altaic, Paleo-Siberian (except YEN)
6. Use of nominalisation to form subordinate sentences, instead of 'that' + finite verb form — SLA, BAL, Uralic, Altaic, YEN, ESK
7. Hortative/imperative suffix -*k* — SLA, BAL, Uralic,[14] TUR, TUN, ESK, CHU
8. Construction 'to be afraid' + ablative — IE, GER, Uralic, YUK, YEN, ESK

["Nostratic East"]

9. Locative/ablative suffix -*t* — Uralic, Altaic, Paleo-Siberian
10. Avoidance of word-initial consonant clusters — Uralic, Altaic, Paleo-Siberian
11. Locative suffix -*n* — Uralic, Altaic, Paleo-Siberian
12. Lative/dative/locative suffix -*k* — Uralic, Altaic, Paleo-Siberian
13. Causative/factitive suffix -*t* — Uralic, Altaic, Paleo-Siberian (except YEN)
14. Use of possessive suffixes — Uralic, Altaic, Paleo-Siberian (except CHU)
15. Use of comparative with ablative — Uralic, Altaic, Paleo-Siberian (except CHU)
16. Plural suffix -*t* — Uralic (except HUN), Altaic, Paleo-Siberian (except YEN)
17. Suffix of reflexive verbs -*u/w* — Uralic, Altaic, YUK, CHU
18. Diminutive suffix -*k* — Uralic, Altaic, ESK, CHU

["Uralic-Altaic"]

19. Suffix of infinitive and other infinite verbal forms -*m* — Uralic, Altaic, YUK
20. Suffix of preterite/perfect -*t* — Uralic, Altaic
21. Suffix of adjectives/actors -*j/i* — Uralic, TUR, MON
22. Suffix of adjectives/diminutives -*c/ć* — BAL, Uralic, TUR, TUN
23. Intensifying/conjoining suffix -*k* — Uralic, TUR, TUN
24. Diminutive suffix -*j/i* — Uralic, TUR, TUN
25. Plural suffix -*n* — Uralic (except HUN), MON, TUN, ESK

---

[11] HUN features a compound verb that joins the roots 'vagy' and 'len-' that leads to an expression not quite equivalent to 'to be at', although such exclusion may be debatable.
[12] Occurrence of 'to be at' need not exclude a more formal 'to possess'. Most languages of the world lack a verb for 'have', but in the light of its occurrence in most contemporary IE languages (and in YUK, CHU and ESK), its absence is worth noting.
[13] Use of postpositions does not perforce imply absence of prepositions.
[14] In some Finno-Ugric languages -*k* > -*j* or vowel and in some Samoyedic languages -*k* > glottal stop (Here we may have slipped into diachronic reconstruction).

["Uralic"]

26. Collective/plural suffix *-k* — Uralic, CHU
27. Suffix of preterite/perfect *-m* — Uralic, ESK
28. Suffix of frequentative verbs *-l* — Uralic, TUR
29. Use of negative auxiliary verbs — Uralic (except HUN), TUN, CHU
30. Coaffix of local case suffixes / locative / ablative suffix *-l* — Uralic (except SAA, SAM), TUN
31. Possessive suffix precedes case suffix at least in some cases — MAR, PER, HUN, OU, TUR
32. Accusative suffix *-m/b* — SAA, MAR, OU, SAM, TUN
33. Three dimensions of local case — Uralic, Paleo-Siberian (except YEN)
34. Plural suffix *-j/i* — Uralic (except MAR, PER), TUR, ESK, CHU

["Altaic"]

35. Collective/plural suffix *-l* — MAR, OU, SAM, Altaic, CHU
36. Genitive/oblique cases suffix *-n* — FIN, SAA, MOR, SAM, Altaic, YUK, CHU
37. Directive/lative/translative suffix *-ksi* — FIN, SAA, MOR, SAM, MON, TUN
38. Lative suffix *-s/ś* — FIN, SAA, MAR, MOR, MON, TUN
39. So-called paired verbs — MAR, PER, MOR, SAM, TUR, MON

["Altaic-Paleo-Siberian"]

