Anton Perdih, Giancarlo Tomezzoli, Vinko Vodopivec

# COMPARISON OF CONTEMPORARY AND ANCIENT LANGUAGES

**Key words:** multidimensional analysis, sound frequencies, linguistic distances, ancient languages, Venetic, Rhaetic, Old Phrygian, Old Slovenian, Old Church Slavonic, Etruscan, Latin, Venetian, Greek, Basque, Estonian, Finnic, Hittite, Luvian, Mycenean, Oscan, Umbrian.

**Ključne besede:** večdimenzionalna analiza, pogostost glasov, jezikovne razdalje, stari jeziki, venetski, retijski, frigijski, staroslovenski, staro cerkveno slovanski, etruščanski, latinski, beneški, grški, baskovski, estonski, finski, hetitski, luvijski, mikenski, oskijski, umbrijski.

## Abstract

Determining the agreement in grammatical structure and in the language material that bears the structure in some ancient languages is questionable. Short and damaged inscriptions which are written in continuous manner, in dialects and with many abbreviations are always subject to potential error in exact translation. This is the case among the Venetic, Rhaetic, and Phrygian inscriptions, where it is useful at the moment to only focus on the comparison of sound frequencies.

Unidimensional as well as multidimensional analyses of sound frequencies in 16 languages, mostly ancient, where in some of them the division of the continuous text into words is still questionable, support the previous observation that Venetic and Rhaetic are by sound frequencies closer to Old Slovenian than to Old Italic languages (Latin, Oscan, Umbrian). Close to Venetic and Rhaetic are in these characteristics also Old Phrygian and Etruscan. Interesting is (by this criterion) also the closeness of Estonian resp. Finnic to most of these languages. Latin, Oscan, and Umbrian form a different cluster than the Etruscan, Rhaetic and Venetic. Whereas Etruscan is close to Rhaetic, Old Slovenian, Venetic, etc, it is not close to Hittite and Luvian, from which it is sometimes supposed to derive. Present Venetian dialect is by the sound frequencies closer to Old Slovenian than to Latin. This indicates that the sound frequencies are very resistant to phonemic changes.

Analyses of frequencies of sounds and their combinations in various languages give thus results, which contribute additional light into knowledge of them. They contribute it from a different and independent point of view than the agreement in grammatical structure and in the language material that bears the structure.

## Introduction

Many computational techniques were used in the past for calculating the linguistic distances between languages, dialects or variants in same or different family languages.

Nerbonne [1-3], Kessler [4], Heeringa [5] were successful in measuring Dutch and Irish dialects distances, in which the phonetics and the meaning of the words were known. The Levenshtein distance technique presented by Kruskal [6] and used by many other authors is also extremely reliable in the calculation of the phonetic distance when applied to Corpuses of well-known words having well known phonetics and grammar rules.

The problem of some ancient languages, like e.g. the Venetic, Rhaetic etc. is that a large Corpus of words cannot be accessed. Even more important is the fact that the exact pronunciation rules are not definitely known, some of them being only supposed. The subdivision of the inscriptions written *in continuo* into words, their understanding, the exact meaning of the punctuation system, and the precise grammar rules are practically unknown. Additionally, the linguistic evolution is unknown.

Thus, for some of the ancient languages it is not possible to use the usual techniques: the Levenshtein distance, the frequency of phonetic features, the frequency per word, the Manhattan distance or hybrid techniques used by Vieregge et al. [7], for evaluating the linguistic distances between these and other ancient or present languages.

For these reasons, in previous contributions [8, 9] a much more simple and direct one-dimensional method for evaluation of said linguistic distances was applied. After putting together several additional language databases, we have the chance in our search to apply in their evaluation not only the simplest one- or two-dimensional techniques but also a multidimensional one.

## Impetus

The motivation to start a search about the linguistic distances between ancient languages was provided by the debate about the origin of the Venetic language started recently by many authors such as Pellegrini and Prosdocimi [10], Marinetti [11], Lejeune [12], Šavli, Bor and Tomažič [13]. All these authors agreed that the Venetic language is an Indo-European (IE) language, but they disagree about the fundamental question of the linguistic distance of the Venetic with respect to the Latin and the Slovenian language.

On one side, Lejeune [12] affirmed that: "This language (the Venetic) is "italic" and, …, closer to Latin than to any other language". On the other side, Bor [13] affirmed that: "I was unable to find a single (Venetic) inscription that could not be deciphered on the basis of the Slavic languages and the surviving Slovenian dialects, above all the Slovenian archaisms" and Šavli and Tomažič [13] agreed that the Venetic is closer to the Slovenian.

The problems in interpreting the Venetic consists in the relatively small number of inscriptions (about 400) which are in categorically short, broken or incomplete, making the composition of an extended and comprehensive linguistic Corpus difficult. In addition, the majority of the Venetic inscriptions are written *in continuo*, i.e. without separation in words, and are mainly of funerary or votive content, so that they do not give us any suitable clue about Venetic toponyms, verbs, and frequently used words that could be used for computational comparisons between Venetic and other languages.

The punctuation rules of the Venetic, provided by Lejeune [12] and Vetter [14] are far from indicating clear word separations. Moreover, a further problem facing the use of computational techniques for comparing Venetic with other languages is the contemporary ignorance of possible pronunciation rules. Another challenge is the use of abbreviations in ancient languages, so that without understanding such abbreviations any division of continuous texts is problematic. The problem arises because these texts are written in unknown dialects. So the derivation of any grammatical rules would be questionable. Therefore, any attempt of classifying the Venetic by using phonetic symbolic techniques would be problematic.

On one side, using the Lejeune [12] and Vetter [14] punctuation rules and possible similarities between Venetic and Latin, Pellegrini, Prosdocimi [10] and Marinetti [11] provided translations of a great number of the Venetic inscriptions. However, as clearly visible in their works, the translation in the majority of the cases is more an extrapolation of the possible meaning of the inscriptions than a clear translation.

On the other side, Vodopivec [15] made a remarkable comparison between Venetic, Latin, Slovenian, as well as other languages: Croatian, English, German, French, Italian, Greek. By considering different Venetic roots: vrv, trt, krk, …, grg; prap, …, prup, observed in the Venetic alphabetic tablets Es23 – Es26, he found that such roots exist mainly in Slovenian and Croatian which are Slavic languages, and to a much lesser degree in non-Slavic languages. The publication [13], pp. 185-443, Engl. ed. pp. 172-340, as well as [16-21] represent an exhaustive list of works presenting the results of use of Slavic, especially Slovenian, as a catalyst for understanding of Venetic.

## Linguistic Distance

For measuring the linguistic distances between Venetic, Latin and Slovenian [8], three electronic databases were developed:

• The Latin Language Database (LLD) comprising the works of the following authors: Plautus (250 – 184 BC) - Stichus, Cato (234 – 149 BC) - De Agri Cultura, Terence (195/185 – 159 BC - Hecyra, Cicero (106 – 43 BC) – Catilinariae I - IV, Caesar (100 – 44 BC) - De Bello Gallico I - VIII, Vergil (70 – 19 BC) - Aeneids I - XII, Propertius (50 – 16 BC) - Elegiae I - IV. All these Latin authors were active in the period 300 ~ 1 BC, a period in which the Latin and the Venetic languages were spoken almost independently. The texts of said authors were acquired from the Internet site [22].

• The Slovenian Language Database (SLD) comprising the texts of the most ancient available Slovenian manuscripts: the Brižinski Spomeniki or Freisinger Denkmäler I-III (972 – 1093 AD), the Rateški Rokopis or Ratetischer Handschrift (1362 – 1390 AD), the Stiški Rokopis or Sitticher Handschrift (1428 – 1440 AD), the Starogorski Rokopis or Handschrift von Castelmonte (1450 – 1520 AD). Although the spoken Venetic is of much greater antiquity than the Slovenian manuscripts from the $10^{th}$, $11^{th}$, $14^{th}$ and $15^{th}$ Centuries, these (Slovenian) writings serve well as templates for linguistic comparisons. No written text has been found in Slovenian or Slovenian dialects earlier than these texts. The texts of the manuscripts were acquired from the Internet sites [23-26].

• The Venetic Language Database (VLD), which comprises all the Venetic inscriptions in the works of Pellegrini, Prosdocimi [10] and Marinetti [11] and the Internet sites [27-30].

The evaluation of the Pythagorean Linguistic Distance in the limit of the above mentioned databases shows [8] that the Venetic is linguistically closer to Slovenian than to Latin.

In the attempt of improving the knowledge about the linguistic distances between ancient languages, the Rhaetian Language Database (RLD) containing all the Rhaetian inscriptions published and revised by S. Schumacher [31] was prepared [9]. The evaluation of the Pythagorean Linguistic Distance using the above-mentioned databases has shown [9] that the Rhaetian is very close to Venetic. Thus geographic proximity is consistant with linguistic distance even if the chronological distance is measured in millenia. The publications [13], pp. 397-408, [32], pp. 61-70, [33-40] represent an exhaustive list of works about the Rhaetian / Slavic problematics.

## The linguistic databases

In view of a linguistic multidimensional analysis further linguistic databases (LDs; in the text, the designations LD or database or the specific labels given below are used for the same purpose, as more appropriate) have been prepared. For each database, two electronic versions were prepared: a basic version containing inscriptions or texts together with information and explanations, and a working database containing only the inscriptions or the texts according to general and specific conversion rules (see later) suitable for electronic computations.

- The Basque Language Database (BqLD) containing the *San Benoaten Bizitzea* text from [41]. The working BqLD was prepared according to general conversion rules and the specific conversion rules derived from [42].
- The Estonian Language Database (EsLD) containing the *Kalevipoeg* text from [43]. The working EsLD was prepared according to the general conversion rules and the specific conversion rules indicated by M. Smolej.
- The Etruscan Language Database (EtLD) containing the Etruscan inscriptions from Pallottino [44]. In order to take into account the pronunciation rules by Pallottino [44] and by Bor [32], pp. 11-60, two working EtLDs were prepared: the EtTLD and the EtBLD. The publications [13], pp. 342-396, [32], pp. 11-60, [45-47] represent an exhaustive list of works about the Etruscan / Slavic relationship.
- The Finnic Language Database (FiLD) containing the *Kalevala* text from [48] which contains a scanned version of [49]. The working FiLD was prepared according to the general conversion rules and the specific conversion rules indicated by M. Smolej.
- The Greek Language Database (GrLD) containing the Homer's *Iliad* text, books 1 to 5 from [50]. The working GrLD was prepared by conversion of Greek letters into Latinic ones.
- The Hittitic Language Database (HiLD) containing Hittitic texts from [51,52]. The working HiLD was prepared by removing Sumeric and Akkadic words and by applying the general conversion rules and the specific conversion rules according to generally accepted sound values including interpretation of intervocalic consonants [53,54]

- The Latin Language Database (LaLD) containing the texts from the LLD used in the past contributions [8,9]. In order to take into account the Classical and Semiclassical pronunciation rules, two working LaLDs were prepared: the LaCLD and the LaSLD, both according to the general conversion rules and respectively according to the Classical and Semiclassical pronunciation rules derived from [55,56].
- The Luvian Language Database (LuLD) containing the text from [57]. The working LuLD was prepared by eliminating the Hittite and Palaic parts of texts and by using the syllabic transcription from cuneiform to Latinic for obtaining the alphabetic writing on which the general conversion rules were applied [54]. Applying Hittite conversion rules, a parallel database LuHLD was prepared.
- The Mycenean Language Database (MyLD) containing the text of few tablets and the Glossary from [58]. The working MyLD was prepared according to the general conversion rules.
- The Old Church Slavonic Language Database (CsLD) containing the *Codex Suprasliensis*, taken from [59,60]. The working CsLD was prepared by eliminating recognised loanwords and foreign names, by transliterating the Cyrillic writing into the Latinic one on which the general conversion rules and the specific conversion rules from [61] were applied.
- The Oscan Language Database (OsLD) containing the *Cippus Abellanus* and *Tabula Bantina* texts from [62]. The working OsLD was prepared on the basis of the Oscan writing and language information acquired from [63-66].
- The Old Phrygian Language Database (PhLD) containing the texts from [67]. Two working PhLD were prepared: PhLD according to instruction in [67], and PhALD according to instruction in [68].
- The Rhaetic Language Database (RtLD) containing the Rhaetian inscriptions from the RLD used in the past contribution [9]. In order to take into account the incertitude in reading some of the characters, three working RtLD were prepared: the RtTLD using the reading in [31], the RtPLD using the reading in [32], pp. 11-20, and the RtVLD using the reading in [33].
- The Old Slovenian Language Database (SlLD) containing the *Brižinski spomeniki* [69] and Slovenian texts from [70]. The working SlLD was prepared by transliterating the texts into modern Slovenian notation, taking into account the diplomatic, critical and phonetic transcription and the translation into modern Slovenian by following the original text as much as possible.
- The Umbrian Language Database (UmLD) containing the *Tables of Iguvium*, taken from TITUS Text collection: Inscr.OU, Oscan and Umbrian Inscriptions [71]. The working UmLD was prepared according to the transliteration in [72] on which the general conversion rules and the specific conversion rules from [72] were applied.
- The Venetic Language Database (VeLD) containing the Venetic inscriptions from the VLD used in the past contributions [8,9]. In order to take into account the incertitude in reading some of the characters three working VeLD were prepared: the VePLD according to [13], the VeTLD according to [10] and the VeVLD according to [21] where repeating parts of texts on Atestine tablets are eliminated.

– The Venetian Language Database (VzLD) containing the C. Goldoni's commedies in Venetian language: *I Rusteghi; Le Baruffe Chiozzotte; Sior Todero Brontolon; Il Campiello; La Casa Nova; Una delle Ultime Sere di Carnevale; Il Gondoliere Veneziano; Gli Sdegni Amorosi*, all from [73]. The working VzLD was prepared according to the general conversion rules and by eliminating italianisms. In order to avoid the mistake for Venetic, in the text will not be written Venetian but Venezian.

For reaching a uniform linguistic database representation, all the working language databases were converted according to the rules of the Slovenian literary notation [74]. The sentences, when recognised by the use of dots, commas, etc. were placed in separate lines. Then, all said signs, the brackets, etc. were removed. The capital letters were replaced by lowercase ones. The signs indicating missing or incertain characters were retained but not counted.

There was made no distinction between open, closed, long, short, stressed and non-stressed vowels; they are grouped together and presented by one corresponding vowel

**Table 1.** Number of countable alphabetic characters, their pairs and triplets in the LDs.

| Language | LD | No. of countable characters | | |
|---|---|---|---|---|
| | | single | pairs | triplets |
| Basque | Bq | 160,177 | 130,866 | 101,577 |
| Old Church Slavonic | Cs | 458,319 | 362,444 | 278,990 |
| Estonian | Es | 90,742 | 76,108 | 61,485 |
| Etruscan [44] | EtT | 30,421 | 24,227 | 18,445 |
| Etruscan [32] | EtB | 30,421 | 24,227 | 18,445 |
| Finnic | Fi | 449,075 | 381,686 | 314,298 |
| Greek | Gr | 117,109 | 93,503 | 71,502 |
| Hittite | Hi | 14,001 | 11,509 | 9,025 |
| Latin Classic | LaC | 1,029,312 | 848,168 | 667,718 |
| Latin Semiclassic | LaS | 1,019,977 | 838,833 | 658,383 |
| Luvian | Lu | 32,626 | 27,254 | 21,942 |
| | LuH | 33,843 | 28,471 | 23,159 |
| Mycenean | My | 26,330 | 22,474 | 18,618 |
| Oscan | Os | 3,057 | 2,418 | 1,841 |
| Old Phrygian [67] | Ph | 2,242 | 1,834 | 1,459 |
| Old Phrygian [68] | PhA | 2,290 | 1,698 | 1,172 |
| Rhaetic [31] | RtP | 2,102 | 1,719 | 1,394 |
| Rhaetic [32] | RtT | 1,948 | 1,572 | 1,265 |
| Rhaetic [33] | RtV | 2,097 | 1,754 | 1,440 |
| Old Slovenian | Sl | 19,834 | 15,428 | 11,301 |
| Umbrian | Um | 25,063 | 20,657 | 16,288 |
| Venetic [13] | VeP | 7,651 | 6,083 | 4,965 |
| Venetic [10] | VeT | 7,427 | 6,119 | 4,843 |
| Venetic [21] | VeV | 7,113 | 4,855 | 2,993 |
| Venezian | Vz | 320,794 | 234,563 | 153,903 |

character to give a five-vowels notation system. Among the consonants, the affricate are denoted by the C or Č sign. To denote the fricatives, the signs F, S, Z, Š, Ž, H are used. Plosives are notated as B, P; T, D; K, G. Laterals are noted by L, rhotics are notated by R, nasals by M or N. Summarising, all the LDs are prepared using a common notation of the 24 alphabetic characters of the Slovenian (a b c č d e f g h i k l m n o p r s š t u v z ž).

The specific conversions rules for vowels and consonants of each LD are collected in the file Rules-09.doc [75].

The survey of EtLD, PhLD, RtLD, VeLD shows that the number of uncertain characters in said databases is less than 10% of the total number of characters (EtLD 8.6%, PhLD 5.5%, RtLD 1.2%, VeLD 3.6%). The results of the survey are in the file Rules-09.doc [75] as well. Uncertain signs were not taken into any counting.

Table 1 presents some characteristics of the LDs, i.e.: the number of counted characters, pairs and triplets of them. In all respects the smallest databases are the Rhaetic and Old Phrygian, followed by Venetic and Oscan, whereas the largest are the Latin databases.

# Methods

## Counting

Counted were the number of alphabetic characters, of pairs and of triplets of characters, as well as some last characters in the words and some last pairs of them. Spaces and markers of missing or unreadable signs were not taken into any counting. From these data their respective frequencies were calculated using MS Excel.

## Principal Component Analysis

To draw relevant conclusions more easily, all frequencies were evaluated also using the Principal Component Analysis (PCA).

PCA is a multivariate method used for displaying data in cases where each sample (object) is described using several parameters (variables). In such cases, it is hard to extract the relevant information from the dataset (typically, a table) by investigating one variable at a time. Furthermore, in most cases the „independent“ variables, which we measure, are not really independent. They usually correlate at least partially to each other, which makes interpretation even harder. Graphical presentation of such datasets is also impossible, because we can display only two- or three-dimensional graphs. The PCA method enables us to present the information contained in the datasets using a small number of graphs. These graphs show us similarities and dissimilarities between objects and variables. Similar objects are grouped together, while dissimilar ones are scattered around. The same is true for variables. The graphs where the grouping of objects is presented are called score plots; while the graphs, which present the grouping of variables are called loading plots. From the patterns on the score plots and loading plots one can extract the information contained in the analysed dataset.

From the mathematical point of view, the PCA method is a rotation of the old co-ordinate system of variables. The co-ordinate system is rotated in such a way that the relevant information - i.e., the largest portion of the variance in the dataset - is presented using only a few variables of the new co-ordinate system. The new variables are called latent variables or principal components. The other latent variables of the new co-ordinate system represent noise – noise due to defects, in our case, in: inscriptions and texts, their transcription and transliteration in preparing the working LD for computation. The principal components are truly independent variables, i.e., they are orthogonal, which means that they do not correlate with each other. Score plots represent objects in the space defined by the principal components, while the loading plots represent the old (measured) variables in the space of principal components.

The PCA method is usually performed in three steps. In the first step, the dataset variables are normalised to variance 1 and the correlation matrix is calculated. The correlation matrix shows how the variables from the dataset correlate to each other. In the second step, eigenvalues and eigenvectors of the matrix are calculated, i.e. the matrix is diagonalised. Eigenvectors are the principal cmponents of the new coordinate system while the eigenvalues show the information content (relevance) of each principal component. In the third step the co-ordinates of samples (objects) in the new co-ordinate system are calculated. More detailed description about the method can be found in Wold et al. [76], Massart et al. [77], Brereton [78], and Graham [79]. The method is sufficiently simple for one to program it by oneself, as was done in our case.

At the end one can import the new calculated coordinates of the objects and eigenvectors into one of the spreadsheet programs available on the market to create score plots and loading plots.

Due to the normalisation of the variables to variance 1, the latent variables are dimensionless. The other consequence of the normalisation is that total variance of the dataset becomes equal to the number of variables.

## Distances

The PC axes are by definition orthogonal to each other. Thus the results of PCA are useful to estimate relative (dimensionless) distances between the objects or latent variables.

The dimensionless distance of a Language Database A from the centre of the PC space is:

$D(A) = (\sum(v(i)\times PC(A,i))^2)^{1/2}$

The distance between Language Database A and B is:

$d(A, B) = (\sum(v(i)\times(PC(A,i) - PC(B,i)))^2)^{1/2}$

where v(i) is the variance contained in the PC(i), whereas PC(A,i) is the coordinate of the Language Database A on the PC axis PC(i).

# Results

The results of comparing the frequencies of particular characters in the databases prepared for counting can be presented in different ways.

