Anton Perdih

COMPARISON OF SOME METHODS OF ESTIMATION OF LINGUISTIC DISTANCES

Povzetek

PRIMERJAVA NEKAJ METOD ZA OCENO JEZIKOVNE RAZDALJE

Primerjanih je bilo 6 metod za oceno jezikovne razdalje med 17 jeziki, z enajstimi različnimi branji nekaterih starih jezikov. Različne metode so dale različne numerične vrednosti, zato je bilo treba primerjati njihove normalizirane in sortirane rezultate. Primerjave kažejo, da je starim jezikom, kot so to etruščanščina, stara frigijščina, retijščina in venetščina, po uporabljenih metodah za oceno jezikovne razdalje stara slovenščina večinoma bliže kot latinščina in grščina. Slovenščino torej lahko upravičeno uporabimo za razvozlavanje nekaterih starih jezikov.

Introduction

In a previous paper [1] we demonstrated the usefulness of the method Principal Component Analysis (PCA) for estimation of linguistic distances between 17 mostly ancient languages. The bases for estimation were the frequencies of particular sounds, their pairs and triplets presented as frequencies of signs in the Slovene notation. These frequencies were used since some ancient languages are known from a relatively small number of inscriptions, which are mostly short, broken or incomplete, making the composition of an extended and comprehensive linguistic Corpus difficult. In addition, the majorities of some groups of inscriptions are written *in continuo*, i.e. without separation in words, and do not give any suitable clue about toponyms, verbs, and frequently used words that could be used for computational comparisons between these old languages and other better known languages.

The PCA method gave good results using frequencies of single sounds, whereas using frequencies of sound pairs and especially of sound triplets the results were far from those usually given by this method. One of the possible causes of this effect could be the inability of the available computer to process the large matrices of intermediate data. For this reason we are trying other methods which would be able to process adequately also the frequencies of sound pairs as well as triplets. One possible alternative, the average sum of absolute values of frequency differences was demonstrated earlier on few sets of data and on data for single sounds only [2, 3]. In present paper the results obtained by the PCA method are compared to the results obtained using as the method the sums of absolute values of frequency differences, their root mean squares and root mean quadrics, as well as the measures of quality of the regression.

Data and methods

The sound frequency data of languages Bq, Cs, Es, Et, Fi, Gr, Hi, La, Lu, My, Os, Ph, Rt, Sl, Um, Ve, and Vz, were used as prepared for the previous study [1]. The meaning of these abbreviations is presented in Table 1. The results of distances are presented here from the languages, i.e. their reading variants EtB, EtT, LaC, LaS, PhT, PhA, RtB, RtT, RtV, Sl, VeB, VeT, or VeV, or averages of these variants marked as Et, La, Ph, Rt, Sl, or Ve, which were taken as the origins. The third character in these combinations indicates the following, cf. [1] for detailed references:

A in PhA - the reading according to A. Ambrozic was applied to all considered inscriptions by A. Perdih,

B in EtB, RtB, VeB - the reading according to M. Bor was applied to all considered inscriptions by A. Perdih,

C in LaC - classical reading of Latin,

S in LaS - semiclassical reading of Latin,

T in EtT, PhT, RtT, VeT - the reading according to western scholars was prepared by G. Tomezzoli,

V in RtV and VeV - the reading by V. Vodopivec.

Table 1: Language abbreviations

Language	Abbreviation	Language	Abbreviation
Basque	Bq	Mycenean	My
Old Church Slavonic	Cs	Oscan	Os
Estonian	Es	Old Phrygian	Ph; PhT, PhA
Etruscan	Et; EtB, EtT	Rhaetic	Rt; RtB, RtT, RtV
Finnic	Fi	Old Slovene	Sl
Greek	Gr	Umbrian	Um
Hittite	Hi	Venetic	Ve; VeB, VeT, VeV
Latin	La; LaC, LaS	Venezian	Vz
Luvian	Lu		

The PCA results in the previous study were obtained after normalization of input data. In present study the input data were normalized (PCA_n) as well as not normalized (PCA_nn). The dimensionless distances obtained are in both cases in fact main diagonals of the 10-dimensional squares.

