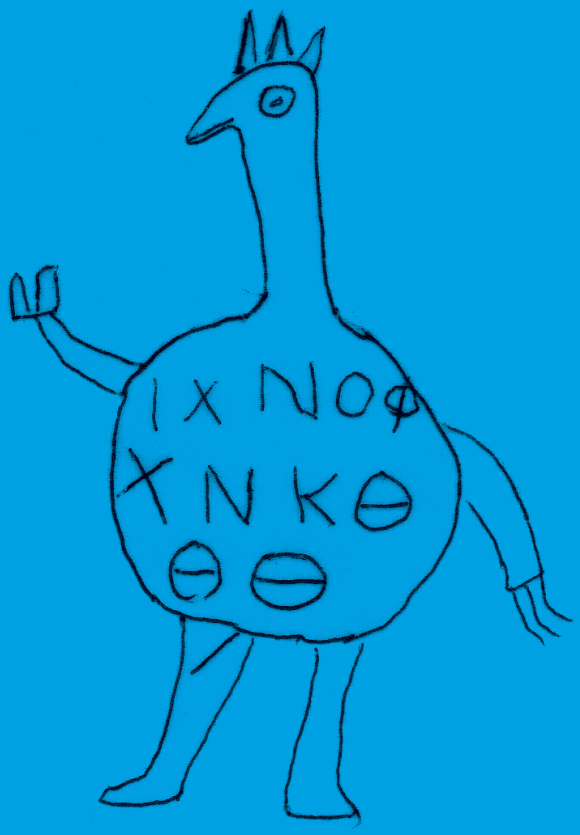
HUNGARIAN POLIS STUDIES Nr. 22

From Polites to Magos

Studia György Németh sexagenario dedicata



Budapest - Debrecen

2016

HUNGARIAN POLIS STUDIES
(HPS)
University of Debrecen
Dept. of Ancient History and Class. Phil.
H–4010 Debrecen, Egyetem tér 1.

Nr. 22
.....
.....

Supporters
Hungarian Society for Antique Studies; Kódex Könyvgyártó Kft.;
Non Omnis Moriar Foundation; University of Debrecen

HU ISSN 1417-1708

2016

FROM POLITES TO MAGOS

Studia György Németh sexagenario dedicata

BUDAPEST – DEBRECEN

HPS 22
From Polites to Magos
Studia György Németh sexagenario dedicata

Redaction
Ádám Szabó

Cover photo:
An image of the daimon Abraxas in bird-form inscribed on one of the six small lead containers found in the cistern of the Fountain of Anna Perenna, Rome in 1999-2000 (IV inscription on the body, IXNOΦ/INKΘ/ΘΘ has been resolved by Gy. Németh (2016) as a slightly inaccurate acronym for an invocation of Christ:

Ἰησοῦς Χριστὸς Ναζωραῖος ὁ [π]αῖς
Ἰησοῦς Ναζωραῖος καὶ Θεός.
Θεὸς, Θεός.

The container is in the Museo Nazionale Romano delle Terme, Dipartimento Epigrafico, inv. no. SAR 475555. Drawing based on Blänsdorf 2012, 624 no. IX.49.6.

ISBN 978-963-284-796-2

Printed by
KÓDEX KÖNYVGYÁRTÓ KFT.
Leader: *Attila Marosi*
Budapest

CONTENTS

Béla Adamik

COMPUTERIZED HISTORICAL LINGUISTIC DATABASE OF THE LATIN INSCRIPTIONS OF THE IMPERIAL AGE: SEARCH AND CHARTING MODULES*

The present paper serves as an introduction and basic presentation of the Computerized Historical Linguistic Database of the Latin Inscriptions of the Imperial Age. Our intention is to let the wider public of professionals interested in the project learn more about how it can assist investigations in the dialectology of the Latin language concerning Roman inscriptions through a special linguistic approach, with the help of modern information technology.

After briefly outlining the aims, the antecedents and the theoretical foundations of the project,[1] some practical elements, first of all the search and charting modules of the Database will be described, since the basic units or building blocks of the Database, the data forms (or data sheets), together with the (pre)history of the project, have already been presented by Adamik 2009 in detail.[2]

1. Aims

As it is displayed on the main page of the Database (http:||lldb.elte.hu|): "The aim of the project 'Computerized Historical Linguistic Database of the Latin Inscriptions of the Imperial Age' is to develop and digitally publish (at http:||lldb.elte.hu|) a comprehensive, computerized historical linguistic database that contains and manages the Vulgar Latin material of the Latin inscriptions found in the so-called Latin part of the Roman Empire

*The present paper has been prepared within the framework of the project OTKA (Hungarian Scientific Research Fund) No. K 108399 entitled "Computerized Historical Linguistic Database of Latin Inscriptions of the Imperial Age" (see: http://lldb.elte.hu/) and of the project entitled "Lendület ('Momentum') Research Group for Computational Latin Dialectology" (Research Institute for Linguistics of the HAS). I wish to express my gratitude to Katalin Horváth and Zsuzsanna Sarkadi for their help in the revision of the English text.

[1] In detail see Adamik 2009 and 2012.

[2] See Adamik 2009, 16–20. Since then, a new module was developed and installed for internal use which serves for checking recorded data forms and improving the data recording process with the help of a complex problem warning and feedback system operating between the data collectors and the principal investigator within the system of the Database. All this aims at achieving the best quality of the recorded data forms, thus yielding solid foundations for the entire work with the Database.