## Unidimensional approaches

### *Frequency of particular characters*

Ten most frequent characters in the LDs prepared as presented above are given in Tables 2-4. In all tested cases a vowel is the most frequent; in ten cases A, in four cases E and in eleven cases I. In two cases (Hittite and Umbrian) one vowel is more frequent than the most frequent consonant, in eleven cases there are three vowels more frequent than the most frequent consonant, in ten cases four vowels are more frequent than the most frequent consonant, and in two cases (Mycenean and Old Slovenian) all five vowels are more frequent than the most frequent consonant. The most frequent consonant is N in Hittite, followed by Luvian, Basque, Finnic and Venezian. Next most frequent consonant is R in Umbrian and Mycenean, followed by S in Old Phrygian, Oscan and Estonian,

**Table 2.** The most frequent particular character is A

| Bq | | Vz | | EtB | | EtT | | Ph | | PhA | | Lu | | LuH | | My | | Hi | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 0.16 | a | 0.14 | a | 0.14 | a | 0.14 | a | 0.18 | a | 0.18 | a | 0.29 | a | 0.28 | a | 0.15 | a | 0.24 |
| e | 0.15 | e | 0.14 | i | 0.11 | i | 0.11 | i | 0.13 | i | 0.13 | i | 0.15 | i | 0.14 | o | 0.14 | n | 0.12 |
| i | 0.10 | o | 0.11 | e | 0.10 | e | 0.10 | e | 0.11 | e | 0.11 | u | 0.10 | u | 0.10 | i | 0.13 | u | 0.12 |
| n | 0.09 | i | 0.08 | l | 0.08 | t | 0.09 | s | 0.09 | s | 0.09 | n | 0.09 | n | 0.09 | e | 0.12 | i | 0.10 |
| r | 0.07 | n | 0.06 | n | 0.08 | l | 0.08 | o | 0.08 | o | 0.08 | t | 0.08 | d | 0.08 | u | 0.09 | s | 0.06 |
| s | 0.07 | r | 0.06 | u | 0.08 | n | 0.08 | t | 0.07 | t | 0.07 | s | 0.06 | z | 0.07 | r | 0.08 | t | 0.05 |
| t | 0.06 | s | 0.05 | s | 0.05 | u | 0.08 | n | 0.06 | n | 0.06 | r | 0.04 | t | 0.05 | k | 0.07 | r | 0.04 |
| u | 0.05 | l | 0.05 | r | 0.05 | r | 0.06 | k | 0.05 | v | 0.05 | l | 0.03 | r | 0.04 | t | 0.07 | m | 0.04 |
| o | 0.05 | k | 0.05 | c | 0.05 | s | 0.05 | v | 0.04 | k | 0.04 | z | 0.03 | l | 0.03 | p | 0.04 | h | 0.04 |
| k | 0.05 | t | 0.04 | o | 0.05 | c | 0.05 | m | 0.04 | m | 0.04 | p | 0.03 | h | 0.03 | n | 0.03 | d | 0.04 |

**Table 3.** The most frequent particular character is E

| Es | | LaS | | Gr | | Um | |
|---|---|---|---|---|---|---|---|
| e | 0.15 | e | 0.12 | e | 0.16 | e | 0.14 |
| a | 0.14 | i | 0.11 | o | 0.11 | r | 0.11 |
| i | 0.11 | u | 0.10 | a | 0.11 | u | 0.10 |
| s | 0.08 | a | 0.08 | i | 0.10 | i | 0.09 |
| l | 0.07 | t | 0.08 | n | 0.08 | t | 0.09 |
| u | 0.07 | s | 0.07 | s | 0.08 | a | 0.09 |
| k | 0.06 | r | 0.07 | t | 0.07 | s | 0.08 |
| t | 0.06 | n | 0.06 | r | 0.04 | o | 0.05 |
| n | 0.04 | o | 0.06 | u | 0.04 | n | 0.05 |
| d | 0.04 | m | 0.05 | p | 0.03 | p | 0.05 |

**Table 4.** The most frequent particular character is I

| RtT | | RtP | | RtV | | Fi | | Sl | | Cs | | LaC | | VeT | | VeP | | VeV | | Os | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| i | 0.17 | i | 0.18 | i | 0.19 | i | 0.14 | i | 0.14 | i | 0.17 | i | 0.12 | i | 0.13 | i | 0.15 | i | 0.18 | i | 0.13 |
| a | 0.15 | a | 0.15 | a | 0.14 | e | 0.13 | e | 0.11 | e | 0.12 | e | 0.11 | o | 0.13 | o | 0.12 | o | 0.13 | u | 0.10 |
| e | 0.10 | u | 0.09 | u | 0.09 | a | 0.12 | a | 0.10 | o | 0.10 | u | 0.10 | a | 0.10 | a | 0.10 | a | 0.10 | e | 0.09 |
| u | 0.09 | e | 0.09 | e | 0.09 | n | 0.09 | o | 0.07 | a | 0.07 | a | 0.09 | e | 0.09 | e | 0.09 | t | 0.09 | s | 0.09 |
| t | 0.07 | t | 0.07 | t | 0.08 | t | 0.08 | u | 0.07 | t | 0.06 | t | 0.08 | t | 0.08 | t | 0.08 | e | 0.09 | a | 0.08 |
| s | 0.07 | n | 0.06 | s | 0.06 | l | 0.07 | t | 0.06 | s | 0.05 | s | 0.08 | n | 0.07 | n | 0.07 | n | 0.07 | t | 0.08 |
| n | 0.06 | s | 0.06 | n | 0.06 | s | 0.07 | s | 0.06 | n | 0.05 | r | 0.07 | r | 0.06 | s | 0.05 | s | 0.06 | m | 0.06 |
| l | 0.05 | l | 0.05 | l | 0.05 | k | 0.06 | n | 0.06 | v | 0.05 | k | 0.06 | s | 0.06 | r | 0.05 | r | 0.05 | n | 0.06 |
| r | 0.04 | r | 0.04 | r | 0.04 | o | 0.06 | r | 0.04 | r | 0.04 | n | 0.06 | k | 0.05 | k | 0.05 | v | 0.04 | k | 0.05 |
| k | 0.04 | k | 0.04 | k | 0.04 | u | 0.05 | m | 0.04 | m | 0.03 | o | 0.06 | l | 0.04 | l | 0.04 | u | 0.04 | p | 0.04 |

T in Etruscan, Venetic, Latin, Rhaetic, Old Slovenian and Old Church Slavonic, and L in Etruscan read according to Bor [32], pp. 11-60.

*Frequency of pairs of characters*

In Tables 5-7 are presented the ten most frequent pairs of characters in the LDs.

The most frequent vowel pairs contain either I, or U, and this indicates that the resulting frequencies contain the information of vowels I, and U, as well as of semi-vowels similar to them. This is the consequence of the way of preparation of the LDs. Finnic, Estonian, Umbrian, Venezian, but also Etruscan and Rhaetic being read in the way of the Italian respectively German scientists, contain among most frequent ten character pairs no such doublet.

The most frequent vowel-consonant pair is AN in Old Anatolian languages (Hittite, Luvian), where also few other pairs are quite frequent, followed by ER in Umbrian, EN in Basque, ON in Old Greek, etc.

The most frequent consonant pair is RS in Umbrian, ST in Oscan, Old Slovenian, Old Church Slavonic, Estonian and Umbrian, LL in Finnic, and NT in Latin.

**Table 5.** The most frequent pair of characters is a vowel pair

| VeT | | VeP | | VeV | | Os | | Cs | | My | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ai | 0.026 | ai | 0.028 | ai | 0.033 | ei | 0.034 | ie | 0.063 | ii | 0.031 |
| ei | 0.023 | ii | 0.027 | ti | 0.031 | is | 0.034 | ni | 0.022 | io | 0.026 |
| on | 0.023 | ka | 0.025 | on | 0.029 | st | 0.026 | st | 0.019 | eu | 0.025 |
| ii | 0.023 | ti | 0.025 | ei | 0.029 | ud | 0.020 | ii | 0.019 | ro | 0.024 |
| ti | 0.022 | on | 0.023 | to | 0.028 | in | 0.019 | ti | 0.019 | ra | 0.024 |
| to | 0.022 | ei | 0.022 | na | 0.025 | us | 0.019 | ri | 0.017 | er | 0.023 |
| os | 0.020 | to | 0.022 | oi | 0.023 | ik | 0.018 | vi | 0.016 | ke | 0.023 |
| na | 0.019 | ek | 0.020 | ia | 0.023 | tu | 0.018 | že | 0.014 | ko | 0.023 |
| ia | 0.019 | ia | 0.020 | os | 0.023 | er | 0.017 | go | 0.013 | ia | 0.023 |
| ke | 0.018 | os | 0.019 | st | 0.022 | um | 0.016 | li | 0.013 | ta | 0.022 |

**Table 6.** The most frequent pair of characters is a vowel-consonant pair

| Hi | | Lu | | LuH | | Ph | | PhA | | Bq | | Fi | | Um | | Gr | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| an | 0.060 | an | 0.061 | an | 0.059 | at | 0.028 | at | 0.028 | en | 0.044 | en | 0.035 | er | 0.055 | on | 0.039 |
| nu | 0.046 | ta | 0.055 | ua | 0.039 | as | 0.028 | as | 0.027 | er | 0.032 | an | 0.025 | tu | 0.035 | en | 0.028 |
| ar | 0.034 | at | 0.044 | ta | 0.037 | oi | 0.026 | oi | 0.027 | ar | 0.031 | ta | 0.022 | pe | 0.024 | te | 0.027 |
| ua | 0.033 | ua | 0.041 | za | 0.036 | es | 0.025 | os | 0.027 | re | 0.027 | in | 0.021 | es | 0.023 | ei | 0.025 |
| as | 0.027 | ti | 0.039 | ar | 0.036 | os | 0.025 | ta | 0.027 | an | 0.026 | si | 0.021 | rs | 0.022 | os | 0.024 |
| ta | 0.026 | ar | 0.037 | ad | 0.035 | ta | 0.025 | ei | 0.025 | ta | 0.022 | le | 0.020 | ar | 0.020 | oi | 0.024 |
| za | 0.025 | as | 0.036 | dz | 0.032 | an | 0.024 | es | 0.024 | te | 0.022 | te | 0.020 | re | 0.017 | ai | 0.024 |
| ma | 0.022 | sa | 0.028 | ia | 0.026 | ei | 0.024 | te | 0.022 | be | 0.020 | ll | 0.017 | st | 0.016 | to | 0.023 |
| ia | 0.022 | ia | 0.027 | nd | 0.025 | te | 0.021 | an | 0.021 | ra | 0.020 | ne | 0.017 | se | 0.016 | es | 0.020 |
| un | 0.021 | pa | 0.026 | al | 0.024 | io | 0.021 | io | 0.019 | ai | 0.019 | ka | 0.016 | at | 0.016 | me | 0.019 |

**Table 7.** The most frequent pair of characters is a consonant-vowel pair

| LaC | | LaS | | Vz | | EtT | | EtB | | Es | | RtP | | RtT | | RtV | | Sl | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ku | 0.028 | ku | 0.028 | la | 0.025 | na | 0.028 | na | 0.028 | ta | 0.023 | ti | 0.027 | ti | 0.027 | ti | 0.034 | ti | 0.026 |
| er | 0.025 | er | 0.026 | ar | 0.025 | ar | 0.023 | la | 0.023 | le | 0.021 | ri | 0.024 | ri | 0.025 | ii | 0.027 | ie | 0.025 |
| is | 0.021 | re | 0.020 | de | 0.021 | ti | 0.023 | al | 0.020 | ka | 0.020 | na | 0.023 | na | 0.024 | it | 0.027 | st | 0.021 |
| um | 0.019 | um | 0.019 | er | 0.021 | la | 0.023 | in | 0.018 | is | 0.019 | is | 0.022 | is | 0.022 | ri | 0.026 | in | 0.020 |
| it | 0.018 | is | 0.019 | ko | 0.019 | al | 0.020 | ar | 0.017 | se | 0.019 | in | 0.021 | it | 0.022 | is | 0.022 | ni | 0.019 |
| re | 0.018 | ue | 0.018 | ke | 0.018 | tu | 0.019 | an | 0.017 | al | 0.018 | ie | 0.020 | nu | 0.020 | na | 0.021 | ri | 0.017 |
| in | 0.018 | it | 0.017 | el | 0.018 | in | 0.018 | ve | 0.016 | st | 0.017 | an | 0.020 | in | 0.020 | nu | 0.019 | se | 0.017 |
| ue | 0.017 | in | 0.017 | ve | 0.018 | an | 0.017 | ni | 0.015 | as | 0.016 | nu | 0.019 | an | 0.018 | in | 0.018 | na | 0.015 |
| nt | 0.016 | te | 0.016 | en | 0.018 | ta | 0.017 | ce | 0.015 | ma | 0.016 | ii | 0.019 | es | 0.018 | ta | 0.018 | ma | 0.015 |
| te | 0.015 | nt | 0.015 | ra | 0.017 | ve | 0.016 | ia | 0.014 | el | 0.016 | it | 0.018 | la | 0.018 | an | 0.017 | et | 0.015 |

### *Frequency of triplets of characters*

In Tables 8-11 there are presented the ten most frequent triplets of characters in the LDs.

The most frequent vowel triplets (v-v-v) appear in Hittite (iia>uua), Luvian (uua>iia), and Mycenean (iio>iia>eue), but also in Venetic (iio>ioi>iia>iai), Umbrian (oui>iou), Old Phrygian (eia), Rhaetic (iii), and Old Church Slavonic (iie). Also here they contain either I, or U, and this indicates that the resulting frequencies contain the information of vowels I, and U, as well as of semi-vowels similar to them.

Among combinations of two vowels and a consonant, we have three different possibilities: two vowels followed by a consonant (v-v-c), a consonant between two vowels (v-c-v), as well two vowels following a consonant (c-v-v). The first possibility, v-v-c, occurs most often in the Venetic (oek; i.e. in the so-called AKEO when read from above), Oscan (eis), Old Phrygian (ios>aes), followed by Venezian (ior), Old Greek (ion), Rhaetic (ies resp. iit), Old Church Slavonic (iem>ies), and Etruscan (ial). Except in Venetic and Old Phrygian, there is in all cases present an I. The second possibility, v-c-v, occurs most often in Rhaetic (eight times), seven times in Basque; six times in Old Slovenian; five times in Luvian, Mycenean; four times in Umbrian; three times in Old Phrygian, Finnic, Old Church Slavonic, Estonian; once in Venetic, Oscan, Latin, Venezian, Etruscan and Greek. In Hittite, no such combination has been observed among ten most frequent triplets. The third possibility, c-v-v, occurs less frequently. In combination with I: three times in Old Church Slavonic; twice in Venetic, Old Greek; once in Venezian, Old Phrygian, Estonian. In combination with U: three times in Latin, where it derives mostly from qu-; two times in Hittite, Oscan, Mycenean and Rhaetic. It appears also in Venetic from AKEO and none is observed among ten most frequent triplets in Basque, Etruscan, Finnic, Luvian, Old Slovenian, Rhaetic and Umbrian.

The combinations of one vowel and two consonants appear in three combinations as well. These are: v-c-c, c-v-c, and c-c-v. The first possibility, v-c-c, occurs among ten most frequent triplets three times in Estonian; two times in Umbrian, Hittite, Venezian, Old Slovenian, Finnic, and Old Greek; once in Luvian, Etruscan, and Latin. None is observed among the ten most frequent triplets in Basque, Mycenean, Old Church Slavonic, Old Phrygian, Oscan, Rhaetic and Venetic. The second possibility, c-v-c, occurs among ten most frequent triplets five times in Etruscan; four times in Oscan; three times in Basque, Old Greek, Old Phrygian, Venezian, and Latin; twice in Umbrian, Finnic; once in Hittite, Old Slovenian, Estonian. None is observed among the ten most frequent triplets in Luvian, Mycenean, Old Church Slavonic, and Rhaetic, whereas in Venetic it depends on the way of reading.

The third possibility, c-c-v, occurs three times in Hittite and Finnic; two times in Luvian; once in Venetic, Old Church Slavonic, Estonian, Old Slovenian, Venezian, and Etruscan. None is observed among the ten most frequent triplets in Basque, Latin, Mycenean, Old Phrygian, Old Greek, Oscan, Rhaetic and Umbrian.

Among the ten most frequent triplets, only one consonant triplet (c-c-c) is observed, in Luvian.

**Table 8.** The most frequent triplet of characters is a vowel triplet or a vowel-vowel-consonant triplet

| Hi | | My | | Os | | Vz | |
|---|---|---|---|---|---|---|---|
| iia | 0.020 | iio | 0.018 | eis | 0.016 | ior | 0.009 |
| uua | 0.016 | iia | 0.016 | ust | 0.011 | sio | 0.009 |
| dza | 0.014 | ara | 0.011 | tud | 0.011 | per | 0.009 |
| nua | 0.012 | ere | 0.011 | ere | 0.010 | ent | 0.008 |
| and | 0.011 | oro | 0.010 | ini | 0.010 | ave | 0.008 |
| uan | 0.011 | ata | 0.009 | pis | 0.009 | kos | 0.007 |
| sta | 0.011 | eue | 0.008 | uae | 0.009 | sta | 0.007 |
| nda | 0.011 | reu | 0.007 | sua | 0.009 | eni | 0.007 |
| nun | 0.011 | rii | 0.007 | nim | 0.008 | kon | 0.007 |
| arh | 0.011 | ake | 0.007 | ter | 0.008 | and | 0.007 |

**Table 9.** The most frequent triplet of characters is a vowel-consonant-vowel triplet

| Ph | | PhA | | Lu | | Sl | | VeV | | VeP | | Bq | | RtP | | RtT | | RtV | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ata | 0.016 | ata | 0.020 | ati | 0.027 | ati | 0.011 | ego | 0.025 | eka | 0.021 | eta | 0.020 | ina | 0.011 | iti | 0.013 | iti | 0.015 |
| ios | 0.014 | ate | 0.014 | uua | 0.026 | ine | 0.008 | ioi | 0.022 | oek | 0.019 | ere | 0.015 | iti | 0.011 | ina | 0.011 | esi | 0.010 |
| ate | 0.012 | ios | 0.013 | ata | 0.023 | eni | 0.008 | don | 0.021 | tii | 0.013 | ber | 0.014 | esi | 0.010 | esi | 0.010 | iii | 0.010 |
| eia | 0.011 | eia | 0.012 | iia | 0.022 | mar | 0.008 | tei | 0.018 | ego | 0.013 | ren | 0.013 | anu | 0.009 | anu | 0.009 | ina | 0.010 |
| mat | 0.008 | mat | 0.010 | nza | 0.015 | ari | 0.007 | sto | 0.018 | iai | 0.012 | era | 0.011 | inu | 0.008 | inu | 0.008 | iit | 0.008 |
| toi | 0.008 | ter | 0.008 | asa | 0.014 | sta | 0.007 | ast | 0.015 | sto | 0.012 | ten | 0.011 | ani | 0.007 | ita | 0.008 | inu | 0.008 |
| man | 0.008 | tes | 0.008 | tar | 0.013 | ete | 0.007 | ona | 0.015 | don | 0.012 | ari | 0.010 | avi | 0.007 | ale | 0.007 | ita | 0.008 |
| avo | 0.007 | toi | 0.008 | anz | 0.013 | ega | 0.007 | iai | 0.014 | iio | 0.011 | are | 0.010 | iii | 0.007 | avi | 0.007 | anu | 0.008 |
| tes | 0.007 | ais | 0.007 | ali | 0.012 | est | 0.007 | nas | 0.014 | iia | 0.011 | egi | 0.009 | ale | 0.006 | nua | 0.007 | ale | 0.007 |
| aes | 0.006 | aba | 0.007 | apa | 0.011 | ost | 0.006 | rei | 0.013 | ioi | 0.010 | ela | 0.009 | ies | 0.006 | pit | 0.007 | avi | 0.007 |

**Table 10.** The most frequent triplet of characters is a consonant-vowel-vowel or consonant-vowel-consonant triplet

| VeT | | LaC | | LaS | | Cs | | EtT | | Gr | | Um | | EtB | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| keo | 0.014 | kue | 0.012 | kue | 0.014 | vie | 0.011 | lar | 0.012 | men | 0.012 | per | 0.019 | vel | 0.010 |
| ake | 0.014 | ere | 0.007 | ere | 0.008 | pri | 0.010 | vel | 0.010 | kai | 0.009 | ers | 0.014 | ado | 0.008 |
| ego | 0.012 | ent | 0.007 | ent | 0.007 | rie | 0.010 | art | 0.009 | ton | 0.008 | oui | 0.012 | lad | 0.007 |
| iio | 0.012 | kui | 0.006 | kui | 0.006 | nii | 0.009 | tur | 0.007 | ion | 0.007 | etu | 0.011 | nas | 0.006 |
| tii | 0.012 | ter | 0.006 | ter | 0.006 | ago | 0.007 | nas | 0.006 | ont | 0.006 | est | 0.009 | ina | 0.006 |
| ioi | 0.011 | kua | 0.006 | per | 0.006 | eni | 0.007 | ina | 0.006 | toi | 0.006 | itu | 0.009 | ial | 0.006 |
| sto | 0.011 | per | 0.006 | eri | 0.005 | iem | 0.007 | tna | 0.006 | ron | 0.005 | tot | 0.009 | eri | 0.005 |
| iia | 0.011 | ant | 0.005 | ant | 0.005 | ako | 0.006 | uti | 0.006 | isi | 0.005 | iou | 0.009 | lar | 0.005 |
| tei | 0.010 | eri | 0.005 | tur | 0.005 | ies | 0.006 | tin | 0.006 | all | 0.005 | ina | 0.008 | vil | 0.005 |
| iai | 0.010 | tur | 0.005 | kon | 0.005 | iie | 0.006 | ial | 0.006 | ene | 0.005 | atu | 0.008 | arn | 0.005 |

**Table 11.** The most frequent triplet of characters is a consonant-consonant-vowel triplet

| Fi | | Es | |
|---|---|---|---|
| lle | 0.011 | sta | 0.009 |
| ine | 0.009 | ist | 0.006 |
| nen | 0.008 | aie | 0.006 |
| ehe | 0.007 | ast | 0.006 |
| lla | 0.007 | ene | 0.006 |
| sta | 0.007 | mai | 0.006 |
| ill | 0.007 | ale | 0.006 |
| sen | 0.006 | kal | 0.006 |
| ell | 0.006 | ema | 0.005 |
| aha | 0.006 | est | 0.005 |

*Average frequency*

The next simplest unidimensional presentation after the frequency of particular sounds, their pairs and triplets, is the average vowel and semivowel frequency versus average consonant frequency. This is presented in Figure 1. Mycenean is placed more to the right and it is omitted. The sequence of languages is the same as when the vowel-to-consonant frequency ratio is used, Figure 2.

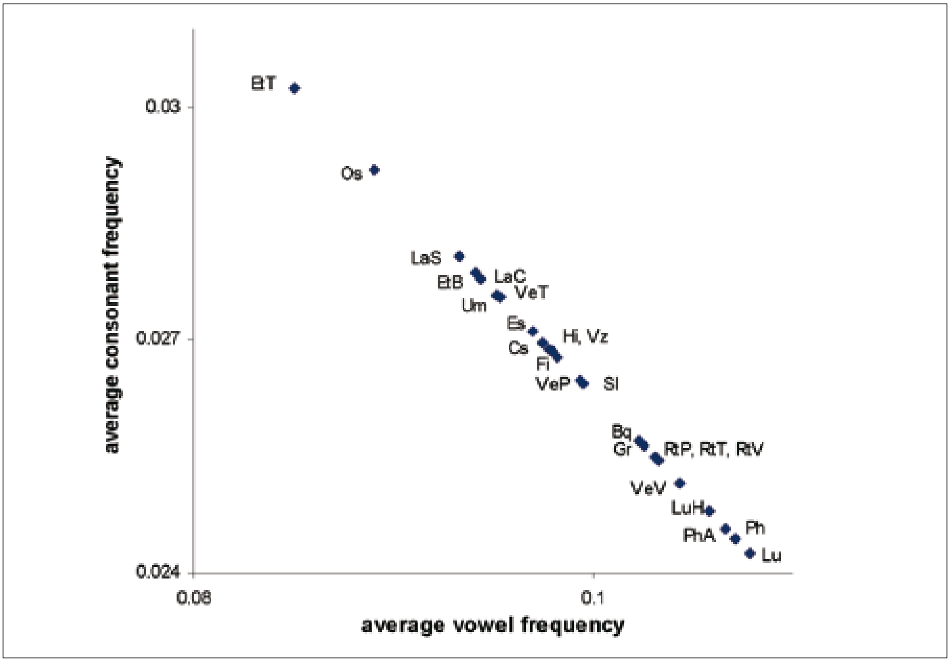


*Figure 1. Average frequency of consonants vs. average frequency of vowels+semivowels.*

*Vowel-to-consonant ratio*

The next simplest unidimensional presentation is the ratio of ∑(vowel and semivowel frequencies)/∑(consonant frequencies). This is presented in Figure 2.