The sum of absolute values of frequency differences was obtained in the following way:

$SuD = \sum abs(y_i - x_i)$

where y is the frequency of the i-th sound in the language in question, whereas x is the frequency of the i-th sound in another language, and i means in these cases one and the same sound or a sound pair or a triplet.

The root-of-sum-of-square frequency differences were obtained as follows:

$SuS = (\sum (y_i - x_i)^2)^{1/2}$

It is in fact the main diagonal of the 24-dimensional squares of frequency data under the assumption of orthogonality of data sets.

The mark Su is used to present the average value of SuD and SuS.

Besides them, also the regression quality indicators were used: The correlation coefficient R, the standard error of estimation STE, and the Fisher ratio F. STE increases with decreasing correlation of the two datasets, whereas R and F are decreasing. To indicate the distance are thus to be used STE, 1-R, and 1/F.

To normalize the obtained results, calculation of I_i/I_{max} was used, except for F, where F_{min}/F_i was used, and R, where $(1\text{-}R_i)/(1\text{-}R_{min})$ was used to obtain normalized results.

Average largest observed distance is the average value of largest observed distances from the languages EtB, EtT, LaC, LaS, PhT, PhA, RtB, RtT, RtV, Sl, VeB, VeT, resp. VeV as the origins, to the most distant of all tested languages.

The use of the normalized space is the distance between the closest non-same language to the last but one language determined after normalization of data. For example, when EtB is taken as the origin, the normalized distance to it is zero, to EtT it is 0.110, to RtT it is 0.303, to My it is 0.907, and to Cs it is 1. The closest non-same language is thus RtT, the last but one is My, and the use of normalized space is in this case equal to 0.604.

RESULTS

Results on frequency data of single sounds

The characteristics of the tested functions for the purpose of estimation of distance between languages based on their sound frequencies were tested on frequency data of single sounds. In Table 2. are presented the average data.

Table 2: Average largest observed distance and the use of normalized space

Method	Largest observed distance	Method	Use of normalized space
PCA_nn	8.038	PCA nn	0.740
PCA_n	1.394	F	0.693
SuD	0.671	R	0.659
1-R	0.388	PCA n	0.633
SuS	0.224	SuS	0.491
1/F	0.083	SuD	0.461
STE	0.035	STE	0.430

The largest distance values give the methods resp. functions PCA and SuD, whereas the best usage of the available normalized data space perform PCA, F, and R. In both cases STE gives rise to the lowest values.

In Table 3 are presented correlations between lumped results of tested methods.

Table 3: Correlations between lumped results of methods

	F	1/F	R	1-R	STE	**PCA nn**	**SuD**	**SuS**
F	1							
1/F	–0.270	1						
R	0.409	**–0.934**	1					
1-R	–0.409	**0.934**	–1	1				
STE	–0.610	0.706	–0.865	0.865	1			
PCA nn	–0.315	0.822	–0.836	0.836	0.723	1		
SuD	–0.591	0.750	–0.874	0.874	**0.905**	0.747	1	
SuS	–0.559	0.817	**–0.921**	**0.921**	**0.921**	0.868	**0.953**	1

The best correlation (> 0.9) is observed among SuD and SuS, 1/F and R resp. 1-R, R and SuS, as well STE to SuS and SuD.

The sequencies of increasing distance

Parameters of regression quality, R, F, STE

Among the parameters of regression quality, not only is the correlation coefficient between lumped data of 1/F and 1-R high, > 0.9. Also the sequence of languages of increasing distance (dissimilarity) is in all but one tested case the same:

F=R=STE: Sl

F=R≈STE: VeB, VeV; EtB, EtT, LaC, LaS, RtB, RtT, RtV

F=R≠STE: PhT, PhA

F≠R≠STE: VeT

STE, on the other hand, correlates to them much less and gives rise to the same sequence of languages only in one case whereas in 9 of 13 cases the sequence is not too much different. Thusly, STE is a different measure of language distance than 1/F and 1-R.