40. Conjugation of nouns — MOR, SAM, Altaic, Paleo-Siberian
41. Negative particle / negative auxiliary verb *ä(lä)/e(le)* — FIN, SAA, MOR, Altaic, Paleo-Siberian (except YEN)
42. Absence of word initial *r-* — SAM, MON, TUN, Paleo-Siberian (except CHU)

["Paleo-Siberian"]

43. Two functionally different types of conjugation — SLA, BAL, HUN, OU, MOR, SAM, Paleo-Siberian
44. Marking of object in verbal forms — OU, MOR, SAM, Paleo-Siberian (except YUK)
45. Use of object in local cases — MOR, Paleo-Siberian (except CHU)
46. The 1st person ending *-k* — HUN, SAM, ESK, CHU

Strengths and weaknesses of the set of features

The overriding value of this set of structural features supplied by P. Klesment, A. Künnap, S.-E. Soosaar and R. Taagepera (2003) is that it has been collected at all and made available, even while their quality may vary. The choice of features was made by Uralists, and it shows: All 46 features do occur in at least one Uralic language. Thus the extent of occurrences of these features in a given language is in some ways a measure of its "Uralicity". Outside the conventional Uralic group, where the number of occurrences ranges from 31 (HUN) to 41 (MOR and SAM), the most "Uralic" would seem to be TUN (33), TUR (30) and ESK (28). Further structural features should be collected, including those that occur in Uralic languages rarely or not at all, so as to obtain a more balanced picture of the language groups in northern Eurasia. This would need input from specialists on Basque, Indo-European, Altaic, Paleo-Siberian, Ainu and even Dravidian.

The "occurrence" of a feature in a language or group does not necessarily indicate extensive occurrence. It may not be typical of the groups involved and thus may overstate group proximity. On the other hand, it may still indicate that these groups have a predisposition for the rise of such a rare feature.

Abbreviations of language groupings/subgroups/languages

A — Altaic languages; **BAL** — Baltic languages; **CHU** — Chukotka-Kamchatkan (or Chukchi-Kamchatkan) languages; **ESK** — Eskimo-Aleut (or Eskaleutic) languages; **FIN** — Finnic (or Balto-Finnic) languages; **GER** — Germanic languages; **HUN** — Hungarian language; **IE** — Indo-European languages; **MAR** — Mari language; **MON** — Mongolic languages; **MOR** — Mordvin languages; **OU** — Ob-Ugric languages; **PS** — Paleo-Siberian languages; **PER** — Permic languages; **SAA** — Saamic (or Lapp) languages; **SAM** — Samoyedic languages; **SLA** — Slavic languages; **TUN** — Tungusic (or Tungus-Manchu) languages; **TUR** — Turkic languages; **U** — Uralic languages; **YEN** — Yeniseian languages; **YUK** — Yukaghir language.

R E F E R E N C E S

A i k h e n v a l d, A. Y. 2003, Mechanisms of Change in Areal Diffusion. New Morphology and Language Contact. — Journal of Linguistics 39, 1—29.

B i b e r, D. 1986, Spoken and Written Textual Dimensions in English. — Language 62, 384—414.

D i x o n, R. M. W. 1999, The Rise and Fall of Languages, Cambridge.

E m b l e t o n, S. M. 1986, Statistics in Historical Linguistics, Bochum.

F o r t e s c u e, M. 1998, Language Relations across Bering Strait. Reappraising the Archaeological and Linguistic Evidence, London—New York.

G e o r g, S., V o v i n, A. 2003, From Mass Comparison to Mess Comparison. Greenberg's Indo-European and Its Closest Relatives. — Diachronica 20, 331—362.

G r e e n b e r g, J. 2000, Indo-European and Its Closest Relatives. The Eurasiatic Language Family I. Grammar, Stanford.

H a s p e l m a t h, M. 1998, How Young is Standard Average European? — Language Sciences 20, 271—287.

K l e s m e n t, P., K ü n n a p, A., S o o s a a r, S.-E., T a a g e p e r a, R. 2003, Common Phonetic and Grammatical Features of Uralic Languages and Other Languages in Northern Eurasia. — Journal of Indo-European Studies 31, 363—390.