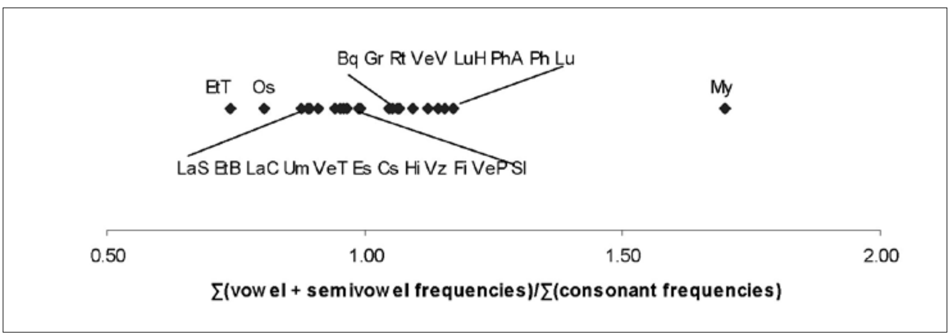


*Figure 2. Vowel-to-consonant ratio in the LDs*

Figure 2 shows that the languages are grouped into two main clusters, whereas Mycenean has a much higher vowel frequency than the other languages, while in Etruscan and Oscan the consonants prevail much more than in the other languages. Other languages are clustered around the point of equal frequency of vowels and consonants. The consonants are slightly prevailing in the Latin, Umbrian, Venetic, Estonian, Old Church Slavonic, Hittite, Venezian, Finnic, and Old Slovenian language, whereas the vowels are slightly prevailing in the Basque, Greek, Rhaetic, Luvian, and Old Phrygian language. However, it should not be forgotten that in the CsLD, the characters for half-sounds ‚jer' (ь) and ‚jor' (ъ) were eliminated and that in SlLD the half-sounds are in several instances not written, so that the real position of these Slavic languages is more in the vowel-prevailing side.

*K/S ratio*

Another group of simple comparisons is the ratio of sum frequencies of k, g, h sounds to the sum of frequencies of sibilants s, š, z, ž and affricate c, č. The results are presented in Figure 3.

In our LDs, the signs for k, g, h are prevailing over sibilants and affricate especially in Mycenean, but also in Finnic, Estonian and Latin in its classical notation. In other LDs, the sibilants and affricate are prevailing, especially in Etruscan, but also in Old Church Slavonic, Old Slovenian, etc.

The comparison of frequencies of above-mentioned characters in combination with the vowel that follows it is presented in Figure 4.

kw/cw presents the ratio of all pairs of k, g, and h with any vowel following it to all pairs of sibilants and affricate with any vowel following it.

ke/ce means the ratio of sums of frequencies of:

(ke+ki+ge+gi+he+hi)/(ce+ci+če+či+se+si+še+ši+ze+zi+že+ži),

ka/ca those containing vowels a, o, and u,

ke/ca presents the ratio of frequencies of k, g, h in combination with vowels e or i while the sibilants and affricate are in combination with vowels a, o, or u.

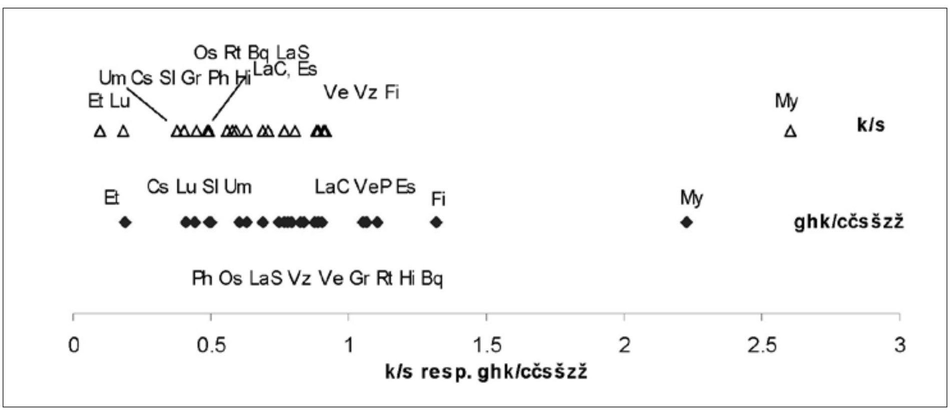


*Figure 3. The ratio of frequencies of g, h, k sounds vs. sibilants and affricate*

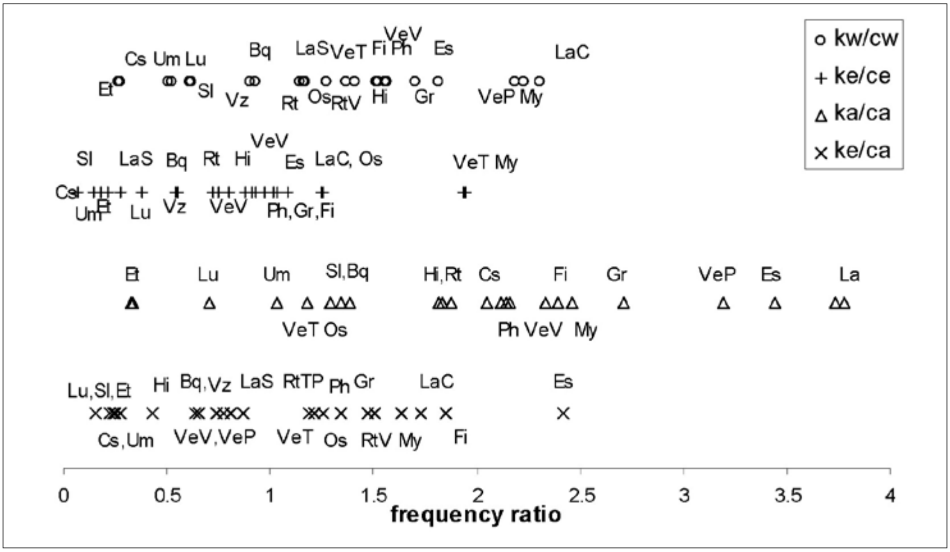


*Figure 4. Frequency ratio of k, g, h to c, č, s, š, z, ž in combinations with vowels*

In these combinations, in Latin, Estonian, Greek, Finnic, etc., the sounds k, g, h prevail, while in Etruscan, Old Church Slavonic, Umbrian, Luwian, Old Slovenian, Venezian and Basque the sibilants and affricate prevail in most cases.

## *Last character in the word*

Interesting are also the frequencies of the last character in a word. Their determination is straightforward in the languages known in detail, while it may be only a supposition for

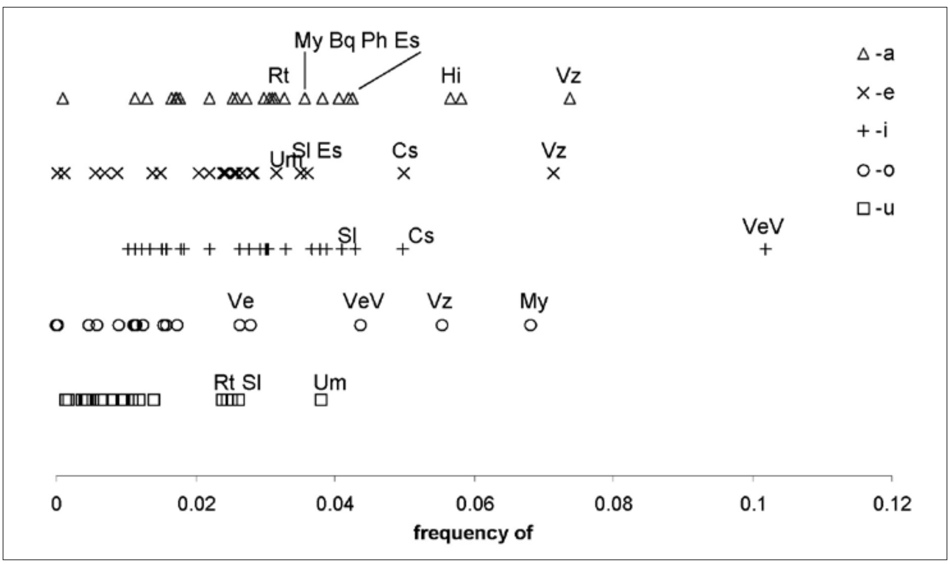


*Figure 5. Frequency of vowels as the last character in the words.*

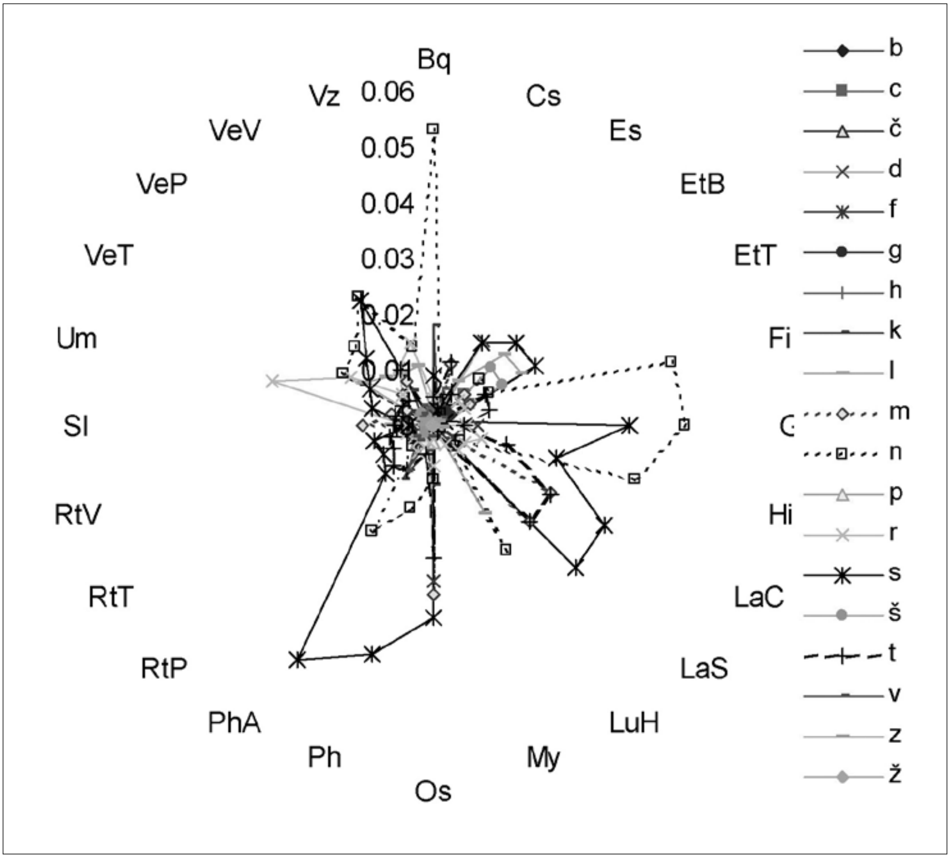


*Figure 6. Frequency of consonants as the last character in the words.*

inscriptions written *in continuo*. Some of these comparisons are presented in Figures 5 and 6. Figure 5 presents the frequency of vowels as the last sign in a word. Figure 6 presents the frequency of consonants as the last sign in the word.

Figure 7 presents the ratio of sums of frequencies of last vowels to those of last consonants in the words. Here it is evident that in Oscan, Latin, and Greek the consonants prevail as the last character in the words. In Mycenean (not shown due to the ratio higher for orders of magnitude), Venezian, Slavic, etc, the vowels prevail as the last character. Significant are results for different readings of some languages. For Latin the classic and semiclassic pronunciation give rise to almost the same result. Also for Rhaetic different decipherments give similar results, close to Old Slovenian. The largest differences are among different readings of Venetic. Among Venetic, Etruscan and Old Phrygian, the decipherments based on Latin and Greek give results closer to those of Latin and Greek, whereas those based on Slavic are reflecting a greater separation from these classic languages.

The frequency of the most frequent final consonants s, n, and t, and in connection to the vowels is presented in Figures 8-10.

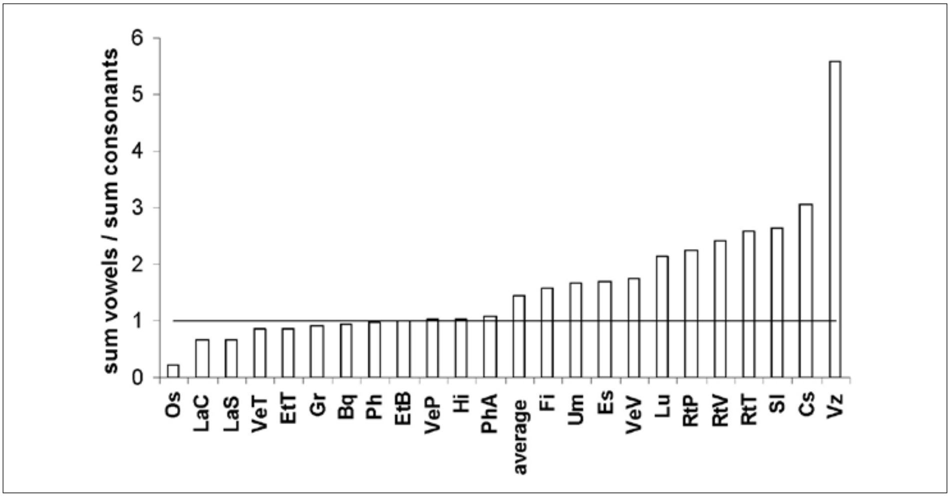


*Figure 7. Ratio of sum of frequencies of last vowels vs. last consonant in the words.*

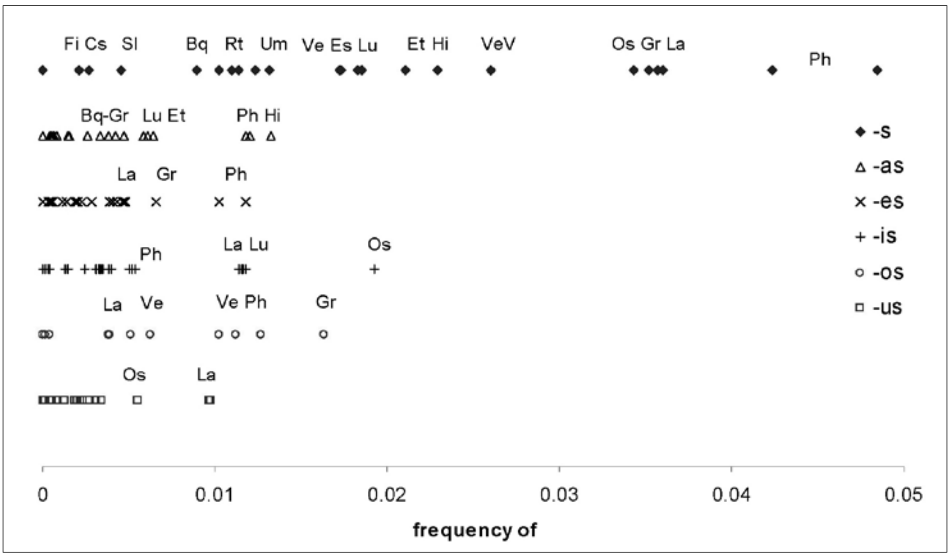


*Figure 8. The frequency of consonant -s and its combinations with vowels as the last character in the words.*

The sibilant s is in general one of the most frequent final consonants in words in Phrygian, Latin, Greek, and Oscan, and the least frequent in Finnic, Old Church Slavonic and Old Slovenian. In combination with vowels, the well-known characteristics of Greek resp. Latin are expressed as well.

The nasal n is in general one of the most frequent final consonants in words in Basque, followed by Greek, Finnic, Hittite, etc.

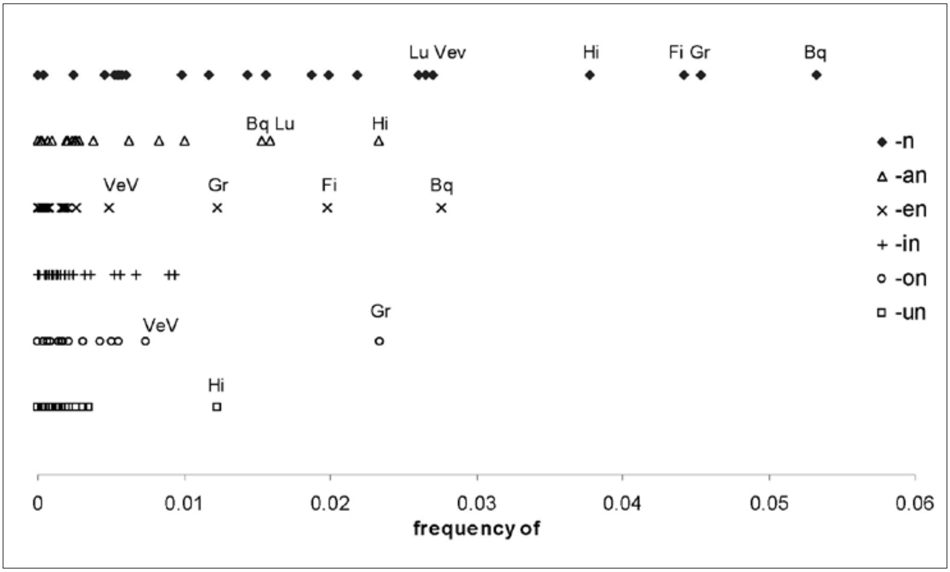


*Figure 9. The frequency of consonant -n and its combinations with vowels as the last character in the words.*

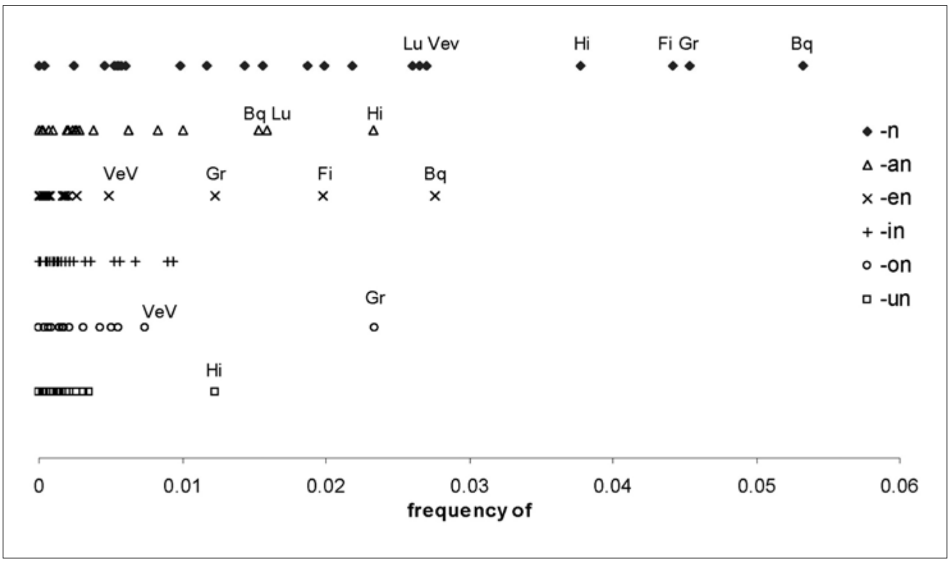


*Figure 10. The frequency of consonant -t and its combinations with vowels as the last character in the words.*

The plosive t is in general one of the most frequent final character in words in Latin and Oscan, followed by Hittite. In Old Church Slavonic this is the case in all events only due to the way of preparation of the database, cf. Methods.

## Bidimensional presentation

The root mean square difference approach of Silvestri and Tomezzoli [8,9] gives a bidimensional result. The root mean square differences of sound frequencies relative to Classical Latin are presented in Figure 11. Here we see the languages from Classical Latin to Etruscan read in the Pallottino's [44] way (EtT) placed near a straight line, then in the centre a cluster containing Estonian, Greek, Old Church Slavonic and between them Mycenean, Basque, Rhaetic, Old Phrygian, Venetic, and Venezian. On the right side of this cluster, Hittite and Luvian form another but loose cluster.

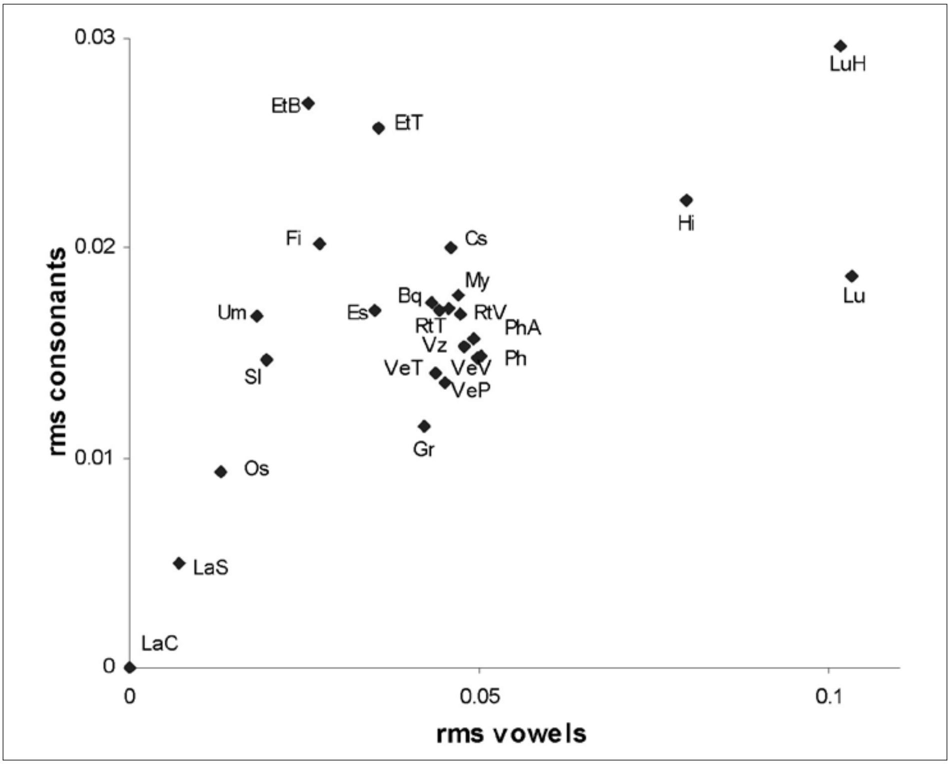


*Figure 11. Root mean square difference to Classical Latin*

## Multidimensional approach

For a presentation, which takes into account not only one or two parameters as the above frequencies or ratios of them but the frequencies of all of 24 signs used for the common notation of all databases in question, we used the Principle Component Analysis (PCA). Its results are presented below in the following Tables and Figures. As the main ways of presenting them we use the amount of variance (%) contained in each PC axis, the spread of languages in the two-dimensional spaces defined by particular pairs of the PC axes, as well as the dimensionless distances of languages from the origin of the multidimensional PC space and the dimensionless distances between the languages in question.

*Using frequencies of single characters*

The amount of variance (%) contained in each PC axis is presented in Table 12.

Table 12. Amount of variance contained on particular PC axes (#), % of total

| PC | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 | #11 | #12 | #13 | #14 | #15 | #16 | #17 | #18 | #19 | #20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (%) | 22.1 | 17.1 | 11.8 | 9.5 | 9.0 | 6.9 | 6.0 | 4.8 | 3.4 | 2.7 | 1.8 | 1.5 | 1.1 | 0.8 | 0.5 | 0.3 | 0.3 | 0.3 | 0.1 | 0.1 |

The dimensionless distance of a language from the origin of the multidimensional PC space is presented in Figure 12.