The sequence of increasing distances between the languages, based on averages of regression parameters is (the left-most language is that from which the distance is estimated):

Et << Rt < Fi < Es, Bq < Sl < Ph < Lu < La, Hi < Vz < Os < Um, Gr < Ve << My < Cs
La << Os < Um < Sl << Gr, Fi < Es, My < Bq < Rt < Ph < Ve < Vz < Et < Cs << Hi < Lu
Ph << Sl, Vz < Bq < Fi < Ve, Es, Gr < Rt < My < La, Cs < Et < Os << Lu < Um < Hi
Rt << Sl < Et, Fi, Os, Es < La, Ph < Lu < Bq < Hi < Ve, Cs < Um << My < Gr < Vz
Sl << Cs < La < Os < Ve, Ph, Gr < Fi < Rt < My, Es < Bq < Vz < Um < Et << Lu < Hi
Ve << Cs < Sl < Gr < Ph, My, Fi < Vz < La < Os < Rt, Bq < Um, Es, Et << Lu < Hi

Methods of geometric distance, PCA_nn, SuD, SuS

The sequence of increasing distances between the languages, based on PCA_nn results is:
Et << Es < Rt < Bq < Ph, Os, Fi < Sl < La < Um < Vz < My < Ve < Gr < Hi < Cs < Lu
La << Um < Sl, Os < Fi < Es < Bq, Vz, My < Gr, Ph < Ve < Cs < Et < Rt << Hi < Lu
Ph << Bq < Fi < Es, Sl < Os < La < Vz, Et, My < Um < Ve < Rt, Gr < Cs << Hi < Lu

Rt << Et < Es < Bq < Os, Ph < Fi < Sl < La < Um < Hi < Vz, My < Ve < Gr < Cs < Lu
Sl << Fi, La < Os < Um < Vz, My, Ph < Es, Ve, Bq, Gr < Cs < Et < Rt << Hi < Lu
Ve << Cs < Gr < My < Vz < Sl < La, Fi, Um < Os, Ph < Bq < Es < Et < Rt << Hi < Lu

The sequence of increasing distances between the languages, based on averages of SuD and SuS is:
Et << Rt < Es < Sl < Fi < La < Vz < Bq < Os < Ve < Ph < Gr < Um < Cs < Hi < My < Lu
La << Os < Sl < Um < Gr < Es < Fi < Bq < Ve < Ph < Vz < My < Rt < Et < Cs < Hi < Lu
Ph << Sl < Vz < Ve < Es < Fi < Gr < Bq < Cs < La < Os < Rt < My < Et < Um < Hi < Lu
Sl << Cs < La < Os < Ve < Gr < Ph < Vz < Es < Fi < Rt < Et < Bq < Um < My < Hi < Lu
Rt << Sl < Et < Os < Es < Fi < La < Ph < Ve < Bq < Cs < Um < Gr < Vz < Lu < My < Hi
Ve << Cs < Sl < Gr < Ph < Vz < Fi < La < Os < Rt < My < Et < Bq < Es < Um < Hi < Lu

Looking only for the most important reference languages,
R&F: **Et** < Sl < La < Gr; **Ph** < Sl < Gr < La; **Rt** < Sl < La < Gr; **Ve** < Cs < Sl < Gr < La
PCA: **Et** < Sl < La < Gr; **Ph** < Sl < La < Gr; **Rt** < Sl < La < Gr; **Ve** < Cs < Gr < Sl < La
Su: **Et** < Sl < La < Gr; **Ph** < Sl < Gr < La; **Rt** < Sl < La < Gr; **Ve** < Cs < Sl < Gr < La

As can be seen from the above presentations of sequences of increasing distances between the languages, different methods give rise to different sequences and different values of distances between languages. Some characteristics are, however, common to them. For example, **Et** < Sl < La < Gr; **Ph** < Sl < Gr, La; **Rt** < Sl < La < Gr; **Ve** < Cs, Sl, Gr < La. Ancient Anatolic languages Hi and Lu are in most cases among the most distant ones.