K ü n n a p, A. 2000, Contact-Induced Perspectives in Uralic Linguistics, [München—Newcastle] (LINCOM Studies in Asian Linguistics 39).

M a r c a n t o n i o, A. 2002, The Uralic Language Family. Facts, Myths and Statistics, Oxford—Boston (Publications of the Philological Society 39).

M c M a h o n, A., M c M a h o n, R. 2003, Finding Families. Quantitative Methods in Language Classification. — Transactions of the Philological Society 101, 7—55.

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

Ö s t m a n, J.-O., R a u k k o, J. 1995, The 'Pragmareal' Challenge to Language Tree Models. — The Fenno-Baltic Cultural Area, Helsinki, 31—69.

P u s z t a y, J. 1994, A Linguistic Model of Determining the Original Home of the Uralic Nations. — Финно-угроведение 1 [Йошкар-Ола], 23—40.

—— 1995, Diskussionsbeiträge zur Grundsprachenforschung (Beispiel: das Protouralische), Wiesbaden (Veröffentlichungen des Societas Uralo-Altaica 43).

R e n f r e w, C. 2000, At the Edge of Knowability. Towards a Prehistory of Languages — Cambridge Archeological Journal 101, 7—34.

R i n g e, D. 1999, How Hard Is It to Match CVC-Roots? — Transactions of the Philological Society 97, 213—244.

R i n g e, D., W a r n o w, T., T a y l o r, A. 2002, Indo-European and Computational Cladistics. — Transactions of the Philological Society 100, 59—129.

S e e f l o t h, U. 2000, Die Entstehung polypersonaler Paradigmen im Uralo-Sibirischen. — Zentralasiatische Studien 30, 163—191.

T a a g e p e r a, R. 1994, The Linguistic Distances between Uralic Languages. — LU XXX, 161—167.

—— 2000, Uralic as a *lingua franca* with Roots. — FU 23, 381—395.

T h o m a s o n, S. G., K a u f m a n, T. 1988, Language Contact, Creolization, and Genetic Linguistics, Berkeley.

V i i t s o, T.-R. 2000, Finnic Affinity. — CIFU IX. Pars I, 153—178.

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

РЕЙН ТААГЕПЕРА (Тарту—Ирвайн, Калифорния), *АГО КЮННАП* (Тарту)

РАССТОЯНИЯ МЕЖДУ УРАЛЬСКИМИ
И ДРУГИМИ ЯЗЫКАМИ СЕВЕРНОЙ ЕВРАЗИИ

В статье рассматривается наличие или отсутствие 46 структурных черт в языковых группах — от прибалтийско-финских языков до эскимосско-алеутских. Разработаны нормы измерения общих черт и расстояний между двумя языками, которые использованы для их графического изображения. Вырисовываются конвенциональные группы уральских и алтайских языков, а также совокупность палеосибирских языков (юкагирский/эскимосско-алеутские/чукотско-камчатские/енисейские), более разрозненная, чем даже урало-алтайские языки вместе взятые, но с некоторой близостью из уральских языков к самодийским и мордовским. Среди уральских языков выявляются четыре подразделения: самодийские, мордовские/саамские/прибалтийско-финские, марийский/пермские/обско-угорские и — как самое отдаленное от остальных — венгерский язык. Неожиданным представляется то обстоятельство, что венгерский язык как уральский оказывается наиболее отдаленным именно от тюркских и других алтайских языков, не говоря уже о его роли якобы соединительного моста между уральскими и алтайскими языками. Тунгузские языки проявляют примерно столько же общих черт с уральскими языками (исключая венгерский), как и между собой. Общее происхождение не может служить обяснением всех этих пересекающихся конфигураций; следует считаться с наличием контактов и селективных миграций, с участием в ареалах использования связывающего языка (*lingua franca*). Необходимо расширить сеть имеющихся структурных черт, включить в нее черты баскского, индоевропейских, семито-хамитских и дравидийских языков, а также дополнительные черты алтайских и палеосибирских языков.