Looking at Figure12, one should be aware that it is scalar and not vectorial, thus it presents only the distance but not the direction in which the distance is realized. For this

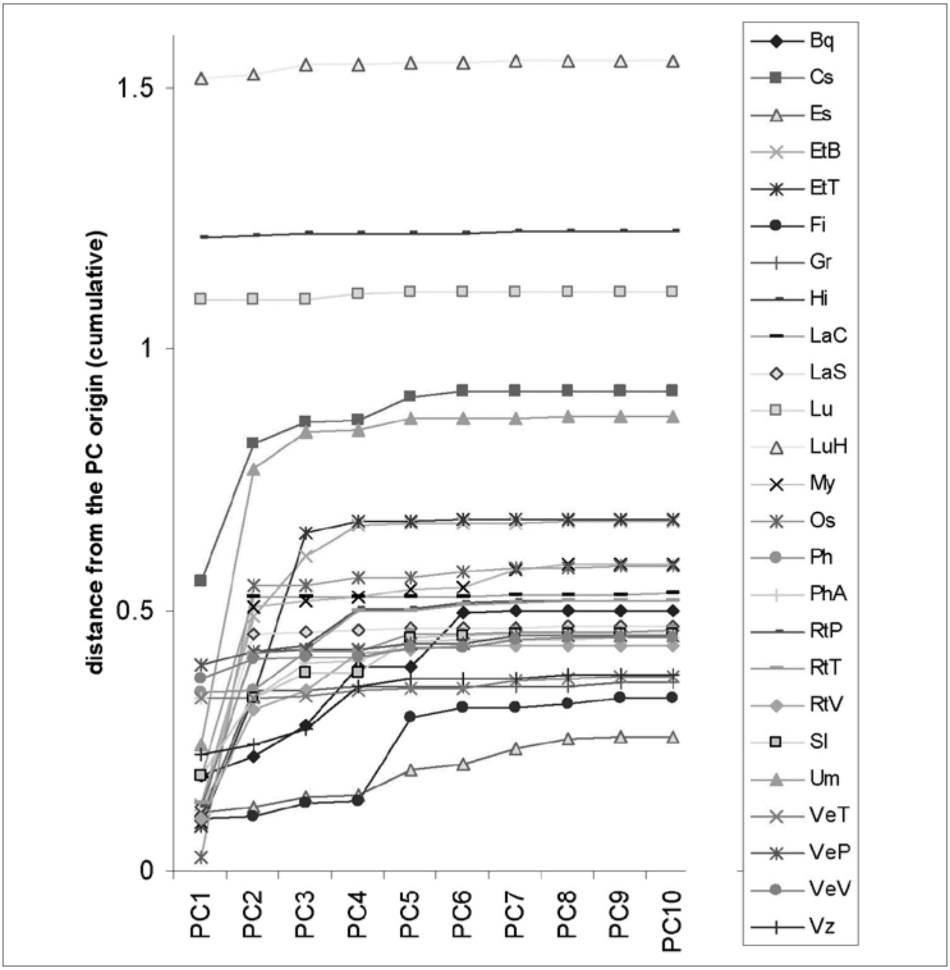


*Figure 12. Dimensionless distance of languages from the PC origin using the frequencies of single characters*

reason Figure 12 is not appropriate to estimate the distances between languages. Two languages, which are far apart from one another in Figure 12, are in reality at least that much apart and usually more. Two languages, which may appear proximate on Figure 12, may be really either proximate or distant. Figure 12, however, makes a clear indication that to estimate the distances between languages, the information about them collected on the first four PC axes (i.e. PC1 to PC4), which contain the information of approx. 60% of variance, may be sufficient for a number of them. The information contained on the first six PCA axes (i.e. PC1 to PC6), which contain the information of over 75% of variance, is sufficient for most of them. The first ten PC axes (i.e. PC1 to PC10) contain the information of over 93% of variance and, there, a clear levelling-off is seen in all cases. Thus, the evaluations in a six- to eight-dimensional PC space, containing information about over 75% resp. over 87% of variance, are sufficient. Because of the clear levelling-off, we use later on data of ten PC axes to evaluate the distances.

In Figure 12 can be however clearly seen that the Old Anatolian languages, Luvian and Hittite are the most distant from the origin of present PC space, whereas Estonian is the least distant. However, a different collection of languages may give a different distribution of them.

In Figure 13, the axis PC1, to which about 22% of total variance of data is associated, separates the Old Anatolian languages (Hittite and Luvian) from European ones. For several European languages, except for Old Church Slavonic, Venetic, Old Phrygian, Venezian, and Umbrian, there is little information on the axis PC1.

The axis PC2, to which about 17% of total variance of data is associated, separates the European languages into two groups:

a. The Greco-Italic group consisting of (in the series of decreasing amount of information) Umbrian, Oscan, Mycenean, Latin, and Old Greek,

b. The other European group consisting of Old Church Slavonic, Etruscan, Rhaetic, Old Slovenian, Venetic, Old Phrygian, Venezian, but also Basque, Estonian and Finnic. However, for the latter three languages as well as for Old Phrygian and Venezian there is little information on the axis PC2.

The axis PC3, to which about 12% of total variance of data is associated, separates the Old Church Slavonic, Old Phrygian and Old Slovenian into a group, to which are close also Basque, Venezian and Mycenean. Another group form the Etruscan and Umbrian, followed by Rhaetic, whereas for the other languages there is little information on the axis PC3.

The axis PC4, to which about 9.5% of total variance of data is associated, separates on the one hand Basque, Etruscan and Venezian, and on the other hand the Rhaetic from the rest of the languages.

The axis PC5, to which about 9% of total variance of data is associated, separates on the one hand Finnic from Phrygian and Estonian, etc, and on the other hand Old Church Slavonic from Old Slovenian and Umbrian, etc.

The axis PC6, to which about 7% of total variance of data is associated, separates first of all Basque from the other languages.

The axis PC7, to which about 6% of total variance of data is associated, separates on the

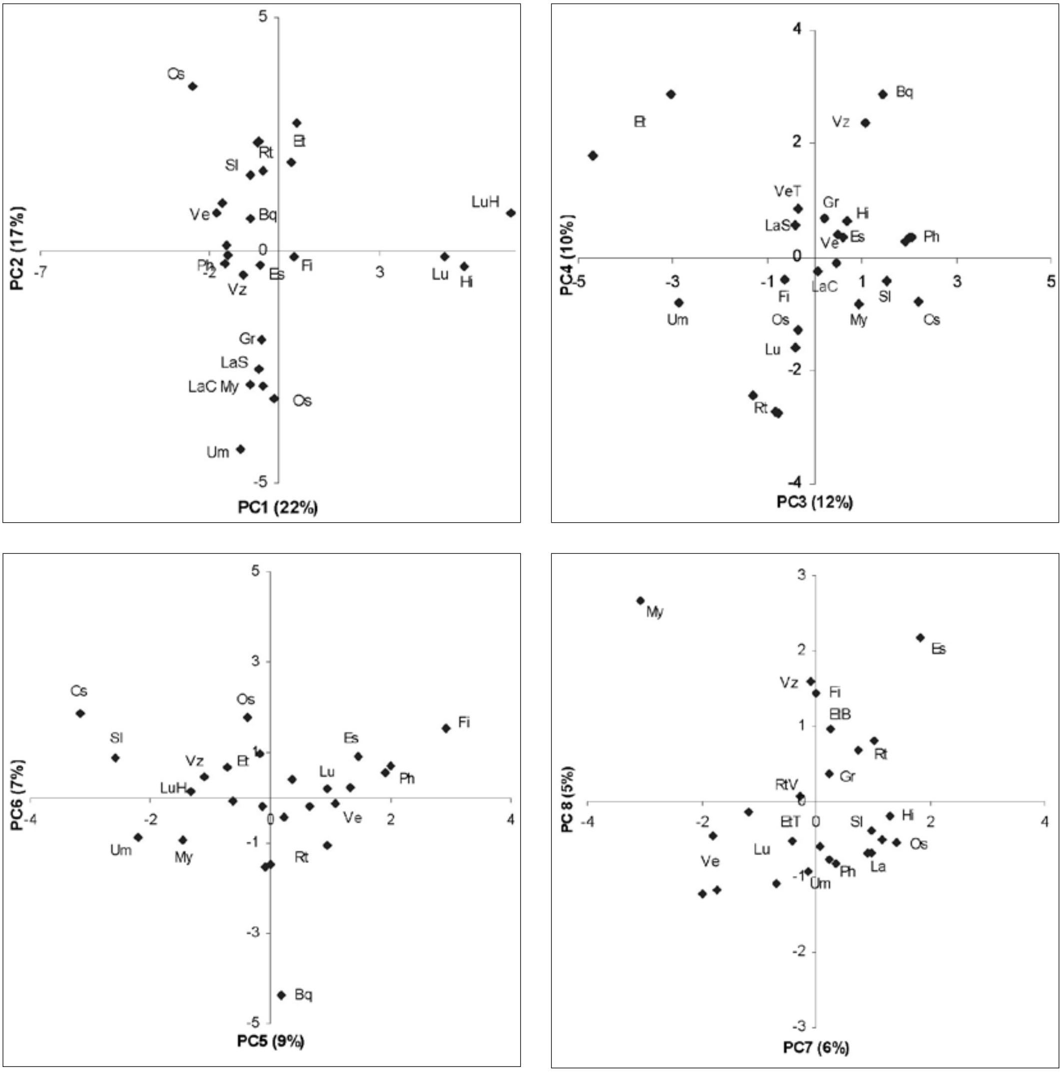


*Figure 13. Information presented by the PC axes PC1 to PC8.*

one hand Estonian and on the other hand Mycenean and Venetic from the other ones.

The axis PC8, to which about 5% of total variance of data is associated, separates on the one hand Mycenean and Estonian from the other languages.

Another point of view gives us the Pythagorean distance between languages in the ten-dimensional PC space. Table 13 gives us the matrix of all data, whereas Table 14 gives us the distance between different interpretations of the same language, and Table 15 the smallest distances between some ancient languages and other languages taken into account. Table 16 gives the smallest distances for some other languages.

**Table 13.** Dimensionless distances between languages in the ten-dimensional PC space.

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bq | 0 | | | | | | | | | | | | | | | | | | | | | | | | |
| Cs | 0.88 | 0 | | | | | | | | | | | | | | | | | | | | | | | |
| Es | 0.52 | 0.94 | 0 | | | | | | | | | | | | | | | | | | | | | | |
| EtB | 0.80 | 1.03 | 0.79 | 0 | | | | | | | | | | | | | | | | | | | | | |
| EtT | 0.89 | 1.14 | 0.80 | 0.28 | 0 | | | | | | | | | | | | | | | | | | | | |
| Fi | 0.71 | 1.12 | 0.34 | 0.74 | 0.70 | 0 | | | | | | | | | | | | | | | | | | | |
| Gr | 0.62 | 1.12 | 0.34 | 0.95 | 0.91 | 0.46 | 0 | | | | | | | | | | | | | | | | | | |
| Hi | 1.45 | 1.93 | 1.34 | 1.31 | 1.36 | 1.17 | 1.35 | 0 | | | | | | | | | | | | | | | | | |
| LaC | 0.76 | 1.23 | 0.50 | 1.12 | 1.05 | 0.64 | 0.24 | 1.46 | 0 | | | | | | | | | | | | | | | | |
| LaS | 0.71 | 1.21 | 0.47 | 1.02 | 0.95 | 0.61 | 0.20 | 1.40 | 0.13 | 0 | | | | | | | | | | | | | | | |
| Lu | 1.41 | 1.83 | 1.25 | 1.22 | 1.23 | 1.03 | 1.26 | 0.32 | 1.37 | 1.31 | 0 | | | | | | | | | | | | | | |
| LuH | 1.76 | 2.14 | 1.68 | 1.57 | 1.66 | 1.52 | 1.71 | 0.45 | 1.83 | 1.77 | 0.61 | 0 | | | | | | | | | | | | | |
| My | 0.83 | 1.25 | 0.62 | 1.18 | 1.14 | 0.73 | 0.39 | 1.43 | 0.35 | 0.38 | 1.33 | 1.75 | 0 | | | | | | | | | | | | |
| Os | 0.92 | 1.33 | 0.58 | 1.15 | 1.07 | 0.64 | 0.35 | 1.36 | 0.24 | 0.27 | 1.25 | 1.73 | 0.42 | 0 | | | | | | | | | | | |
| Ph | 0.51 | 0.84 | 0.34 | 0.98 | 1.00 | 0.58 | 0.46 | 1.58 | 0.57 | 0.58 | 1.48 | 1.90 | 0.68 | 0.71 | 0 | | | | | | | | | | |
| PhA | 0.51 | 0.82 | 0.33 | 0.96 | 0.98 | 0.56 | 0.47 | 1.57 | 0.59 | 0.60 | 1.47 | 1.88 | 0.70 | 0.72 | 0.04 | 0 | | | | | | | | | |
| RtP | 0.70 | 0.73 | 0.61 | 0.67 | 0.69 | 0.64 | 0.82 | 1.46 | 0.94 | 0.91 | 1.31 | 1.73 | 0.98 | 0.99 | 0.71 | 0.68 | 0 | | | | | | | | |
| RtT | 0.69 | 0.72 | 0.61 | 0.68 | 0.70 | 0.64 | 0.82 | 1.47 | 0.93 | 0.90 | 1.31 | 1.74 | 0.98 | 0.99 | 0.71 | 0.68 | 0.03 | 0 | | | | | | | |
| RtV | 0.68 | 0.83 | 0.54 | 0.65 | 0.62 | 0.51 | 0.73 | 1.42 | 0.84 | 0.81 | 1.25 | 1.71 | 0.90 | 0.89 | 0.65 | 0.62 | 0.17 | 0.18 | 0 | | | | | | |
| Sl | 0.56 | 0.51 | 0.53 | 0.75 | 0.84 | 0.71 | 0.70 | 1.47 | 0.82 | 0.78 | 1.38 | 1.72 | 0.85 | 0.90 | 0.56 | 0.54 | 0.48 | 0.47 | 0.53 | 0 | | | | | |
| Um | 1.10 | 1.51 | 0.90 | 1.32 | 1.18 | 0.97 | 0.63 | 1.68 | 0.48 | 0.49 | 1.57 | 2.06 | 0.59 | 0.50 | 0.98 | 1.00 | 1.20 | 1.19 | 1.10 | 1.15 | 0 | | | | |
| VeT | 0.48 | 0.81 | 0.39 | 0.77 | 0.75 | 0.54 | 0.45 | 1.57 | 0.58 | 0.54 | 1.45 | 1.89 | 0.67 | 0.72 | 0.35 | 0.34 | 0.59 | 0.59 | 0.51 | 0.52 | 0.87 | 0 | | | |
| VeP | 0.49 | 0.69 | 0.44 | 0.81 | 0.83 | 0.60 | 0.58 | 1.64 | 0.70 | 0.68 | 1.52 | 1.94 | 0.77 | 0.84 | 0.32 | 0.30 | 0.55 | 0.55 | 0.49 | 0.48 | 1.03 | 0.18 | 0 | | |
| VeV | 0.53 | 0.68 | 0.45 | 0.80 | 0.81 | 0.59 | 0.60 | 1.62 | 0.73 | 0.71 | 1.49 | 1.91 | 0.80 | 0.86 | 0.34 | 0.31 | 0.51 | 0.51 | 0.45 | 0.47 | 1.06 | 0.22 | 0.08 | 0 | |
| Vz | 0.43 | 0.87 | 0.35 | 0.82 | 0.87 | 0.60 | 0.36 | 1.46 | 0.52 | 0.46 | 1.40 | 1.77 | 0.57 | 0.65 | 0.39 | 0.40 | 0.76 | 0.75 | 0.71 | 0.49 | 0.87 | 0.36 | 0.43 | 0.48 | 0 |
| | Bq | Cs | Es | EtB | EtT | Fi | Gr | Hi | LaC | LaS | Lu | LuH | My | Os | Ph | PhA | RtP | RtT | RtV | Sl | Um | VeT | VeP | VeV | Vz |

**Table 14.** Dimensionless distances in the ten-dimensional PC space between different presentations of the same language

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ph | PhA | 0.04 | RtP | RtT | 0.03 | VeP | VeV | 0.08 |
| LaC | LaS | 0.13 | RtP | RtV | 0.17 | VeT | VeP | 0.18 |
| EtB | EtT | 0.28 | RtT | RtV | 0.18 | VeT | VeV | 0.22 |

Table 14 presents the dimensionless distances between different presentations of the same language. In our case they span from 0.03 in the case of Rhaetian to almost 0.3 in the case of Etruscan. Except the distances between Latin and Greek resp. Oscan, they are smaller than the distances between different languages, cf. Table 15 and 16.

**Table 15.** The distances of some ancient languages to other ones in the ten-dimensional PC space

| EtT | | EtB | | Ph | | PhA | | RtT | | RtP | | RtV | | VeT | | VeP | | VeV | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RtV | 0.62 | RtV | 0.65 | VeP | 0.32 | VeP | 0.30 | Sl | 0.47 | Sl | 0.48 | VeV | 0.45 | PhA | 0.34 | PhA | 0.30 | PhA | 0.31 |
| RtP | 0.69 | RtP | 0.67 | Es | 0.34 | VeV | 0.31 | VeV | 0.51 | VeV | 0.51 | VeP | 0.49 | Ph | 0.35 | Ph | 0.32 | Ph | 0.34 |
| Fi | 0.70 | RtT | 0.68 | VeV | 0.34 | Es | 0.33 | VeP | 0.55 | VeP | 0.55 | VeT | 0.51 | Vz | 0.36 | Vz | 0.43 | RtV | 0.45 |
| RtT | 0.70 | Fi | 0.74 | VeT | 0.35 | VeT | 0.34 | VeT | 0.59 | VeT | 0.59 | Fi | 0.51 | Es | 0.39 | Es | 0.44 | Es | 0.45 |
| VeT | 0.75 | Sl | 0.75 | Vz | 0.39 | Vz | 0.40 | Es | 0.61 | Es | 0.61 | Sl | 0.53 | Gr | 0.45 | Sl | 0.48 | Sl | 0.47 |
| Es | 0.80 | VeT | 0.77 | Gr | 0.46 | Gr | 0.47 | Fi | 0.64 | Fi | 0.64 | Es | 0.54 | Bq | 0.48 | Bq | 0.49 | Vz | 0.48 |
| VeV | 0.81 | Es | 0.79 | Bq | 0.51 | Bq | 0.51 | EtB | 0.68 | EtB | 0.67 | PhA | 0.62 | RtV | 0.51 | RtV | 0.49 | RtP | 0.51 |
| VeP | 0.83 | Bq | 0.80 | Sl | 0.56 | Sl | 0.54 | PhA | 0.68 | PhA | 0.68 | EtT | 0.62 | Sl | 0.52 | RtT | 0.55 | RtT | 0.51 |
| Sl | 0.84 | VeV | 0.80 | LaC | 0.57 | Fi | 0.56 | Bq | 0.69 | EtT | 0.69 | EtB | 0.65 | Fi | 0.54 | RtP | 0.55 | Bq | 0.53 |
| Vz | 0.87 | VeP | 0.81 | LaS | 0.58 | LaC | 0.59 | EtT | 0.70 | Bq | 0.70 | Ph | 0.65 | LaS | 0.54 | Gr | 0.58 | Fi | 0.59 |
| Bq | 0.89 | Vz | 0.82 | Fi | 0.58 | LaS | 0.60 | Ph | 0.71 | Ph | 0.71 | Bq | 0.68 | LaC | 0.58 | Fi | 0.60 | Gr | 0.60 |
| Gr | 0.91 | Gr | 0.95 | RtV | 0.65 | RtV | 0.62 | Cs | 0.72 | Cs | 0.73 | Vz | 0.71 | RtT | 0.59 | LaS | 0.68 | Cs | 0.68 |
| LaS | 0.95 | PhA | 0.96 | My | 0.68 | RtT | 0.68 | Vz | 0.75 | Vz | 0.76 | Gr | 0.73 | RtP | 0.59 | Cs | 0.69 | LaS | 0.71 |
| PhA | 0.98 | Ph | 0.98 | RtT | 0.71 | RtP | 0.68 | Gr | 0.82 | Gr | 0.82 | LaS | 0.81 | My | 0.67 | LaC | 0.70 | LaC | 0.73 |
| Ph | 1.00 | LaS | 1.02 | Os | 0.71 | My | 0.70 | LaS | 0.90 | LaS | 0.91 | Cs | 0.83 | Os | 0.72 | My | 0.77 | My | 0.80 |
| LaC | 1.05 | Cs | 1.03 | RtP | 0.71 | Os | 0.72 | LaC | 0.93 | LaC | 0.94 | LaC | 0.84 | EtT | 0.75 | EtB | 0.81 | EtB | 0.80 |
| Os | 1.07 | LaC | 1.12 | Cs | 0.84 | Cs | 0.82 | My | 0.98 | My | 0.98 | Os | 0.89 | EtB | 0.77 | EtT | 0.83 | EtT | 0.81 |
| Cs | 1.14 | Os | 1.15 | Um | 0.98 | EtB | 0.96 | Os | 0.99 | Os | 0.99 | My | 0.90 | Cs | 0.81 | Os | 0.84 | Os | 0.86 |
| My | 1.14 | My | 1.18 | EtB | 0.98 | EtT | 0.98 | Um | 1.19 | Um | 1.20 | Um | 1.10 | Um | 0.87 | Um | 1.03 | Um | 1.06 |
| Um | 1.18 | Lu | 1.22 | EtT | 1.00 | Um | 1.00 | Lu | 1.31 | Lu | 1.31 | Lu | 1.25 | Lu | 1.45 | Lu | 1.52 | Lu | 1.49 |
| Lu | 1.23 | Hi | 1.31 | Lu | 1.48 | Lu | 1.47 | Hi | 1.47 | Hi | 1.46 | Hi | 1.42 | Hi | 1.57 | Hi | 1.64 | Hi | 1.62 |
| Hi | 1.36 | Um | 1.32 | Hi | 1.58 | Hi | 1.57 | | | | | | | | | | | | |

**Table 16.** The distances of some classical languages in the ten-dimensional PC space

| Greek | | Latin Class. | | Latin Semi. | | Oscan | | Umbrian | | Mycenean | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LaS | 0.20 | Gr | 0.24 | Gr | 0.20 | LaC | 0.24 | LaC | 0.48 | LaC | 0.35 |
| LaC | 0.24 | Os | 0.24 | Os | 0.27 | LaS | 0.27 | LaS | 0.49 | LaS | 0.38 |
| Es | 0.34 | My | 0.35 | My | 0.38 | Gr | 0.35 | Os | 0.50 | Gr | 0.39 |
| Os | 0.35 | Um | 0.48 | Vz | 0.46 | My | 0.42 | My | 0.59 | Os | 0.42 |
| Vz | 0.36 | Es | 0.50 | Es | 0.47 | Um | 0.50 | Gr | 0.63 | Vz | 0.57 |
| My | 0.39 | Vz | 0.52 | Um | 0.49 | Es | 0.58 | Vz | 0.87 | Um | 0.59 |
| VeT | 0.45 | Ph | 0.57 | VeT | 0.54 | Fi | 0.64 | VeT | 0.87 | Es | 0.62 |
| Ph | 0.46 | VeT | 0.58 | Ph | 0.58 | Vz | 0.65 | Es | 0.90 | VeT | 0.67 |
| Fi | 0.46 | PhA | 0.59 | PhA | 0.60 | Ph | 0.71 | Fi | 0.97 | Ph | 0.68 |
| PhA | 0.47 | Fi | 0.64 | Fi | 0.61 | VeT | 0.72 | Ph | 0.98 | PhA | 0.70 |
| VeP | 0.58 | VeP | 0.70 | VeP | 0.68 | PhA | 0.72 | PhA | 1.00 | Fi | 0.73 |
| VeV | 0.60 | VeV | 0.73 | Bq | 0.71 | VeP | 0.84 | VeP | 1.03 | VeP | 0.77 |
| Bq | 0.62 | Bq | 0.76 | VeV | 0.71 | VeV | 0.86 | VeV | 1.06 | VeV | 0.80 |