Results on frequency data of sound pairs and triplets

The PCA method gave in a previous trial [1] suspect results. This seems to be the consequence of software limitations in addressing the memory. Here we test the other methods mentioned above.

Sound pairs
Methods R, STE
Et << Rt < Es < Sl < Bq < Fi < Vz < La < Um < Ph < Lu < Ve < Os < Gr < Hi < Cs < My
La << Um < Bq < Os < Sl < Vz < Rt < Es < My < Gr < Fi < Et < Ph < Ve < Lu < Cs < Hi
Ph << Gr < Ve < Es < Fi < Sl < Vz < Bq < Et < My < Rt < La < Cs < Os < Lu < Um < Hi
Rt << Et < Es < Sl < La < Fi < Lu < Ph < Cs < Os < Bq < Um < Hi < Ve < Vz < My < Gr
Sl << Cs < Rt < La < Vz < Es < Et < Fi < Ph < Bq < Um < Ve < Gr < Os < My < Lu < Hi
Ve << Gr < Ph < My < Vz < Sl < Fi < Cs < La < Et < Rt < Bq < Os < Es < Um < Lu < Hi
Methods SuD, SuS
Et << Rt < Es < Sl < La < Fi < Vz < Um < Bq < Os < Ph < Ve < Gr < Cs < Hi < My < Lu
La << Os < Um < Sl < Es < Vz < Gr < Bq < Fi < Rt < My < Et < Ph < Ve < Cs < Hi < Lu
Ph << Gr < Ve < Es < Fi < Vz < Sl < My < Bq < La < Rt < Os < Et < Um < Cs < Hi < Lu
Rt << Et < Es < Fi < Sl < La < Os < Ph < Um < Ve < Bq < Vz < Hi < Gr < Cs < My < Lu
Sl << Cs < Vz < Es < La < Fi < Rt < Os < Et < Ph < Gr < Bq < Ve < Um < My < Hi < Lu
Ve << Gr < Ph < Vz < Fi < Sl < La < My < Os < Rt < Et < Cs < Es < Um < Bq < Hi < Lu

Sound triplets

Methods R, STE

Et << La < Rt < Sl < Es < Cs < Gr < Bq < Vz < Fi < Um < Lu < Ve < Ph < My < Os < Hi
La << Vz < Gr < Um < Bq < Et < Es < Sl < Os < Fi < Cs < My < Rt < Ph < Lu < Ve < Hi
Ph << Gr < Cs < La < Es < Fi < Et < My < Vz < Bq < Sl < Ve < Lu < Um < Rt < Hi < Os
Rt << Et < La < Cs < Es < Sl < Fi < Gr < Vz < Bq < Um < Lu < Ve < My < Ph < Hi < Os
Sl << Cs < La < Vz < Et < Gr < Es < Fi < Rt < Bq < Um < My < Lu < Ve < Ph < Os < Hi
Ve << Cs < Gr < La < Et < Vz < My < Es < Fi < Sl < Bq < Um < Ph < Rt < Lu < Hi < Os

Methods SuD, SuS

Et << La < Es < Rt < Fi < Gr < Sl < Vz < Um < Cs < Bq < Ph < My < Os < Ve < Hi < Lu
La << Vz < Gr < Es < Et < Fi < Um < Sl < Bq < My < Rt < Os < Cs < Ph < Hi < Ve < Lu
Ph << Gr < Fi < Es < My < La < Rt < Et < Vz < Sl < Cs < Bq < Ve < Um < Os < Hi < Lu
Rt << Et < Es < La < Fi < Sl < Gr < Ph < Um < Cs < Vz < My < Bq < Os < Hi < Ve < Lu
Sl << Cs < Es < Gr < Vz < Fi < La < Et < Um < My < Os < Rt < Bq < Ph < Hi < Ve < Lu
Ve << Gr < Fi < La < Ph < Vz < Et < Es < My < Cs < Sl < Rt < Um < Bq < Os < Hi < Lu

Putting all together and looking only for the most important reference languages, we have the situation presented in Table 4.