**Table 16.** Continued

| Greek | | Latin Class. | | Latin Semi. | | Oscan | | Umbrian | | Mycenean | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Um | 0.63 | Sl | 0.82 | Sl | 0.78 | RtV | 0.89 | RtV | 1.10 | Bq | 0.83 |
| Sl | 0.70 | RtV | 0.84 | RtV | 0.81 | Sl | 0.90 | Bq | 1.10 | Sl | 0.85 |
| RtV | 0.73 | RtT | 0.93 | RtT | 0.90 | Bq | 0.92 | Sl | 1.15 | RtV | 0.90 |
| RtT | 0.82 | RtP | 0.94 | RtP | 0.91 | RtT | 0.99 | EtT | 1.18 | RtT | 0.98 |
| RtP | 0.82 | EtT | 1.05 | EtT | 0.95 | RtP | 0.99 | RtT | 1.19 | RtP | 0.98 |
| EtT | 0.91 | EtB | 1.12 | EtB | 1.02 | EtT | 1.07 | RtP | 1.20 | EtT | 1.14 |
| EtB | 0.95 | Cs | 1.23 | Cs | 1.21 | EtB | 1.15 | EtB | 1.32 | EtB | 1.18 |
| Cs | 1.12 | Lu | 1.37 | Lu | 1.31 | Lu | 1.25 | Cs | 1.51 | Cs | 1.25 |
| Lu | 1.26 | Hi | 1.46 | Hi | 1.40 | Cs | 1.33 | Lu | 1.57 | Lu | 1.33 |
| Hi | 1.35 | | | | | Hi | 1.36 | Hi | 1.68 | Hi | 1.43 |

**Table 17.** The distances of some other languages in the ten-dimensional PC space

| Basque | | O. Sloven. | | OChSl | | Estonian | | Finnic | | Venezian | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Vz | 0.43 | VeV | 0.47 | Sl | 0.51 | PhA | 0.33 | Es | 0.34 | Es | 0.35 |
| VeT | 0.48 | RtT | 0.47 | VeV | 0.68 | Ph | 0.34 | Gr | 0.46 | VeT | 0.36 |
| VeP | 0.49 | VeP | 0.48 | VeP | 0.69 | Gr | 0.34 | RtV | 0.51 | Gr | 0.36 |
| PhA | 0.51 | RtP | 0.48 | RtT | 0.72 | Fi | 0.34 | VeT | 0.54 | Ph | 0.39 |
| Ph | 0.51 | Vz | 0.49 | RtP | 0.73 | Vz | 0.35 | PhA | 0.56 | PhA | 0.40 |
| Es | 0.52 | Cs | 0.51 | VeT | 0.81 | VeT | 0.39 | Ph | 0.58 | VeP | 0.43 |
| VeV | 0.53 | VeT | 0.52 | PhA | 0.82 | VeP | 0.44 | VeV | 0.59 | Bq | 0.43 |
| Sl | 0.56 | RtV | 0.53 | RtV | 0.83 | VeV | 0.45 | Vz | 0.60 | LaS | 0.46 |
| Gr | 0.62 | Es | 0.53 | Ph | 0.84 | LaS | 0.47 | VeP | 0.60 | VeV | 0.48 |
| RtV | 0.68 | PhA | 0.54 | Vz | 0.87 | LaC | 0.50 | LaS | 0.61 | Sl | 0.49 |
| RtT | 0.69 | Ph | 0.56 | Bq | 0.88 | Bq | 0.52 | RtP | 0.64 | LaC | 0.52 |
| RtP | 0.70 | Bq | 0.56 | Es | 0.94 | Sl | 0.53 | RtT | 0.64 | My | 0.57 |
| Fi | 0.71 | Gr | 0.70 | EtB | 1.03 | RtV | 0.54 | Os | 0.64 | Fi | 0.60 |
| LaS | 0.71 | Fi | 0.71 | Fi | 1.12 | Os | 0.58 | LaC | 0.64 | Os | 0.65 |
| LaC | 0.76 | EtB | 0.75 | Gr | 1.12 | RtT | 0.61 | EtT | 0.70 | RtV | 0.71 |
| EtB | 0.80 | LaS | 0.78 | EtT | 1.14 | RtP | 0.61 | Sl | 0.71 | RtT | 0.75 |
| My | 0.83 | LaC | 0.82 | LaS | 1.21 | My | 0.62 | Bq | 0.71 | RtP | 0.76 |
| Cs | 0.88 | EtT | 0.84 | LaC | 1.23 | EtB | 0.79 | My | 0.73 | EtB | 0.82 |
| EtT | 0.89 | My | 0.85 | My | 1.25 | EtT | 0.80 | EtB | 0.74 | Um | 0.87 |
| Os | 0.92 | Os | 0.90 | Os | 1.33 | Um | 0.90 | Um | 0.97 | EtT | 0.87 |
| Um | 1.10 | Um | 1.15 | Um | 1.51 | Cs | 0.94 | Lu | 1.03 | Cs | 0.87 |
| Lu | 1.41 | Lu | 1.38 | Lu | 1.83 | Lu | 1.25 | Cs | 1.12 | Lu | 1.40 |
| Hi | 1.45 | Hi | 1.47 | Hi | 1.93 | Hi | 1.34 | Hi | 1.17 | Hi | 1.46 |

There is also the question, which characters contribute most variance in the LDs. One possible measure of this is the dimensionless distance of a character from the origin of the ten-dimensional PC space. These data are presented in Table 18. From these data we can conclude that the largest contribution to the variance in the LDs have the characters

**Table 18.** Dimensionless distances of particular characters from the origin of the 10D PC space.

| Sign | App. dist. | Sign | App. dist. | Sign | App. dist. | Sign | App. dist. |
|---|---|---|---|---|---|---|---|
| z | 0.086 | u | 0.075 | d | 0.070 | ž | 0.065 |
| a | 0.085 | n | 0.073 | č | 0.068 | p | 0.065 |
| e | 0.082 | o | 0.073 | i | 0.068 | b | 0.064 |
| h | 0.081 | g | 0.072 | r | 0.066 | s | 0.064 |
| š | 0.076 | k | 0.072 | c | 0.066 | t | 0.062 |
| v | 0.076 | f | 0.070 | l | 0.066 | m | 0.058 |

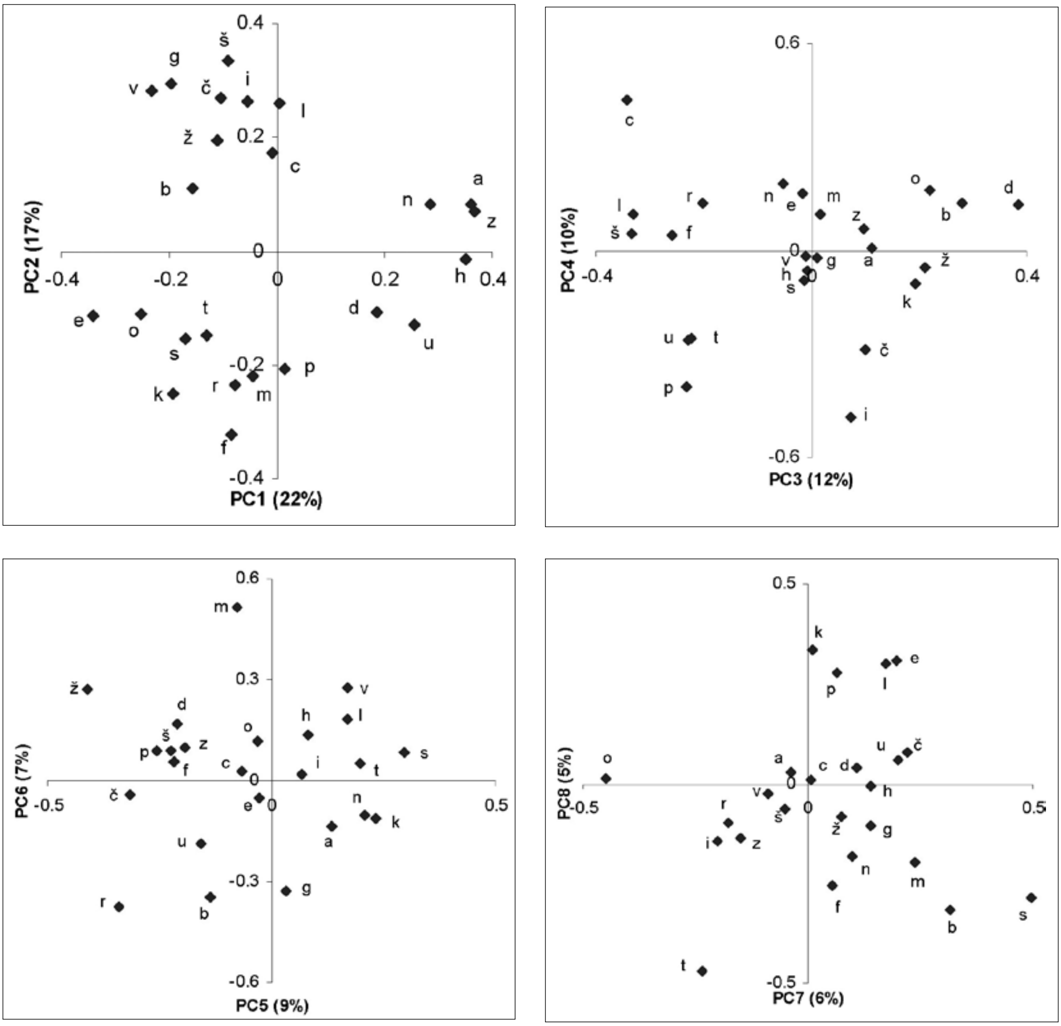


*Figure 14. Information regarding contribution of particular characters to the variance of the system in the first eight PC dimensions.*

z > a > e > h > š > v, and the least contribution have the characters m > t > s > b > p > ž. The difference in contribution is not big; the contribution of m is >2/3 of the contribution of z.

If we compare this with the average frequency of the characters, where the most frequent characters are a > i > e > t > u > n and the least frequent ones are z > c > š > f > č > ž, the a being more than 100 times more frequent than the ž, then wee see that there is little agreement between these series.

To illustrate these contributions, we present the score PCA plots, Figure 14. The characters placed most distant from the origin have the highest contribution to the resulting variance.

*Using frequencies of pairs of characters*

The amount of variance (%) contained in each PC axis is presented in Table 19. Whereas on using frequencies of single characters, Table 12, the variance was distributed in substantial amounts on several PC axes but in a clearly decreasing manner, on using the frequencies of pairs of characters the majority of variance is contained on the first PC axis. On the other PC axes there is contained much less of variance and its decrease is not as steep as in the former case. Cumulative variance in the former case (single characters) is PC1 22 %, PC4 60%, PC6 76%, PC8 87%, PC10 93%, whereas in the present case (pairs of characters) it is PC1 58%, PC4 72%, PC6 78%, PC8 83%, PC10 87%.

**Table 19.** Amount of variance contained on particular PC axes (#), % of total

| PC | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 | #11 | #12 | #13 | #14 | #15 | #16 | #17 | #18 | #19 | #20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (%) | 58.3 | 5.2 | 4.1 | 4.0 | 3.6 | 3.1 | 3.0 | 2.1 | 2.0 | 2.0 | 1.9 | 1.5 | 1.5 | 1.3 | 1.2 | 0.9 | 0.9 | 0.8 | 0.7 | 0.6 |

The dimensionless distance of a language from the origin of the multidimensional PC space does not increase noticeably going cumulatively from the axis PC1 to PC10. The only exception is Basque, where the axis PC4 contributes a substantial increase. The closest to the origin of the PC space is Basque, followed by Old Church Slavonic, Mycenean, classical Latin and Anatolian languages, whereas the most distant from the origin of PC space is Old Slovenian, followed by Rhaetic, Phrygian, Finnic, etc.

The results of PCA of frequencies of pairs of characters are presented in Figure 15.

We see that the axis PC1 separates first of all Basque from Old Church Slavonic, Anatolian languages, Mycenean, classical Latin, and Oscan, and these from all the other languages. The axis PC2 separates on one side Venetic and Venezian and on the other side the Anatolian languages, Mycenean and classical Latin from the others. The axis PC3 separates on one side Mycenean, semiclassical Latin and Oscan, and on the other side the Rhaetic and Etruscan, from the others. The axis PC4 separates first of all Basque from the other languages. The axis PC5 separates on one side Old Slovenian, Rhaetic and Umbrian, and on the other side the Anatolian languages, from the others. The axis PC6 separates first of all Old Church Slavonic, but also Estonian and Oscan, from the other languages.

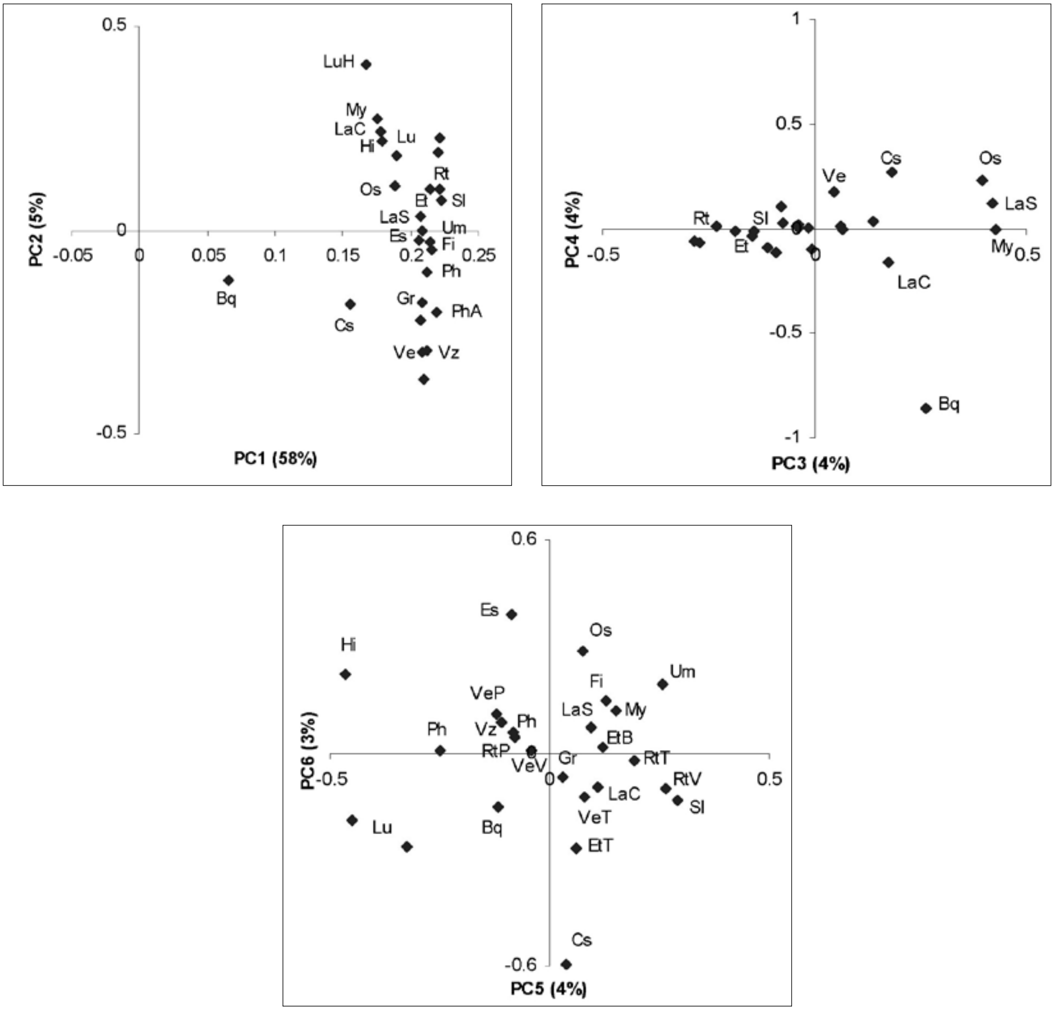


*Figure 15. Distribution of languages by pairs of characters in the PC space; presented by the PC axes PC1 and PC2, respectively PC3 and PC4, respectively PC5 and PC6.*

Considering the distance between languages in the PC space of frequencies of pairs of characters, the dimensionless distances between different presentations of the same language are as a rule not the smallest ones, contrary to the case when single characters were taken into account. The nearest neighbours are presented in Table 21. Using the frequencies of pairs of characters there is much more expressed different interpretation of sounds in the same language than if single characters are considered. Anyway, Table 22-24, it is interesting the closeness in position of Etruscan, Rhaetic, and Old Slovenian, but also Umbrian, Finnic, Estonian and Greek. The closeness in position is interesting as well in the case of Phrygian to Venezian, Greek, Venetic and also Finnic. For Venetic there is expressed the closeness to Venezian, Phrygian, and Greek. Among the other languages, there is interesting small distance between Finnic resp. Estonian, Umbrian and Old Slovenian.

**Table 20.** Dimensionless distances (×100) between languages in the ten-dimensional PC space of frequencies of the character pairs.

| | Bq | Cs | Es | EtB | EtT | Fi | Gr | Hi | LaC | LaS | Lu | LuH | My | Os | Ph | PhA | RtP | RtT | RtV | Sl | Um | VeT | VeP | VeV | Vz |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bq | 0 | | | | | | | | | | | | | | | | | | | | | | | | |
| Cs | 7.1 | 0 | | | | | | | | | | | | | | | | | | | | | | | |
| Es | 9.1 | 4.8 | 0 | | | | | | | | | | | | | | | | | | | | | | |
| EtB | 9.3 | 4.1 | 2.6 | 0 | | | | | | | | | | | | | | | | | | | | | |
| EtT | 9.6 | 4.7 | 1.8 | 1.4 | 0 | | | | | | | | | | | | | | | | | | | | |
| Fi | 9.6 | 4.8 | 1.3 | 2.2 | 1.5 | 0 | | | | | | | | | | | | | | | | | | | |
| Gr | 9.3 | 3.9 | 2.4 | 2.0 | 2.3 | 1.9 | 0 | | | | | | | | | | | | | | | | | | |
| Hi | 8.2 | 4.3 | 3.0 | 3.4 | 3.3 | 3.9 | 3.9 | 0 | | | | | | | | | | | | | | | | | |
| LaC | 7.8 | 4.3 | 3.6 | 3.2 | 3.3 | 3.6 | 3.5 | 3.6 | 0 | | | | | | | | | | | | | | | | |
| LaS | 9.3 | 4.1 | 2.9 | 2.9 | 2.7 | 2.8 | 2.4 | 3.6 | 3.3 | 0 | | | | | | | | | | | | | | | |
| Lu | 8.5 | 4.0 | 2.8 | 3.0 | 2.9 | 3.3 | 3.5 | 2.1 | 3.2 | 3.3 | 0 | | | | | | | | | | | | | | |
| LuH | 7.7 | 4.2 | 4.2 | 4.1 | 4.1 | 4.6 | 4.4 | 3.0 | 2.7 | 4.3 | 2.5 | 0 | | | | | | | | | | | | | |
| My | 7.8 | 4.2 | 3.7 | 4.0 | 3.8 | 3.9 | 3.6 | 3.7 | 1.9 | 2.9 | 3.6 | 2.8 | 0 | | | | | | | | | | | | |
| Os | 8.6 | 3.8 | 3.1 | 3.6 | 3.3 | 3.5 | 3.1 | 3.0 | 3.4 | 1.6 | 3.4 | 4.0 | 2.5 | 0 | | | | | | | | | | | |
| Ph | 9.4 | 4.6 | 2.0 | 2.4 | 2.2 | 1.8 | 1.9 | 3.7 | 3.2 | 2.4 | 2.8 | 4.4 | 3.8 | 3.4 | 0 | | | | | | | | | | |
| PhA | 9.8 | 4.7 | 2.1 | 2.3 | 2.4 | 2.0 | 1.4 | 3.7 | 4.1 | 3.0 | 3.2 | 4.6 | 4.3 | 3.8 | 1.9 | 0 | | | | | | | | | |
| RtP | 10.1 | 5.1 | 2.2 | 1.6 | 1.2 | 2.2 | 2.6 | 3.2 | 3.9 | 3.1 | 3.1 | 4.4 | 4.3 | 3.6 | 2.7 | 2.2 | 0 | | | | | | | | |
| RtT | 10.2 | 5.4 | 2.6 | 2.0 | 1.0 | 2.2 | 3.2 | 3.8 | 3.6 | 3.3 | 3.4 | 4.4 | 4.2 | 3.9 | 2.9 | 3.2 | 1.5 | 0 | | | | | | | |
| RtV | 10.1 | 5.2 | 2.7 | 2.1 | 1.1 | 2.2 | 3.2 | 4.0 | 3.8 | 3.3 | 3.4 | 4.5 | 4.2 | 3.9 | 3.0 | 3.2 | 1.8 | 0.6 | 0 | | | | | | |
| Sl | 10.1 | 4.9 | 2.6 | 2.1 | 1.2 | 1.8 | 2.7 | 4.2 | 3.7 | 2.9 | 3.5 | 4.6 | 4.1 | 3.7 | 2.6 | 2.8 | 2.0 | 1.3 | 0.9 | 0 | | | | | |
| Um | 9.5 | 4.7 | 1.7 | 2.2 | 1.4 | 1.2 | 2.0 | 3.9 | 3.4 | 2.3 | 3.5 | 4.6 | 3.5 | 3.0 | 2.2 | 2.4 | 2.1 | 2.0 | 1.9 | 1.5 | 0 | | | | |
| VeT | 9.4 | 3.8 | 2.7 | 2.2 | 2.3 | 2.4 | 1.9 | 3.9 | 3.6 | 2.3 | 3.3 | 4.7 | 4.0 | 3.2 | 1.5 | 2.2 | 2.8 | 3.1 | 3.1 | 2.5 | 2.3 | 0 | | | |
| VeP | 9.3 | 4.4 | 2.2 | 2.7 | 2.6 | 2.3 | 2.1 | 3.8 | 4.0 | 3.1 | 3.4 | 4.8 | 4.3 | 3.7 | 2.0 | 1.6 | 2.7 | 3.4 | 3.3 | 2.8 | 2.3 | 1.8 | 0 | | |
| VeV | 9.4 | 4.4 | 2.7 | 2.7 | 2.8 | 2.6 | 2.1 | 4.3 | 4.1 | 3.2 | 3.8 | 5.1 | 4.5 | 3.9 | 2.2 | 1.9 | 3.0 | 3.6 | 3.5 | 2.9 | 2.5 | 1.7 | 0.6 | 0 | |
| Vz | 9.4 | 4.5 | 2.3 | 2.5 | 2.5 | 2.3 | 1.8 | 3.9 | 4.0 | 3.0 | 3.5 | 4.8 | 4.3 | 3.7 | 2.1 | 1.3 | 2.5 | 3.3 | 3.3 | 2.8 | 2.2 | 1.8 | 0.6 | 0.7 | 0 |
| | Bq | Cs | Es | EtB | EtT | Fi | Gr | Hi | LaC | LaS | Lu | LuH | My | Os | Ph | PhA | RtP | RtT | RtV | Sl | Um | VeT | VeP | VeV | Vz |

**Table 21.** Dimensionless distances×100 between different presentations of the same language in the ten-dimensional PC space of the frequencies of pairs of characters.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| EtB | EtT | 1.4 | RtT | RtV | 0.6 | VeP | VeV | 0.6 |
| Ph | PhA | 1.9 | RtP | RtT | 1.5 | VeT | VeV | 1.7 |
| LaC | LaS | 3.3 | RtP | RtV | 1.8 | VeT | VeP | 1.8 |

**Table 22.** The distances×100 of some ancient languages to other ones in the ten-dimensional PC space of frequencies of pairs of characters

| EtT | | EtB | | Ph | | PhA | | RtT | | RtP | | RtV | | VeT | | VeP | | VeV | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RtT | 1.0 | RtP | 1.6 | VeT | 1.5 | Vz | 1.3 | EtT | 1.0 | EtT | 1.2 | Sl | 0.9 | Ph | 1.5 | Vz | 0.6 | Vz | 0.7 |
| RtV | 1.1 | Gr | 2.0 | Fi | 1.8 | Gr | 1.4 | Sl | 1.3 | EtB | 1.6 | EtT | 1.1 | Vz | 1.8 | PhA | 1.6 | PhA | 1.9 |
| RtP | 1.2 | RtT | 2.0 | Gr | 1.9 | VeP | 1.6 | Um | 2.0 | Sl | 2.0 | Um | 1.9 | Gr | 1.9 | Ph | 2.0 | Gr | 2.1 |
| Sl | 1.2 | Sl | 2.1 | Es | 2.0 | VeV | 1.9 | EtB | 2.0 | Um | 2.1 | EtB | 2.1 | EtB | 2.2 | Gr | 2.1 | Ph | 2.2 |
| Um | 1.4 | RtV | 2.1 | VeP | 2.0 | Fi | 2.0 | Fi | 2.2 | Fi | 2.2 | Fi | 2.2 | PhA | 2.2 | Es | 2.2 | Um | 2.5 |
| Fi | 1.5 | Um | 2.2 | Vz | 2.1 | Es | 2.1 | Es | 2.6 | Es | 2.2 | Es | 2.7 | Um | 2.3 | Um | 2.3 | Fi | 2.6 |
| Es | 1.8 | VeT | 2.2 | EtT | 2.2 | VeT | 2.2 | Ph | 2.9 | PhA | 2.2 | Ph | 3.0 | LaS | 2.3 | Fi | 2.3 | Es | 2.7 |
| Ph | 2.2 | Fi | 2.2 | Um | 2.2 | RtP | 2.2 | VeT | 3.1 | Vz | 2.5 | VeT | 3.1 | EtT | 2.3 | EtT | 2.6 | EtB | 2.7 |
| VeT | 2.3 | PhA | 2.3 | VeV | 2.2 | EtB | 2.3 | PhA | 3.2 | Gr | 2.6 | Gr | 3.2 | Fi | 2.4 | EtB | 2.7 | EtT | 2.8 |
| Gr | 2.3 | Ph | 2.4 | EtB | 2.4 | Um | 2.4 | Gr | 3.2 | Ph | 2.7 | PhA | 3.2 | Sl | 2.5 | RtP | 2.7 | Sl | 2.9 |
| PhA | 2.4 | Vz | 2.5 | LaS | 2.4 | EtT | 2.4 | Vz | 3.3 | VeP | 2.7 | Vz | 3.3 | Es | 2.7 | Sl | 2.8 | RtP | 3.0 |
| Vz | 2.5 | Es | 2.6 | Sl | 2.6 | Sl | 2.8 | LaS | 3.3 | VeT | 2.8 | LaS | 3.3 | RtP | 2.8 | LaS | 3.1 | LaS | 3.2 |
| VeP | 2.6 | VeP | 2.7 | RtP | 2.7 | LaS | 3.0 | VeP | 3.4 | VeV | 3.0 | VeP | 3.3 | RtV | 3.1 | RtV | 3.3 | RtV | 3.5 |
| LaS | 2.7 | VeV | 2.7 | Lu | 2.8 | RtT | 3.2 | Lu | 3.4 | LaS | 3.1 | Lu | 3.4 | RtT | 3.1 | RtT | 3.4 | RtT | 3.6 |
| VeV | 2.8 | LaS | 2.9 | RtT | 2.9 | Lu | 3.2 | VeV | 3.6 | Lu | 3.1 | VeV | 3.5 | Os | 3.2 | Lu | 3.4 | Lu | 3.8 |
| Lu | 2.9 | Lu | 3.0 | RtV | 3.0 | RtV | 3.2 | LaC | 3.6 | Hi | 3.2 | LaC | 3.8 | Lu | 3.3 | Os | 3.7 | Os | 3.9 |
| Os | 3.3 | LaC | 3.2 | LaC | 3.2 | Hi | 3.7 | Hi | 3.8 | Os | 3.6 | Os | 3.9 | LaC | 3.6 | Hi | 3.8 | LaC | 4.1 |
| LaC | 3.3 | Hi | 3.4 | Os | 3.4 | Os | 3.8 | Os | 3.9 | LaC | 3.9 | Hi | 4.0 | Cs | 3.8 | LaC | 4.0 | Hi | 4.3 |
| Hi | 3.3 | Os | 3.6 | Hi | 3.7 | LaC | 4.1 | My | 4.2 | My | 4.3 | My | 4.2 | Hi | 3.9 | My | 4.3 | Cs | 4.4 |
| My | 3.8 | My | 4.0 | My | 3.8 | My | 4.3 | Cs | 5.4 | Cs | 5.1 | Cs | 5.2 | My | 4.0 | Cs | 4.4 | My | 4.5 |
| Cs | 4.7 | Cs | 4.1 | Cs | 4.6 | Cs | 4.7 | Bq | 0.10.2 | Bq | 0.10.1 | Bq | 0.10.1 | Bq | 9.4 | Bq | 9.3 | Bq | 9.4 |
| Bq | 9.6 | Bq | 9.3 | Bq | 9.4 | Bq | 9.8 | | | | | | | | | | | | |

**Table 23.** The distances×100 of some classical languages in the ten-dimensional PC space of pairs odf characters

| Gr | | LaC | | LaS | | Os | | Um | | My | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PhA | 1.4 | My | 1.9 | Os | 1.6 | LaS | 1.6 | Fi | 1.2 | LaC | 1.9 |
| Vz | 1.8 | Lu | 3.2 | Um | 2.3 | My | 2.5 | EtT | 1.4 | Os | 2.5 |
| VeT | 1.9 | Ph | 3.2 | VeT | 2.3 | Um | 3.0 | Sl | 1.5 | LaS | 2.9 |
| Fi | 1.9 | EtB | 3.2 | Gr | 2.4 | Hi | 3.0 | Es | 1.7 | Um | 3.5 |
| Ph | 1.9 | EtT | 3.3 | Ph | 2.4 | Gr | 3.1 | RtV | 1.9 | Gr | 3.6 |
| EtB | 2.0 | LaS | 3.3 | EtT | 2.7 | Es | 3.1 | RtT | 2.0 | Lu | 3.6 |
| Um | 2.0 | Os | 3.4 | Fi | 2.8 | VeT | 3.2 | Gr | 2.0 | Hi | 3.7 |
| VeP | 2.1 | Um | 3.4 | Es | 2.9 | EtT | 3.3 | RtP | 2.1 | Es | 3.7 |
| VeV | 2.1 | Gr | 3.5 | Sl | 2.9 | Ph | 3.4 | EtB | 2.2 | Ph | 3.8 |
| EtT | 2.3 | Hi | 3.6 | EtB | 2.9 | LaC | 3.4 | Ph | 2.2 | EtT | 3.8 |
| Es | 2.4 | VeT | 3.6 | My | 2.9 | Lu | 3.4 | Vz | 2.2 | Fi | 3.9 |
| LaS | 2.4 | Es | 3.6 | Vz | 3.0 | Fi | 3.5 | VeT | 2.3 | EtB | 4.0 |
| RtP | 2.6 | RtT | 3.6 | PhA | 3.0 | EtB | 3.6 | VeP | 2.3 | VeT | 4.0 |

**Table 23.** Continued

| Gr | | LaC | | LaS | | Os | | Um | | My | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sl | 2.7 | Fi | 3.6 | VeP | 3.1 | RtP | 3.6 | LaS | 2.3 | Sl | 4.1 |
| Os | 3.1 | Sl | 3.7 | RtP | 3.1 | Vz | 3.7 | PhA | 2.4 | Cs | 4.2 |
| RtV | 3.2 | RtV | 3.8 | VeV | 3.2 | Sl | 3.7 | VeV | 2.5 | RtT | 4.2 |
| RtT | 3.2 | RtP | 3.9 | Lu | 3.3 | VeP | 3.7 | Os | 3.0 | RtV | 4.2 |
| Lu | 3.5 | VeP | 4.0 | RtV | 3.3 | PhA | 3.8 | LaC | 3.4 | RtP | 4.3 |
| LaC | 3.5 | Vz | 4.0 | LaC | 3.3 | Cs | 3.8 | Lu | 3.5 | Vz | 4.3 |
| My | 3.6 | PhA | 4.1 | RtT | 3.3 | RtT | 3.9 | My | 3.5 | PhA | 4.3 |
| Hi | 3.9 | VeV | 4.1 | Hi | 3.6 | RtV | 3.9 | Hi | 3.9 | VeP | 4.3 |
| Cs | 3.9 | Cs | 4.3 | Cs | 4.1 | VeV | 3.9 | Cs | 4.7 | VeV | 4.5 |
| Bq | 9.3 | Bq | 7.8 | Bq | 9.3 | Bq | 8.6 | Bq | 9.5 | Bq | 7.8 |

**Table 24.** The distances×100 of some reference languages in the ten-dimensional PC space of pairs of characters

| Bq | | Cs | | Sl | | Es | | Fi | | Vz | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cs | 7.1 | VeT | 3.8 | RtV | 0.9 | Fi | 1.3 | Um | 1.2 | VeP | 0.6 |
| LaC | 7.8 | Os | 3.8 | EtT | 1.2 | Um | 1.7 | Es | 1.3 | VeV | 0.7 |
| My | 7.8 | Gr | 3.9 | RtT | 1.3 | EtT | 1.8 | EtT | 1.5 | PhA | 1.3 |
| Hi | 8.2 | Lu | 4.0 | Um | 1.5 | Ph | 2.0 | Sl | 1.8 | Gr | 1.8 |
| Lu | 8.5 | LaS | 4.1 | Fi | 1.8 | PhA | 2.1 | Ph | 1.8 | VeT | 1.8 |
| Os | 8.6 | EtB | 4.1 | RtP | 2.0 | VeP | 2.2 | Gr | 1.9 | Ph | 2.1 |
| Es | 9.1 | My | 4.2 | EtB | 2.1 | RtP | 2.2 | PhA | 2.0 | Um | 2.2 |
| EtB | 9.3 | Hi | 4.3 | VeT | 2.5 | Vz | 2.3 | RtV | 2.2 | Es | 2.3 |
| Gr | 9.3 | LaC | 4.3 | Es | 2.6 | Gr | 2.4 | RtP | 2.2 | Fi | 2.3 |
| LaS | 9.3 | VeV | 4.4 | Ph | 2.6 | Sl | 2.6 | EtB | 2.2 | EtB | 2.5 |
| VeP | 9.3 | VeP | 4.4 | Gr | 2.7 | RtT | 2.6 | RtT | 2.2 | EtT | 2.5 |
| VeV | 9.4 | Vz | 4.5 | Vz | 2.8 | EtB | 2.6 | Vz | 2.3 | RtP | 2.5 |
| VeT | 9.4 | Ph | 4.6 | PhA | 2.8 | VeT | 2.7 | VeP | 2.3 | Sl | 2.8 |
| Vz | 9.4 | PhA | 4.7 | VeP | 2.8 | RtV | 2.7 | VeT | 2.4 | LaS | 3.0 |
| Ph | 9.4 | EtT | 4.7 | LaS | 2.9 | VeV | 2.7 | VeV | 2.6 | RtV | 3.3 |
| Um | 9.5 | Um | 4.7 | VeV | 2.9 | Lu | 2.8 | LaS | 2.8 | RtT | 3.3 |
| Fi | 9.6 | Es | 4.8 | Lu | 3.5 | LaS | 2.9 | Lu | 3.3 | Lu | 3.5 |
| EtT | 9.6 | Fi | 4.8 | Os | 3.7 | Hi | 3.0 | Os | 3.5 | Os | 3.7 |
| PhA | 9.8 | Sl | 4.9 | LaC | 3.7 | Os | 3.1 | LaC | 3.6 | Hi | 3.9 |
| RtP | 0.10.1 | RtP | 5.1 | My | 4.1 | LaC | 3.6 | My | 3.9 | LaC | 4.0 |
| RtV | 0.10.1 | RtV | 5.2 | Hi | 4.2 | My | 3.7 | Hi | 3.9 | My | 4.3 |
| Sl | 0.10.1 | RtT | 5.4 | Cs | 4.9 | Cs | 4.8 | Cs | 4.8 | Cs | 4.5 |
| RtT | 0.10.2 | Bq | 7.1 | Bq | 0.10.1 | Bq | 9.1 | Bq | 9.6 | Bq | 9.4 |

The largest distance to the other languages have the Basque and Old Church Slavonic.

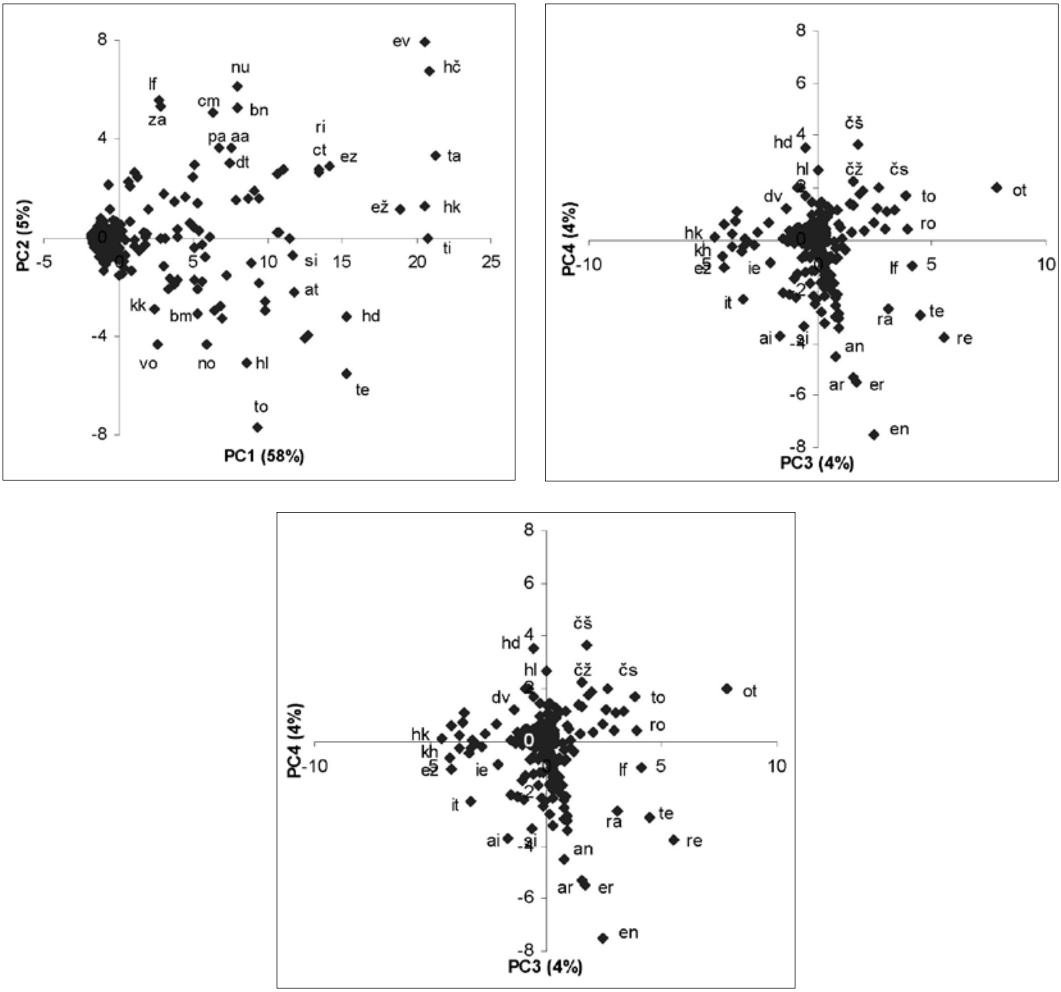


*Figure 16. Information regarding contribution of particular pairs of characters to the variance of the system in the first six PC dimensions.*

Regarding the pairs of characters, Table 25 together with Figure 16 shows that to the variance of data contribute the most the frequencies of pairs of characters ta, hč, ti, ev, hk, and ež.

The majority of pairs of characters are clustered near the origin of the PC space. This means that their frequencies in the languages in question contribute little to the total information contained in the studied system.

**Table 25.** Dimensionless distances from the origin of the 10D PC space of those pairs of characters, which contribute most to the variance of the system.

| Pair | App. dist. | Pair | App. dist. | Pair | App. dist. | Pair | App. dist. |
|---|---|---|---|---|---|---|---|
| ta | 12.42 | ri | 7.83 | re | 5.76 | nu | 4.66 |
| hč | 12.19 | st | 7.43 | čš | 5.75 | sa | 4.59 |
| ti | 12.09 | ga | 7.33 | se | 5.51 | aa | 4.43 |
| ev | 12.01 | at | 6.87 | ra | 5.50 | dt | 4.38 |
| hk | 12.01 | si | 6.84 | to | 5.48 | ia | 4.22 |
| ež | 11.03 | dz | 6.68 | čs | 5.31 | vi | 4.05 |
| te | 8.97 | ai | 6.47 | dv | 5.23 | kh | 4.02 |
| hd | 8.91 | hm | 6.31 | ae | 5.08 | pa | 3.97 |
| ez | 8.30 | tu | 6.22 | hl | 5.03 | ve | 3.75 |
| čt | 7.85 | ie | 6.20 | bn | 4.69 | cm | 3.71 |

*Using frequencies of triplets of characters*

The amount of variance (%) contained in each PC axis is presented in Table 26. Whereas on using frequencies of single characters, Table 12, the variance is distributed in substantial amounts on several PC axes but in a clearly decreasing manner, and on using the frequencies of pairs of characters, Table 19, the majority of variance is contained on the first PC axis, on using the frequencies of triplets of characters the variance is distributed quite evenly on all PC axes with only a slight decrease towards the higher ones. One quarter of cumulative variance is contained till the axis PC6, one half till the axis PC12, two-thirds till the axis PC16, etc.

**Table 26.** Amount of variance contained on particular PC axes (#), % of total

| PC | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 | #11 | #12 | #13 | #14 | #15 | #16 | #17 | #18 | #19 | #20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (%) | 5.0 | 4.6 | 4.4 | 4.4 | 4.2 | 4.2 | 4.2 | 4.1 | 4.1 | 4.0 | 4.0 | 4.0 | 4.0 | 3.9 | 3.9 | 3.9 | 3.8 | 3.8 | 3.8 | 3.8 |

The dimensionless distance of a language from the origin of the multidimensional PC space increases gradually, Figure 17, and it does not level off on different levels as in Figure 12, but the distance of all languages converges to the almost same value. The standard deviation between languages, Figure 18, increases till the fifth PC axis, where less than one quarter of cumulative variance is observed, then it decreases.

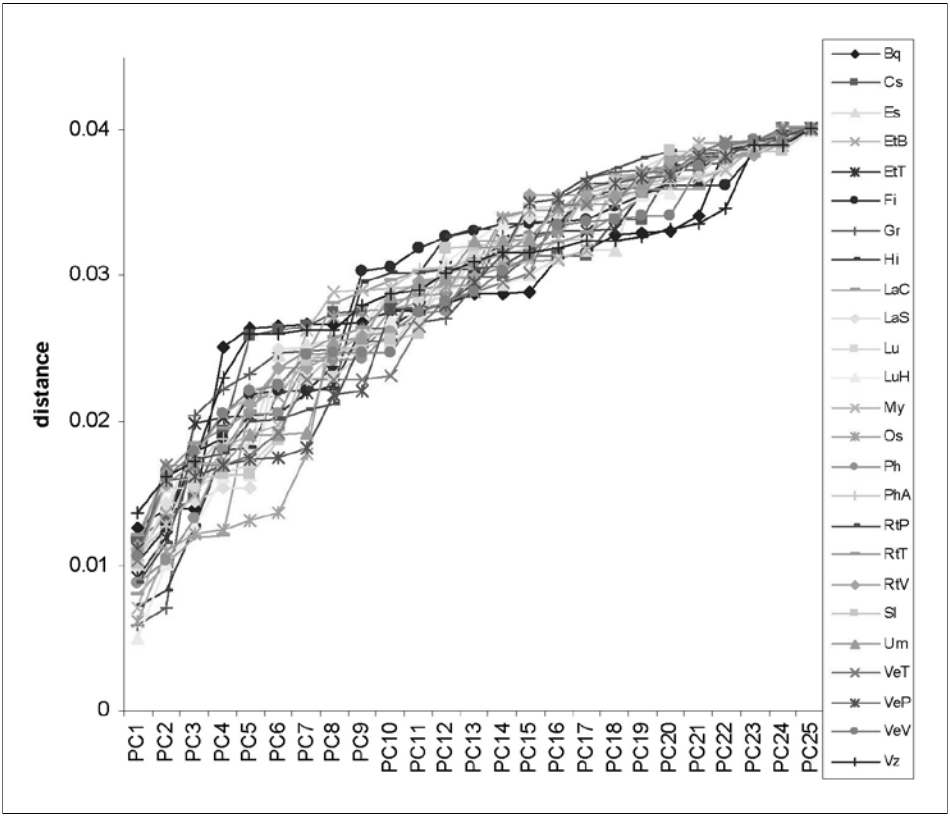


*Figure 17. Dimensionless distance of languages from the PC origin using the frequencies of triplets of characters*

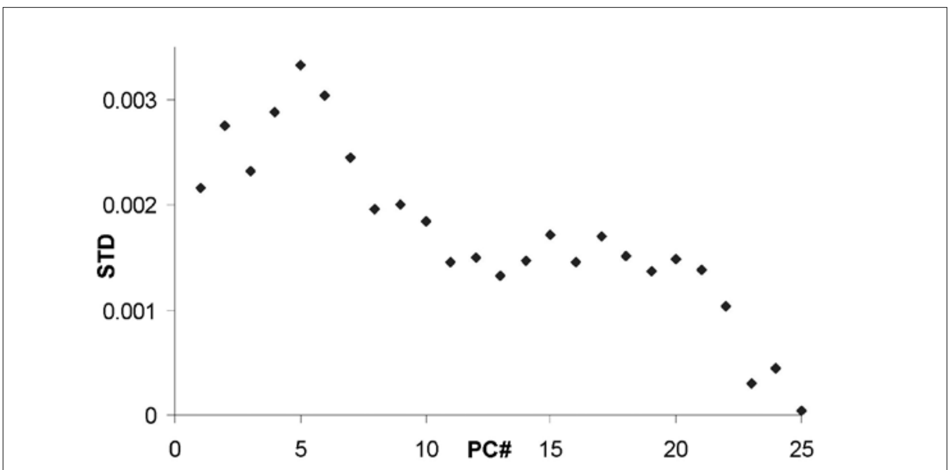


*Figure 18. Standard deviation between languages in Figure 17.*

*Last character in the word*

Besides the characters within the words, there can be used also data about the last character in the words where they are known or reasonably supposed. The results of PCA of these data are presented below. In Table 27 is presented the amount of variance contained on particular PC axes when taken into account only the last character in a word. The variance content is spread among a number of PC axes and it indicates that at least eight axes are to be taken into account.