Table 4: Sequences of increasing average distance of some reference languages from the language in question

Sounds		**Etruscan**	**Old Phrygian**	**Rhaetic**	**Venetic**
Single	PCA	**Et**<Rt<Sl<La<Gr	**Ph**<Sl<La<Gr	**Rt**<Et<Sl<La<Gr	**Ve**<Cs<Gr<Sl<La
	F,R,STE	**Et**<Rt<Sl<La<Gr	**Ph**<Sl<Gr<La	**Rt**<Sl<Et <La<Gr	**Ve**<Cs<Sl<Gr<La
	SuD(S)	**Et**<Rt<Sl<La<Gr	**Ph**<Sl<Gr<La	**Rt**<Sl<Et<La<Gr	**Ve**<Cs<Sl<Gr<La
Pairs	R, STE	**Et**<Rt<Sl<La<Gr	**Ph**<Gr<Sl<La	**Rt**<Et<Sl<La<Gr	**Ve**<Gr<Sl<Cs<La
	SuD(S)	**Et**<Rt<Sl<La<Gr	**Ph**<Gr<Sl<La	**Rt**<Et<Sl<La<Gr	**Ve**<Gr<Sl<La<Cs
Triplets	R, STE	**Et**<La<Rt<Sl<Gr	**Ph**<Gr<La<Sl	**Rt**<Et<La<Sl<Gr	**Ve**<Cs<Gr<La<Sl
	SuD(S)	**Et**<La<Rt<Gr<Sl	**Ph**<Gr<La<Sl	**Rt**<Et<La<Sl<Gr	**Ve**<Gr<La<Cs<Sl

Table 5: Sequences of increasing distance of some reference languages from the Etruscan and Rhaetic

Sounds	**Method**	**Etruscan**	**Rhaetic**
Single	PCA	**EtT**<RtT<LaC<Sl<Gr	**RtT**<EtT<LaC<Sl<Gr
	F, R	**EtT**<RtT<Sl<LaC<Gr	**RtT**<EtT<Sl<LaC<Gr
	STE	**EtT**<RtT<Sl<Gr<LaC	**RtT**<EtT<Sl<LaC<Gr
	SuD	**EtT**<RtT<Sl<LaC<Gr	**RtT**<Sl<EtT<LaC<Gr
	SuS	**EtT**<RtT<Sl<LaC<Gr	**RtT**<EtT<Sl<LaC<Gr
Pairs	R	**EtT**<RtT<Sl<LaC<Gr	**RtT**<EtT<LaC<Sl<Gr
	STE	**EtT**<RtT<Sl<LaC<Gr	**RtT**<EtT<Sl<LaC<Gr
	SuD	**EtT**<RtT<LaC<Sl<Gr	**RtT**<EtT<LaC<Sl<Gr
	SuS	**EtT**<RtT<Sl<LaC<Gr	**RtT**<EtT<Sl<LaC<Gr
Triplets	R	**EtT**<RtT<LaC<Sl<Gr	**RtT**<EtT<LaC<Sl<Gr
	STE	**EtT**<LaC<Gr<Sl<RtT	**RtT**<LaC<EtT<Gr<Sl
	SuD	**EtT**<RtT<LaC<Gr<Sl	**RtT**<EtT<LaC<Sl<Gr
	SuS	**EtT**<LaC<RtT<Gr<Sl	**RtT**<EtT<LaC<Sl<Gr

Table 6: Sequences of increasing distance of some reference languages from the Old Phrygian and Venetic

Sounds	Method	Old Phrygian	Venetic
Single	PCA	**PhT**<Sl<LaC<Gr	**VeT**<Cs<Gr<Sl<LaC
	F, R	**PhT**<Sl<Gr<LaC	**VeT**<Gr<Cs<Sl<LaC
	STE	**PhT**<Sl<Gr<LaC	**VeT**<Cs<Gr<Sl<LaC
	SuD	**PhT**<Sl<LaC<Gr	**VeT**<Gr<Sl<Cs<LaC
	SuS	**PhT**<Sl<Gr<LaC	**VeT**<Cs<Sl<Gr<LaC
Pairs	R	**PhT**<Gr<Sl<LaC	**VeT**<Gr<LaC<Sl<Cs
	STE	**PhT**<Gr<Sl<LaC	**VeT**<Gr<Cs<Sl<LaC
	SuD	**PhT**<Gr<LaC<Sl	**VeT**<Gr<LaC<Cs<Sl
	SuS	**PhT**<Gr<Sl<LaC	**VeT**<Gr<Sl<LaC<Cs
Triplets	R	**PhT**<Gr<LaC<Sl	**VeT**<Gr<LaC<Cs<Sl
	STE	**PhT**<LaC<Gr<Sl	**VeT**<Cs<LaC<Gr<Sl
	SuD	**PhT**<Gr<LaC<Sl	**VeT**<Gr<LaC<Sl<Cs
	SuS	**PhT**<Gr<LaC<Sl	**VeT**<Gr<LaC<Cs<Sl

Taking into account only the original (i.e. non-Slavic) reading of target languages, we have the situation presented in Table 5 and 6. It is not very different from that in Table 4. Taking into account the readings based on Slovene, then Old Slovene appears in about 19 % of cases somewhat closer to the target language than by the non-Slavic reading. But in about 5 % of cases it appears somewhat more distant. Thusly, the conclusions that can be derived from Tables 4-6 regarding the distances between languages based on their sound frequencies, are generally valid.

Discussion

It should not be forgotten that whereas the functions SuD, SuS, R and F are symmetric functions, the STE function is asymmetric and caution is to be taken that there are compared only the results where the data of a language in question are in the proper and always the same position in the equation.

Comparing the frequencies of sounds and especially their pairs and triplets, there can easily emerge another source of error using the STE function. Namely, when a sound or a pair or a triplet is not observed in a text, it is usually noted that its frequency is zero. If both languages, which are compared, have zero frequency of any sound or a pair or a triplet, this does not mean that they are equal in this respect, but only that this information is lacking in both of them.

As can be seen in Table 3, tested methods do not give rise to highly correlating results. The highest are on the one hand between the dimensionless geometric distance representing PCA and SuS, and on the other hand between the regression qualities parameters R and 1/F. Different methods obviously give more weight to different details in the studied frequencies dataset.

Among the tested methods of geometric distance, only PCA uses results obtained after distribution of variance contained in the input dataset to the orthogonal coordinate

axes. Thus it enables the calculation of true main diagonals of multidimensional squares. In other methods of this type the orthogonality is only supposed; in other words, the main diagonals are calculated as if the orthogonality of data would be granted. Therefore, neither good correlation of results with those of PCA nor good similarity of sequencies of languages is to be expected. Table 3 and the sequence data confirm this expectation. This is a general result. In spite of it, some general conclusions can be derived also using those other methods.

An interesting characteristic of the sum-of-frequency-differences methods is that the separation between different ways of reading of the same texts increases with the increasing exponent.

The results using sound pairs and triplets are consistent among the languages, which are written in the same way, e.g. using equal rules to write the words separated from one another. Among the languages, which are written *in continuo* and with no fixed word separation rules, there may be counted, depending on the choice of division of the continuous text into words, also too few or too many doublets resp. triplets; cf the counts in Table 1 in ref. [1]. So the results based on counting sign doublets and triplets must be expected to be less plausible than those based on counting single signs. Anyway, they may give some useful indications.