**Table 27.** Amount of variance contained on particular PC axes (#), % of total

| PC | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 | #11 | #12 | #13 | #14 | #15 | #16 | #17 | #18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (%) | 17.8 | 14.2 | 12.4 | 10.3 | 8.0 | 7.1 | 6.7 | 4.9 | 4.1 | 3.8 | 2.5 | 2.5 | 1.9 | 1.3 | 1.0 | 0.8 | 0.5 | 0.2 |

In Figure 19 is presented the dimensionless distance of particular languages from the origin of this PC space formed by considering the data about the last character in a word. Also here the distances level off when approaching the axis PC10, therefore this one is taken as the limit also here.

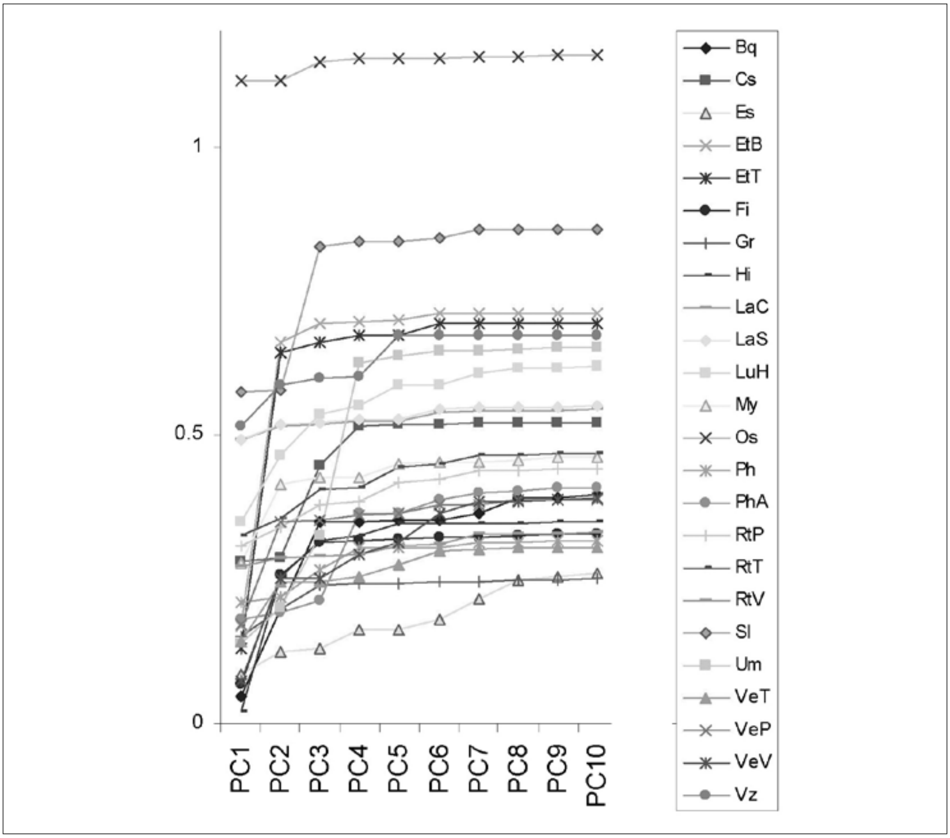


*Figure 19. Last character in a word - dimensionless distance of languages from the PC origin*

The spread of languages in the present PC space is illustrated in Figure 20. The axis PC1 separates the most Old Slovenian from Oscan, whereas the axis PC2 separates first of all Etruscan from Mycenean and Luvian. The axis PC3 separates first of all Slovenian from Basque and Luvian, whereas the axis PC4 separates first of all Umbrian from Old Phrygian and Old Church Slavonic. The axis PC5 separates first of all Venezian from Luvian and Rhaetic, whereas the axis PC6 separates first of all Etruscan from Venetic as well as their variant presentations between them. The axis PC7 separates first of all Slovenian and Luvian from Estonian, Rhaetic and Phrygian, whereas the axis PC8 separates first of all Estonian and Luvian from Basque.

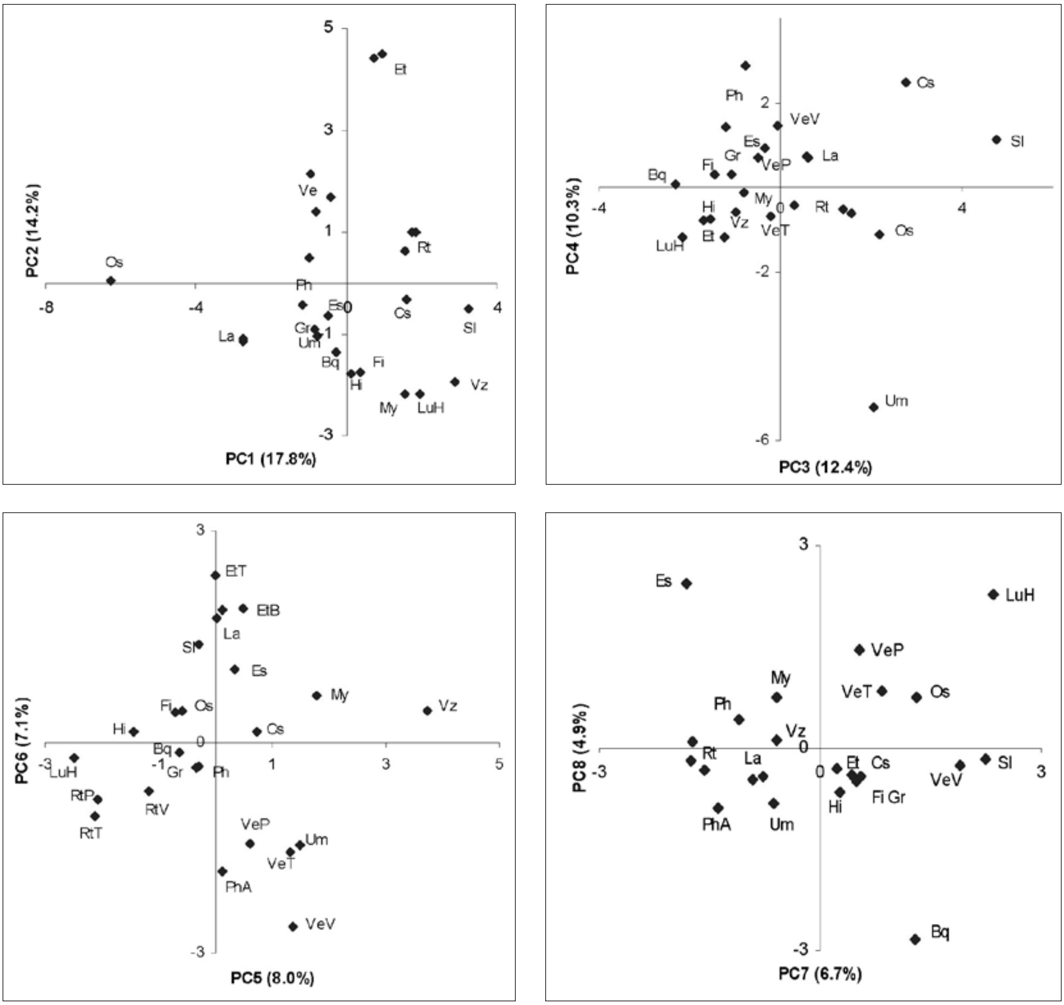


*Figure 20. Last character in a word - information presented by the first eight PC axes.*

Considering the dimensionless distances in the ten-dimensional PC space, different presentations of particular languages are not distant from one another, Table 28.

**Table 28.** Dimensionless distances in the ten-dimensional PC space between different presentations of the same language

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| LaC | LaS | 0.02 | RtP | RtT | 0.04 | VeT | VeP | 0.20 |
| EtB | EtT | 0.10 | RtP | RtV | 0.17 | VeP | VeV | 0.23 |
| Ph | PhA | 0.25 | RtT | RtV | 0.20 | VeT | VeV | 0.27 |

The smallest average distances between different languages, however, are in all cases higher, Table 29. Interesting is the closeness of Old Phrygian to Greek and Estonian, of Venetic to Old Phrygian, Greek and Estonian, of Rhaetic to Old Church Slavonic, Estonian, Finnic and Mycenean, as well as of Etruscan to Venetic and Rhaetic.

**Table 29.** The smallest average distances of some ancient languages in the ten-dimensional PC space

| Etruscan to | | O.Phrygian to | | Rhaetic to | | Venetic to | |
|---|---|---|---|---|---|---|---|
| Venetic | 0.62 | Greek | 0.30 | OChSl | 0.50 | O.Phrygian | 0.42 |
| Rhaetic | 0.70 | Estonian | 0.32 | Estonian | 0.57 | Greek | 0.45 |
| O.Phrygian | 0.84 | Venetic | 0.42 | Finnic | 0.58 | Estonian | 0.49 |
| O. Greek | 0.86 | Basque | 0.45 | Mycenean | 0.59 | Basque | 0.60 |
| Basque | 0.90 | Finnic | 0.47 | Venetic | 0.60 | Rhaetic | 0.60 |
| Finnic | 0.91 | Latin | 0.49 | Hittite | 0.60 | Finnic | 0.62 |
| Hittite | 0.92 | Hittite | 0.50 | Greek | 0.63 | Hittite | 0.63 |
| O.Ch.Sl. | 0.95 | Rhaetic | 0.67 | O.Phrygian | 0.67 | Latin | 0.65 |
| Mycenean | 0.98 | Mycenean | 0.67 | O. Slovenian | 0.69 | O.Ch.Sl. | 0.69 |
| | | OChSl | 0.71 | Basque | 0.70 | Mycenean | 0.74 |
| | | Umbrian | 0.88 | Etruscan | 0.70 | Umbrian | 0.79 |
| | | Venezian | 0.89 | Venezian | 0.70 | Venezian | 0.90 |
| | | | | Umbrian | 0.78 | | |
| | | | | Latin | 0.90 | | |

**Table 30.** The smallest (average) distances of some other languages in the ten-dimensional PC space

| O.Slovenian to | | Venezian to | | Latin to | | Greek to | | Mycenean to | |
|---|---|---|---|---|---|---|---|---|---|
| OChSl | 0.45 | Mycenean | 0.30 | Greek | 0.45 | Basque | 0.25 | Venezian | 0.30 |
| Mycenean | 0.85 | Finnic | 0.59 | Estonian | 0.48 | Finnic | 0.26 | Finnic | 0.32 |
| Venezian | 0.86 | Hittite | 0.65 | Hittite | 0.63 | Hittite | 0.27 | Hittite | 0.41 |
| Finnic | 0.96 | OChSl | 0.70 | Basque | 0.64 | Estonian | 0.29 | Estonian | 0.49 |
| Estonian | 0.97 | Basque | 0.72 | Finnic | 0.64 | Latin | 0.45 | Basque | 0.50 |
| | | Estonian | 0.73 | Oscan | 0.73 | Mycenean | 0.51 | Greek | 0.51 |
| | | Greek | 0.77 | Umbrian | 0.78 | OChSl | 0.70 | OChSl | 0.60 |
| | | O.Slovenian | 0.86 | Mycenean | 0.83 | Venezian | 0.77 | Umbrian | 0.80 |
| | | Umbrian | 0.93 | OChSl | 0.86 | | | Latin | 0.83 |
| | | | | | | | | O.Slovenian | 0.85 |

**Table 31.** Dimensionless distance of particular last characters in words, from the origin of the 10D PC space.

| Sign | App. dist. | Sign | App. dist. | Sign | App. dist. | Sign | App. dist. |
|---|---|---|---|---|---|---|---|
| m | 0.073 | a | 0.068 | u | 0.066 | r | 0.058 |
| š | 0.073 | c | 0.068 | d | 0.065 | k | 0.057 |
| g | 0.072 | h | 0.068 | č | 0.064 | i | 0.056 |
| t | 0.070 | ž | 0.067 | e | 0.059 | b | 0.056 |
| s | 0.069 | f | 0.067 | v | 0.058 | o | 0.055 |
| l | 0.069 | p | 0.066 | n | 0.058 | z | 0.053 |

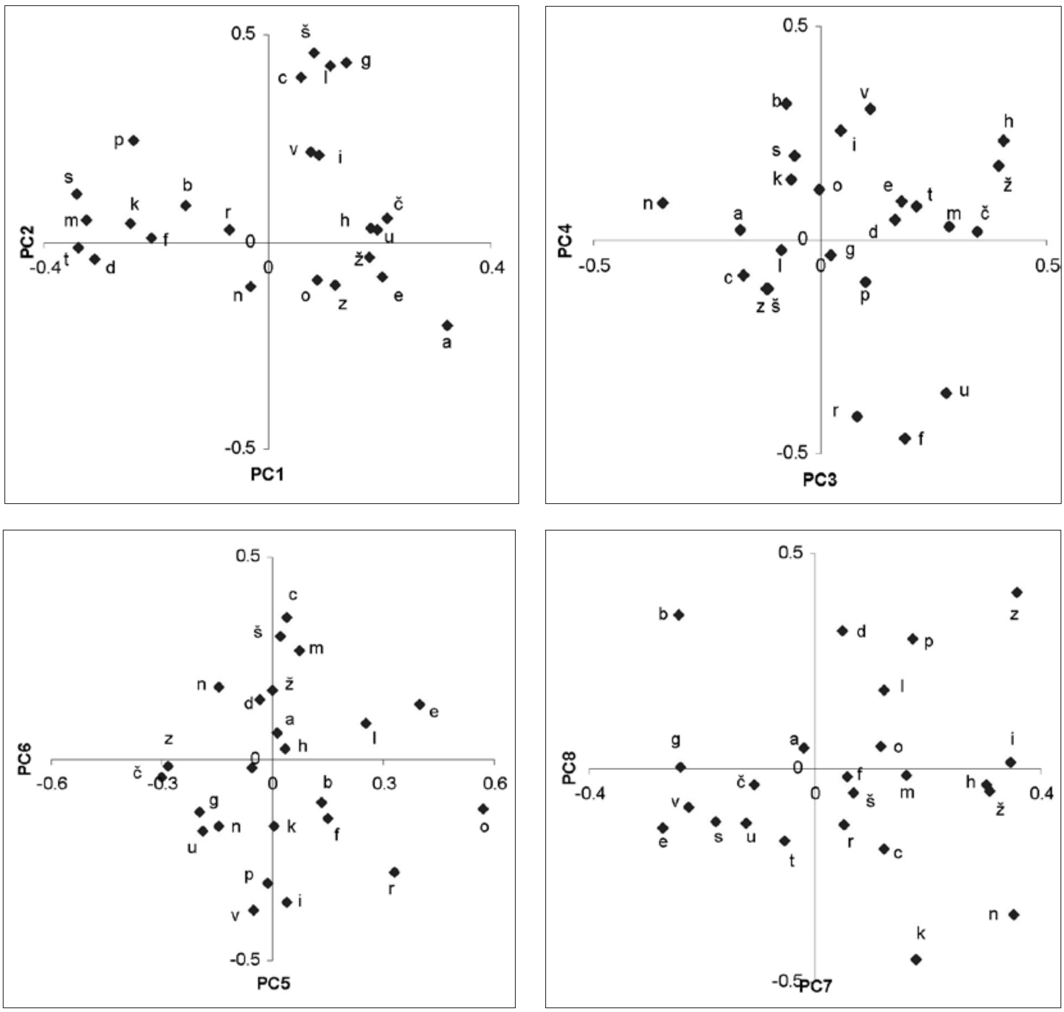


*Figure 21. Information regarding contribution of particular characters to the variance of the system in the first eight PC dimensions.*

The smallest distances among some other languages, Table 30, are e.g. of Old Slovenian to Old Church Slavonic, Venezian and Mycenean; of Latin to Greek and Estonian; of Greek to Basque, Finnic, Hittite and Estonian, as well as of Mycenean to Venezian and Finnic.

In Table 31 there are presented the distances of particular characters from the origin of the PC space. There is not great difference in these distances. Comparing these distances to the highest frequencies of the last characters in the words, which are on average:

a > i > e > s > n > o > u > t > r > m,

which are to be compared with those having here the largest distances and thus contributing the most to the variance of the system:

m > š > g > t > s > l > a > c > h > ž.

We can see in these series little similarity.

In Figure 21 are presented the illustrations of these distances in subsequent pairs of PC axes.

## Discussion

The impetus for the present study was the result of previous attempts to classify the Venetic and Rhaetic language [8,9]. This result indicated that by the linguistic distance based on the mean vowel and mean consonant frequency, Venetic and Rhaetic were closer to Old Slovenian than to Latin, what contradicts the assertion of Lejeune [12]. In order to generate additional significant data and thus to provide more reliable results, this present study includes more languages and additional methodologies.

Besides Venetic and Rhaetic we also included also Etruscan and Old Phrygian. Besides Old Slovenian also Old Church Slavonic was compared. Besides Latin also one of the oldest Greek texts was analysed, together with Oscan, Umbrian, and Mycenean. Since Etruscan may be a transplant from Anatolia or its vicinity, Hittite and Luvian were also included. We have at our disposal also texts in the dialect spoken now in the territory that was formerly Venetic. The history of this territory is well documented, whereas the contemporary dialect known as Venezian belongs to the Romance group. Of the non-Indo-European languages, Basque, Estonian, and Finnic were included for comparison. We also introduce different ways of reading Etruscan, Latin, Old Phrygian, Rhaetic, and Venetic. To all texts taken into consideration, a common notation system was applied in order to assure the applicability of the methods used in the study. The essence of the common notation system is to present different sound values of the same vowel or consonant by one sign only, as well as to join the vowels and semivowels. We are aware of the imperfection of such a common notation system; however, this is at present the best common denominator we know for our purpose and we are cognizant of it also when interpreting the results.

Regarding the methods, besides some unidimensional approaches we also use the multidimensional Principle Component Analysis. For easier understanding of its results we present them in Figures showing their several dimensions as well as in their unidimensional summaries - the dimensionless distance of a language from the origin of the PC space as well as the dimensionless distances between the languages in question. The PCA results are very appropriate for this purpose since the PC axes are orthogonal to each other and thus a simple Pythagorean type of calculation of mentioned distances is possible.

For a realistic classification of some old languages, like the Venetic, Rhaetic and Etruscan there are not available sufficient data to do that in a proper manner. For the purpose of classification of languages into language families are normally used the agreement in grammatical structure and in the language material which bears the structure [56], p. 6. The inscriptions in said languages are mostly short, broken or incomplete, making the extraction of needed data difficult or impossible. Even the sound value of some characters in them is still debatable. For these reasons, we limited ourselves to the comparison of the character structures in these languages transcribed into a common notation system and where more versions of interpretation were known, we took also them into comparison. Other, better-known languages were transcribed into the same notation system to enable their use in this comparison.

For the purpose of comparison we used unidimensional and multidimensional methods. The simplest unidimensional approach is to compare the frequency of particular characters used to notate particular or sufficiently similar sounds. It is followed by various ratios, like the vowel-to-consonant ratio, etc.

As the multidimensional method we used the PCA of the frequencies of particular characters respectively last characters in the words, as well as of pairs and triplets of characters. The selectivity of this approach is indicated in Table 32 as the ratio of the largest and the smallest PC distance of a language from the others.

It is obvious that in our case the PCA of the frequency of character triplets does not contribute any useful information. Regarding the single characters and the pairs of them, we have two situations. Among the Basque, Old Church Slavonic, Old Greek, Latin, Mycenean and Oscan, the selectivity is higher using frequencies of single characters. Using frequencies of character pairs, the selectivity is higher among other languages.

The results presented in the chapter Results allow the following insights.

**Table 32.** Ratio of the largest and the smallest dimensionless distance of a language to the other languages in question in the PC space.

| char. | single | pair | triplet | char. | single | pair | triplet | char | single | pair | triplet |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PC1-.. | 10 | 10 | 25 | PC1-.. | 10 | 10 | 25 | PC1-.. | 10 | 10 | 25 |
| same* | 10.5 | 5.4 | 1.0 | EtB | 2.0 | 5.8 | 1.1 | Es | 4.1 | 6.8 | 1.1 |
| Bq | 3.3 | 1.4 | 1.1 | EtT | 2.2 | 9.7 | 1.1 | Sl | 3.1 | 10.7 | 1.1 |
| Cs | 3.8 | 1.9 | 1.1 | Fi | 3.4 | 7.8 | 1.1 | Um | 3.5 | 7.7 | 1.1 |
| Gr | 6.9 | 6.5 | 1.1 | Ph | 5.0 | 6.1 | 1.1 | VeT | 4.6 | 6.1 | 1.1 |
| LaC | 6.2 | 4.0 | 1.1 | PhA | 5.3 | 7.3 | 1.1 | VeP | 5.5 | 16.6 | 1.0 |
| LaS | 7.2 | 5.8 | 1.1 | RtP | 3.1 | 8.6 | 1.1 | VeV | 5.2 | 13.3 | 1.0 |
| My | 4.1 | 4.1 | 1.1 | RtT | 3.1 | 10.2 | 1.1 | Vz | 4.2 | 16.8 | 1.1 |
| Os | 5.6 | 5.3 | 1.0 | RtV | 3.2 | 10.6 | 1.1 | | | | |

same*: Versions of the same language, cf. elsewhere in the Table.

Regarding the most frequent vowel, Table 2, close to one another are:

- Basque, Venezian, Etruscan, Old Phrygian, Luvian, Mycenean, and Hittite;
- Estonian, Greek and Umbrian;
- Rhaetic, Finnic, Old Slovenian, Old Church Slavonic, Latin and Oscan.

Regarding the most frequent consonant, Table 2, there can be put together:

- Basque, Venezian, Etruscan read in the Bor's [32] way, Luvian, Hittite, Greek, and Finnic;
- Mycenean and Umbrian;
- Old Phrygian, Estonian, Oscan;
- Etruscan read in the Pallottino's [44] way, Latin, Rhaetic, Old Slovenian and Old Church Slavonic, Venetic.

The most frequent pairs of vowels, of vowel-consonant, and consonant-vowel, Tables 5-7, do not give any clear clue. The same holds true for the most frequent vowel triplets.

Regarding the vowel-to-consonant ratio, Figure 2, Etruscan is either placed separately from the other languages or together with Latin and Umbrian. Rhaetic is placed together with Basque, Greek, Old Phrygian and Luvian, whereas Venetic is placed in the cluster with Estonian, Old Church Slavonic, Hittite, Venezian, Finnic, and Old Slovenian.

The K/S ratio must not be confused with the Kentum/Satem division, which has a different basis. The Kentum/Satem characteristics are, however, part of the K/S ratio. The sound k like sounds are prevailing over sibilants and affricate especially in Mycenean, followed by Finnic and Estonian, cf. Figure 3. The reverse is true especially in Etruscan, and also in Luvian, Umbrian, Old Church Slavonic, Old Slovenian, etc. Rhaetic is placed by this criterion together with Oscan, Basque and Latin, whereas Venetic is placed together with Venezian and Finnic. In combination of these sounds with vowels, Figure 4, the position of Venetic versions is governed mainly by the direction of reading the AKEO, therefore only the position of VeV is of some diagnostic value. The most selective is the combination with a, o, and u. In Latin, Mycenean, Estonian, Greek and Finnic the sounds k, g, h prevail in all cases. Close to them are Phrygian, Venetic, and Hittite. In all tested combinations, in Etruscan, Old Church Slavonic, Umbrian, Old Slovenian, Venezian and Basque the sibilants and affricate prevail over k, g, h sounds.

The frequency of the last character in a word is of interest as well, since it reflects also some grammatical features. Their determination is straightforward in languages known in detail, while it may be only a supposition for inscriptions written *in continuo*. In these cases it is especially dangereous that a continuous text would be divided into words due to some suppositions based on one or another well known language. In such situations it is advisable to confront different approaches to the decipherment not only between them but also to results of other independent examinations. In our case this problem is the most evident in Venetic and Etruscan, Figure 7, where the basis for division of continuous text into words appreciably influences the result.

In our databases the frequency of the last character in a word is on average a > i > e > s > n > o > u > t > r > m > others.

**Table 33.** The smallest weighted average dimensionless distances between tested languages.

a. Some ancient languages

| | |
|---|---|
| Etruscan | Rhaetic << Finnic ~ Old Slovenian < Estonian < Greek < Venetic, etc. |
| Old Phrygian | Venetic < Venezian ~ Estonian < Greek << Finnic << Old Slovenian, etc. |
| Rhaetic | Old Slovenian < Etruscan < Finnic < Estonian < Venetic < Old Phrygian, etc. |
| Venetic | Venezian < Old Phrygian << Estonian < Greek < Old Slovenian < Finnic, etc. |

b. Reference languages

| | |
|---|---|
| Latin | Oscan < Greek < Mycenean < Umbrian < Estonian < Venezian, etc. |
| Oscan | Latin < Mycenean < Greek < Umbrian < Estonian < Finnic, etc. |
| Umbrian | Latin < Greek < Oscan < Finnic < Estonian < Mycenean, etc. |
| Greek | Venezian < Latin < Estonian < Old Phrygian < Finnic < Oscan, etc. |
| Mycenean | Latin < Oscan < Greek < Umbrian < Estonian < Venezian, etc. |
| Old Church Slavonic | Old Slovenian < Venetic < Venezian < Old Phrygian < Greek, etc. |
| Old Slovenian | Rhaetic << Venetic ~ Venezian ~ Estonian < Old Phrygian, etc. |
| Estonian | Finnic < Old Phrygian < Venezian ~ Greek < Venetic < Old Slov., etc. |
| Finnic | Estonian << Greek << Old Phrygian < Rhaetic < Venetic ~ Venezian, etc. |
| Venezian | Venetic < Greek < Old Phrygian < Estonian << Old Slovenian < Finnic, etc. |
| Basque | Venezian < Estonian < Venetic < Old Prygian < Old Ch. Slavonic, etc. |

The PCA results give a quantity of data. Taking averages of different readings of some of the languages taken into account and giving equal weight to results derived from frequencies of single characters as well as to those of pairs of them, the linguistic distances are increasing in the following series, Table 33.

The results of present study thus confirm the previous [8,9] conclusions that by their sound structure, Venetic and Rhaetic are closer to Old Slovenian than to Latin. In all tested ways of reading Rhaetic inscriptions, Rhaetic is the closest to Old Slovenian, followed by Etruscan, etc. Also at Venetic, different ways of reading do not give rise to appreciably different results. In respect to Latin, Venetic read in any tested way, even in the LLV [10] way, is closer to Semiclassical Latin than to the Classical Latin, although by its age it is contemporary with the latter and not with the former. Here arises the question whether Venetic influenced Latin to change from the classic to semiclassic pronunciation.

Old Phrygian is also close to Venetic and Rhaetic, in line with the previous observation by Ambrozic [68], pp 5-57. In both ways of reading it is the closest to Venetic. In respect to Latin, Old Phrygian read in any tested way is closer to Semiclassical Latin than to the Classical Latin.

Etruscan is by present results also close to the above group, regardless whether it is read in the Pallottino's [44] way or Bor's [13,32] way. Significantly, it is not close to Hittite and Luvian, from which it might have derived or to its neighbour Old Italic languages. Regarding Etruscan, there should be taken seriously the observation by Bor [13], p. 344; [32], p. 11, that he was able to decipher the older Etruscan inscriptions but not the younger ones. Thusly, for Etruscan an additional study would be needed, where the Etruscan inscriptions would be divided into several groups by their origin and age, and then to repeat the study.

From the above results follows that it is legitimate to use Slovenian with its archaic dialects as a catalyst in deciphering the Rhaetic, Venetic, older Etruscan, Old Phrygian and possibly also other old inscriptions.

The Pääbo's approach to decipher the Venetic inscriptions using Estonian as a catalyst [80] is by present results legitimate as well. It must however sustain the criticism directed to it [81], in order to prove acceptable.

Regarding the reference languages, the PC distances between the reference languages indicate that Latin is the closest to Greek (cf. [56], p. 2), as well as to Oscan, Umbrian, Mycenean and Estonian. By our results, Latin, Oscan, and Umbrian form a different cluster than the Etruscan, Rhaetic and Venetic. Mycenean belongs close to the cluster of Latin, Oscan, and Umbrian, as well.

Estonian is close to Finnic, but also to Old Phrygian, Venezian and Greek, whereas Finnic is close to Estonian, Greek and Old Phrygian. Old Slovenian is close to Rhaetic, Venetic, Venezian, Estonian and Old Phrygian. Old Church Slavonic indicates some closeness besides to Old Slovenian also to Venetic. Basque and Old Anatolian languages are quite distant from all other tested languages.

Surprisingly, Venezian, being a present Romanic dialect on the previous Venetic territory, by its sound system is closer to the ancient Venetic as well as to Old Slovenian than to Latin, of which it contains many other characteristics. In this case geographic proximity seems to be in agreement with linguistic distance and Slavic commonality. The known sequence of events on that territory indicates that Venetic should be considered as a substratum, whereas the later influx of Latin, Celtic, and Germanic formed the superstrata. No Slavic superstratum is recorded on that territory. In spite of that the Venezian sound system is by our results closer to Old Slovenian than to Latin. Does this mean that the sound frequecy is more persistent than other characteristics of a language? Would this explain the closeness of sound frequencies to Estonian and Finnic, which would have its origin in the ‚nostratic' ages?

There is also the question, why the presumably Kentum Venetic [10-12], in contact with Kentum Latin, Kentum Celtic, and Kentum Germanic turned to Romanic Venezian, which contains many Satem-like characteristics? What triggered this direction of development? Which of these components was in fact not Kentum but Satem?

## Acknowledgements

The authors wish to thank to:

- Milan Smolej for his kind help in determining the conversion rules used for the preparation of the EsLD and of FiLD;
- Andrej Perdih for his help in obtaining and preparing the BqLD, GrLD, HiLD, LuLD, Old CsLD, SlLD, OsLD, and UmLD,
- Branislav Perdih for making the necessary computational programs;
- Dr. Marko Perdih for executing the PCA using his program.

## References


1. J Nerbonne, W Heeringa, Measuring Dialect Distance Phonetically, In: J Coleman (ed.): *Workshop on Computational Phonology*, Madrid 1997, 12-15. Available as: http://odur.let.rug.nl/~nerbonne/paper.html
2. J Nerbonne, W Heeringa, Computational Comparison and Classification of Dialects, *2nd International Congress of Dialectologists and Geolinguists*, Amsterdam 2002, 1-16
3. J Nerbonne, W Heeringa, E Van den Hout, P Van der Kooi, S Otten, W Van de Vis, Phonetic Distance between Dutch Dialects. In: Durieux G., Daelemans W. & Gillis (eds.), *CLIN VI, Papers from the sixth meeting*, University of Antwerp, Center for Dutch Language and Speech, Antwerp 1996, 185-202. Available as: http://odur.let.rug.nl/~nerbonne/paper.html
4. B Kessler, Computational Dialectology in Irish Gaelic, *Proc. of the 6th Conference of European ACL*, Dublin 1995, 60-66
5. W Heeringa, C Gooskens, Norwegian Dialects examined Perceptually and Acoustically, *Computers and the Humanities*, Groeningen 2003, *37*, 295-297
6. J B Kruskal, An Overview of Sequence Comparison. In: Sankoff D, Kruskal J (eds.), *Time Warps, String Edits and Macro Molecules. The Theory and Practice of Sequence Comparison*, 2nd Edition, CSLI, Stanford 1999, 1-44
7. W Vieregge, A Rietveld, C Jansen, A Distinctive Feature Based System for the Evaluation of Segmental Transcription in Dutch, *Proceedings of the 10th International Congress of Phonetic Sciences*, Dordrecht 1984, 654-659
8. 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*, Založništvo Jutro, Ljubljana 2005, 77-85 (ISBN 961-6433-51-2)
9. M Silvestri, G Tomezzoli, Linguistic distances between Rhaetian, Venetic, Latin and Slovenian languages, *Proceedings of the Fifth International Topical Conference, Origin of Europeans*, Založništvo Jutro, Ljubljana 2007, 184-190 (ISBN 961-6433-83-9)
10. G B Pellegrini, A L Prosdocimi, *La Lingua Venetica*, Vol. 1, 2, Istituto di Glottologia dell'Univ. di Padova, Circolo Linguistico Fiorentino, Padova-Firenze 1967
11. A Marinetti, Venetico 1976 – 1996. Acquisizioni e Prospettive – Protostoria e Storia del 'Venetorum Angulus', *Atti del XX Convegno di Studi Etruschi ed Italici*, Pisa – Roma MCMXCIX, 391-436
12. M Lejeune, *Les Inscriptions Vénčtes*, Univ. Degli Studi di Trieste, Del Bianco Editore, Udine 1965
13. M Bor, J Šavli, I Tomažič, *Veneti naši davni predniki*, Editiones Veneti, Vienna: German Ed. 1988, Slovenian Ed. 1989, Italian Ed. 1991, English Ed. (*Veneti. First Builders of European Community*) 1996, Russian Ed. Part I 2002.
14. E Vetter, Die Herkunft des venetischen Punktiersystems, *Glotta* 1920, *XXIV*, 114-133
15. V Vodopivec, Atestinske tablice verski in jezikovni pomniki naših prednikov, *Proceedings of the First International Topical Conference, The Veneti within the Ethnogenesis of the Central-European Population*, September 17/18, 2001, Založništvo JUTRO, Ljubljana, 2002, 167-181 (ISBN 961-6433-06-7)
16. A Ambrozic, G Tomezzoli, The "Tavola da Este" Inscription, *Proceedings of the International Workshop, Traces of European Past*, October 10/11, 2003, Založništvo Jutro, Ljubljana 2004, 132-146 (ISBN 961-6433-34-2)
17. V Vodopivec, Študija prečrkovanj in branj najstarejšega venetskega napisa (A Study of Transcriptions and Readings of the Oldest Venetic Inscription), *Zbornik tretje mednarodne konference Staroselci v Evropi* (*Proceedings of the Third International Topical Conference Ancient Settlers of Europe*), Založništvo Jutro, Ljubljana 2005, 121-130
18. F S Smole, Nekaj o venetskih napisih (On Venetic inscriptions), *Zbornik četrte mednarodne*

*konference Evropski staroselci* (*Proceedings of the Fourth International Topical Conference Ancient inhabitants of Europe*), Jutro, Ljubljana 2006, pp. 145-163

19. A Ambrozic, P Serafimov, G Tomezzoli, The Venetic inscription Es 120 on the cup of "Scolo Di Lozzo", *Zbornik četrte mednarodne konference Evropski staroselci* (*Proceedings of the Fourth International Topical Conference Ancient inhabitants of Europe*), Jutro, Ljubljana 2006, pp. 164-171
20. A Kumar, Analiza in razlaga napisov na Vojvodskem stolu (Analysis and interpretation of the inscriptions on the Ducal throne), *Zbornik četrte mednarodne konference Evropski staroselci* (*Proceedings of the Fourth International Topical Conference Ancient inhabitants of Europe*), Jutro, Ljubljana 2006, pp. 181-191
21. V Vodopivec, Nabor venetskih napisov, delitev, prevod in slovar (Collection of Venetic inscriptions, division, translation, vocabulary), *Zbornik četrte mednarodne konference Evropski staroselci* (*Proceedings of the Fourth International Topical Conference Ancient inhabitants of Europe*), Jutro, Ljubljana 2006, pp. 118-144
22. www.thelatinlibrary.com
23. http://www.kortlandt.nl/editions/freis.html
24. http://kodeks.uni-bamberg.de/AltSloven/Quellen/ASL.Ratetsch.htm
25. http://kodeks.uni-bamberg.de/AltSloven/Quellen/ASL.Sittich.htm
26. http://kodeks.uni-bamberg.de/AltSloven/Quellen/ASL.Castelmonte.htm
27. http://members.tripod.com/adolfozavaroni/este.htm
28. http://members.tripod.com/adolfozavaroni/padua.htm
29. http://members.tripod.com/adolfozavaroni/cadore.htm
30. http://www.classicitaliani.it/index100.htm
31. S Schumacher, Die Rätischen Inschriften, Geschichte und heutiger Stand der Forschung, 2., erweitete Auflage, *Archaeolingua*, Innsbrucker Beiträge zur Kulturwissenschaft, Sonderheft 121, Innsbruck 2004 (ISBN 3-85124-214-9)
32. M Bor in: *Etruščani in Veneti* (Tomažič I., Ed.), Editiones Veneti, Wien 1995
33. V Vodopivec, Zbir retijskih napisov, delitev, prevod in besednjak, *Zbornik šeste mednarodne konference Izvor Evropejcev* (*Proceedings of the sixth international topical conference Origin of Europeans*), Jutro, Ljubljana 2008, 118-136
34. G Tomezzoli, About two Magrè-Rhaetic inscriptions in the Civic Natural History Museum in Verona, *Zbornik prve mednarodne konference Veneti v etnogenezi srednjeevropskega prebivalstva* (*Proceedings of the first international topical conference The Veneti within the ethnogenesis of the Central-European population*), Jutro, Ljubljana 2002, pp. 182-187
35. G Tomezzoli, The "Spada di Verona", *Zbornik posveta Praprebivalstvo na tleh Srednje Evrope* (*Proceedings of the Conference Ancient Settlers of Central Europe*), Jutro, Ljubljana 2003, pp. 65-73
36. I Tomažič, G Tomezzoli, The inscription Pauli No. 39, *Proceedings of the International Workshop, Traces of European Past*, Založništvo Jutro, Ljubljana 2004, 147-157 (ISBN 961-6433-34-2)
37. V Vodopivec, Primerjava branj retijskega napisa na meču iz Verone (*Comparison of Readings of a Rhaetian Inscription on the Sword of Verona*), *Proceedings of the International Workshop, Traces of European Past*, October 10/11, 2003, Založništvo JUTRO, Ljubljana 2004, 158-165 (ISBN 961-6433-34-2)
38. V Vodopivec, Prevodi napisa Pauli št. 39 (*Translations of the inscription Pauli No. 39*), *Proceedings of the International Workshop, Traces of European Past*, October 10/11, 2003, Založništvo JUTRO, Ljubljana 2004, 176-177 (ISBN 961-6433-34-2)
39. V Vodopivec, Primerjava branj retijskega napisa na meču iz Verone - popravki (Comparison of Readings of a Rhaetian Inscription on the Sword of Verona - Corrections), *Zbornik tretje mednarodne konference Staroselci v Evropi* (*Proceedings of the Third International Topical Conference Ancient Settlers of Europe*), Založništvo Jutro, Ljubljana 2005, 175-176
40. P Serafimov, Steinberg Inscription, *Zbornik četrte mednarodne konference Evropski staroselci*

(*Proceedings of the Fourth International Topical Conference Ancient inhabitants of Europe*), Jutro, Ljubljana 2006, pp. 172-180

41. http://www.buruxkak.org/pdf/261_SAN%20BENOATEN%20BIZITZEA.pdf;
42. http://en.wikipedia.org/wiki/Basque_language;
43. http://www.kalevipoeg.info/index.html;
44. M Pallottino, *Testimonia Linguae Etruscae*, "La Nuova Italia Editrice", Firenze, 1954;
45. A Berlot, I Rebec, *So bili Etruščani Slovani?* Lipa, Koper 1984
46. V Vodopivec, Primerjava branj pyrgijskih zlatih ploščic (Comparison of readings of Golden Plates of Pyrgi), *Zbornik posveta Praprebivalstvo na tleh Srednje Evrope* (*Proceedings of the Conference Ancient Settlers of Central Europe*), Jutro, Ljubljana 2003, pp. 51-63
47. A Ambrozic, The „Warrior" Stele from Lemnos, *Zbornik tretje mednarodne konference Staroselci v Evropi* (*Proceedings of the Third International Topical Conference Ancient Settlers of Europe*), Založništvo Jutro, Ljubljana 2005, 107-120
48. http://www.sacred-texts.com/neu/kveng/index.htm;
49. J M Crawford, The Kalevala, in two volumes, The Robert Blake Company, Cincinnati, Third Edition, 1910, [Copyright 1888]
50. http://www.perseus.tufts.edu/cgi-bin/ptext?doc=Perseus:text:1999.01.0133:toc
51. http://www.premiumwanadoo.com/cuneiform.languages/index_en.php?page=textes;
52. M Zorman, K Rojko, Ž Antauer, M Cajnko, T Matič, P Pečnik, A Perdih, N Podlipnik, N Zotlar, *Zakaj je bilo treba uničiti Zalpo*, Znanstvenoraziskovalni inštitut Filozofske fakultete, Ljubljana 2005
53. J Friedrich, *Hethitisches Elementarbuch* I, II. Universitätverlag, Heidelberg 1975
54. H C Melchert, *Anatolian historical phonology*, Rodopi, Amsterdam, Atlanta 1994
55. S. Kopriva, Latinska slovnica, Obzorja, Maribor , pp. 12-15].
56. O Szemerenyi, *Introduction to Indo-European Linguistics*, Oxford University Press, Oxford 1996
57. *Cuneiform Luvian Corpus* by H. Craig Melchert (last revised 7/20/01): http://www.unc.edu/~melchert/CLUVIAN.pdf
58. J Chadwick, *Documents in Mycenaean Greek*, Cambridge University Press, Cambridge, 1973, Mycenean Glossary, 527-593
59. http://www.slav.helsinki.fi/ccmh/suprasliensis.html
60. http://titus.fkidg1.uni-frankfurt.de/texte/etcs/slav/aksl/suprasl/supra.htm
61. V Babič, *Učbenik stare cerkene slovanščine*, Filozofska fakulteta Univerze v Ljubljani, Oddelek za slovenistiko, Ljubljana 2003
62. http://titus.uni-frankfurt.de/didact/idg/ital/oskinsc.htm
63. http://www.google.com/gwt/n?u=http%3A%2F%2F
64. www.wordgumbo.com%2Fie%2Fcmp%2Fosca.htm&_gwt_nav=F%7C.1.0.
65. http://www.britannica.com/oscar/print?articleId=109773&fullArticle=true&tocId=74679
66. http://www.forumromanum.org/latin/buck_1.html
67. C Brixhe, M Lejeune, Corpus des Inscriptions Palčo-Phrygiennes, Čditions Recherche sur les Civilisations, Paris 1984 – ISBN-2-86538-089-0, «Memoire» N° 45
68. A Ambrozic, *Gordian Knot Unbound*. Cythera Press, Toronto 2002, pp. 1-57.
69. *Brižinski spomeniki*, znanstveno kritična izdaja, SAZU, Ljubljana, Družina 1993, pp. 36-99
70. N Mikhailov, *Jezikovni spomeniki zgodnje slovenščine, rokopisna doba slovenskega jezika*, prevod H Ošlak, Mladika, Trst 2001, pp. 79-94, 106-108
71. http://titus.uni-frankfurt.de/texte/etcs/ital/oskumb/oskum.htm
72. J W Poultney, *The Bronze Tables of Iguvium*, American Philological Association, Baltimore 1959
73. http://www.classicitaliani.it/index100.htm;
74. *Slovenski pravopis*, J Toporišič et al. (ed.), Založba ZRC, ZRC SAZU, Ljubljana 2003, 135-143

75. Rules-09.doc available on request (gtomezzoli@epo.org)
76. S Wold, K Esbensen, P Geladi, Principal Component Analysis. *Chemometr Intell Lab Syst* 1987, *2*, 37-52
77. D L Massart, B G M Vandeginste, S N Deming, Y Michotte, L Kaufman, *Chemometrics: a textbook*, Elsevier, Amsterdam 1988
78. R G Brereton, *Chemometrics: applications of mathematics and statistics to laboratory systems*, Ellis Horwood, New York 1990
79. R C Graham, *Data analysis for the chemical sciences*,VCH, Weinheim 1993, pp. 329-343
80. A Pääbo, *The Veneti Language*, Apsley 2006
81. A Ambrozic, Commentary on Andres Pääbo's Internet publication of *The Veneti Language,* http://www.veneti.info/index.php?view=article&catid=80%3Areviews&id=152%3Acommentary-on-internet-publication-of-andres-paeaebo&option=com_content&Itemid=233

## Povzetek

*Primerjava sedanjih in nekdanjih jezikov*

Ugotavljati ujemanje slovnične zgradbe in jezikovnega gradiva, ki to nosi, je pri nekaterih starih jezikih zaradi majhnega obsega in poškodb napisov, ki so pisani zvezno, v narečjih in z mnogimi okrajšavami, brez njihovega dobrega razumevanja dvomljivo. Prav pri venetskih, retijskih in frigijskih napisih so zaradi teh razlogov ustreznejše glasovne primerjave.

Enodimenzionalne in večdimenzionalne analize pogostosti glasov v 16 jezikih, večinoma starih, kjer je pri nekaterih od njih vprašljiva še delitev zveznega besedila na besede, potrjujejo prejšnjo ugotovitev, da sta po pogostosti glasov venetščina in retijščina bližje stari slovenščini kot starim italskim jezikom (latinščini, oskijščini, umbrijščini). Po teh lastnostih sta venetščini in retijščini blizu tudi stara frigijščina in etruščina. Zanimiva je po tem kriteriju podobnost estonščine odnosno finščine z večino od teh jezikov. Latinščina, oskijščina in umbrijščina tvorijo poseben skupek, ki je ločen od skupka, ki ga tvorijo etruščina, retijščina in venetščina. Medtem ko je etruščina blizu retijščini, stari slovenščini, venetščini, itd, pa ni blizu hetitščini in luvijščini, iz katerih naj bi po nekaterih domnevah izhajala. Sedanja benečanščina je po pogostosti glasov bližje stari slovenščini kot pa latinščini in ima, čeprav jo štejejo med kentumske jezike, mnogo satemskih prvin, kar daje slutiti, da so glasovne korenine zelo obstojne, in nam lahko nudijo vpogled v izvore jezikov.

Analize pogostosti glasov in njihovih kombinacij v raznih jezikih dajejo rezultate, ki bi lahko neodvisno dopolnjevali tisto vedenje o jezikih, ki izhaja iz ujemanja slovnične zgradbe in jezikovnega gradiva, ki to nosi.