The comparison shows that to the ancient languages: Etruscan, Rhaetic, and Old Phrygian, by the majority of tested methods Old Slovene is closer than Latin or Greek. To Venetic, in several cases the Old Church Slavonic or Homeric Greek is closer than Old Slovene. Another general trend is observed that going from frequencies of single sounds to pairs or triplets of them, the Old Slovene appears to be more and more distant. What is the reason of this effect is not possible yet to conclude in a definite manner. One possible reason could be the differences in word division.

Having all this in mind, the use of Slovene language, with its almost 50, some quite archaic dialects, is a legitimate template for deciphering of some ancient languages.

The closeness of Etruscan and Rhaetic, observed on language distance estimations based on sound frequencies, Tables 4-6, indicates that on the Apennine peninsula there existed during the Bronze Age a clinal cultural and linguistic continuum formed during or after the neolithisation of the peninsula (e.g. the Golasecca culture, the Villanova culture, etc). Some ancestors of the so-called Italic peoples (e.g. Latini, Veneti, etc) seem, according to their mythology, to be late Bronze Age or early Iron Age military intruders into this peninsula, possibly as the rests of the troops of the so-called "Peoples from beyond the Sea" after their unsuccessful attacks on Egypt.

Conclusion

Six methods were tested for estimation of linguistic distance of 17 languages. With different ways of reading of some ancient languages altogether 24 language variants were tested. Different methods give different numerical results. Therefore, their sorted normalized values are to be compared. The comparison shows that to the ancient languages Etruscan,

Old Phrygian and Rhaetic, by tested methods Old Slovene is closer than Latin or Greek. Thus, the use of Slovene language, with its almost 50, some quite archaic dialects, is a legitimate template for deciphering of some ancient languages.

The closeness of Etruscan and Rhaetic, observed on language distance estimations based on sound frequencies indicates that on the Apennine peninsula there existed during the Bronze Age a clinal cultural and linguistic continuum formed during the neolithisation of the peninsula. Some ancestors of the so-called Italic people (e.g. Latini, Veneti, etc.) may have been late Bronze Age or early Iron Age military intruders into this area.

References

1. A. Perdih, G. Tomezzoli, V. Vodopivec, *Comparison of contemporary and ancient languages*, Zbornik šeste mednarodne konference, Izvor Evropejcev, Jutro, Ljubljana **2008**, 40-87, http://www.korenine.si/zborniki/zbornik08/comparison_languages.pdf.
2. M. Silvestri, G. Tomezzoli, *Linguistic Computational Analysis to measure the distances between ancient Venetic, Latin and Slovenian Languages*, Proceedings of the Third International Topical Conference, Ancient Settlers of Europe, Jutro, Ljubljana **2005**, 77-85, http://www.korenine.si/zborniki/zbornik05/tomezzoli_venslolat.htm
3. M. Silvestri, G. Tomezzoli, *Linguistic distances between Rhaetian, Venetic, Latin and Slovenian languages*, Proceedings of the Fifth International Topical Conference, Origin of Europeans, Jutro, Ljubljana **2007**, 184-190, http://www.korenine.si/zborniki/zbornik07/tomezzoli_dist07.pdf

Abstract

Six methods for estimating the linguistic distances were tested. These methods were applied to seventeen languages with eleven variant readings. The different methods give rise to dissimilar quantitative distance values. Therefore their normalized distance values were useful in comparing the degrees of similarities and differences between languages. The ancient languages addressed were: Etruscan, Old Phrygian, Rhaetic and Venetic. These languages were tested against better known models of Latin, Greek, and Old Slovene. Old Slovene proved in most cases to be a closer match to those extinct languages than the antique Greek and Latin. Slovene language therefore is a legitimate template for deciphering some ancient languages.