(Illyricum, Gallia, Britannia, Germania, Hispania, Northern Italy, Rome and Africa). This will allow for a more thorough study of the regional changes and differentiation of the Latin language of the Imperial Age in a wider sense and for a multilayer visualization of the discovered structures concerning linguistic geography. The project is going to be realized with the collaboration of the Latin Department of the Eötvös Loránd University, Budapest and the Lendület ('Momentum') Research Group for Computational Latin Dialectology of the Research Institute for Linguistics of the Hungarian Academy of Sciences, supported by the Hungarian Scientific Research Fund (= OTKA, no. K 108399, 2014–2017; no. K 81864, 2010–2013 and no. K 62032, 2006–2009) and by the 'Momentum' Program of the Hungarian Academy of Sciences (2015–2020)."

2. Antecedents

As for the antecedents of the Database, two facts have to be mentioned. First, there is no other database of the Vulgar Latin material of Latin inscriptions from the period of the Roman Empire, neither in a printed nor in a digital format. Second, the Database was preceded by the Late Latin Data Base founded and developed by József Herman at the beginning of the 1990's, when he was the director of the Research Institute for Linguistics of the Hungarian Academy of Sciences. After a promising start, his venture unfortunately failed. The main reason was that at the time conditions of information technology were not satisfactory; most of all the lack of internet proved to be obstructive to this venture. In the end, the Late Latin Data Base collapsed around the end of the nineties, and only the printed guidelines for data collection published in 1990 or 1991[3] remained, which later served as a starting point for designing the new Database. Fortunately, Herman never gave up on his idea of this very important project, and some years later, in 2004, during several personal consultations he convinced me it was worth recreating his former Database, but this time on new foundations, in a new format, and with a new team. After successfully applying for financial support to the Hungarian Scientific Research Fund (OTKA) in 2005, we were able to start the development of the new Database renamed "Computerized Historical Linguistic Database of the Latin Inscrip-tions of the Imperial Age" in 2006. Unfortunately, Professor Herman did not live to see the revival of his Database, as he died prematurely in late 2005.

[3] See Herman 1991.

Concerning its basic concepts and main features, the Database has to be regarded as a direct continuation of Herman's Late Latin Data Base, which is definitely symbolised by the website address http://lldb.elte.hu/ and the abbreviated prefix LLDB standing before the serial number of each data form, such as LLDB-37952 on the next figure:

LLDB-number:	LLDB-37952
Code:	í > E
Alternative code:	litterae perperam incisae
Relevant text:	PRINCEPIO
Classical text:	Principio
Bibliography:	title: RIU?, volume: 2, fascicle: , inscription number: 604, line number: 6
Equivalent Bibliography:	title: CIL?, volume: 4335, fascicle: , inscription number: 604, line number: 6
Localization	Province Pannonia superior, City/Place Brigetio
Date:	AD 251 AD 350
Type of inscription	non-christian, prose, private, object bearing the inscription: stone
Evaluability factors:	
Remark:	Es fehlt die oberste Hasta der E. Datierung nach EDH039588.
Collector:	Barta Andrea

Fig. 1.: A data form of the Database

3. Theoretical foundations

As for the theoretical foundations, the present Database, just like Herman's former Late Latin Data Base, provides the means for employing Herman's methodology. In short, Herman's methodology is based on a distributional analysis of misspellings (i.e. deviations from the classical norm[4]) to be found on inscriptions: after the registration of all the 'faults' in any region concerned, you can calculate the relative frequencies of the different types of 'faults'. This way every region concerned will have its own profile based on the proportion of deviations. The linguistic differences between the regions may be represented as differences between these profiles, that is, in the

[4] Cf. the Guidelines for Data Collection (http://lldb.elte.hu/admin/doc_guidelines.php): "For the purposes of this project, data are defined as linguistic phenomena that can be isolated at text level (in terms of surface manifestation, such phenomena can be phonetic [orthographic], morphological, syntactic, lexicographical, or lexico-semantic) and that deviate from what is called the 'classical' norm."

diverse distribution of the 'faults'.[5] The present Database with all its modules has been developed for executing various kinds of distributional analyses according to Herman's methodology. The efficiency of Herman's method and of its practical application by the present Database has been proved by the significant results achieved and presented over the last years (see http:||lldb.elte.hu| documents.php). It follows from the above that the Database fills a niche in the field of Latin dialectology at an international level.
Now we turn from the theoretical foundations of the project to the practical tools of the Database. First, we will present the search modules, then we continue with the charting module of the Database. A database can be useful only if you can exploit and evaluate the thousands or myriads of pieces of data recorded in it. As for the linguistic and dialectological aims of the project, it is of vital importance to be able to search the inner structures of the data sets and visualize the inner distributions of these. Therefore, with the help of IT specialist Márton Kiss, we have developed two search modules and a charting module which enable us to discover and visualize the distributional structures of diverse data sets.

4. The Simple Search module

The Simple Search module (see http:||lldb. elte.hu|admin|search.php) enables us to perform a search for each field of the data form. Here the combination of search criteria is limited to one single criterion for a each heading or subheading. For example, Code can be either Vocalismus as a single code type or specified in á > E as a single code, while you may add another criterion for another heading or subheading of the data form (e.g. by setting Localization, Province to Pannonia Inferior). By setting one criterion for each heading, you can easily search for specific data (e.g. data dated from the year 101 to the year 300 A.D., written in prose, on stone etc.). If you wish to set multiple search criteria for a single heading or subheading of the data form, you will need to use the Extended Search Module.

5. The Extended Search module

In order to exploit the linguistic data as much as possible, the Extended Search module (see http://lldb.elte.hu/adminsearch_2.php) enables us to

[5] If a 'fault' is committed at a specific point of the linguistic structure, it cannot have any other reason than a linguistic one for not being committed at other points. As a result, the diverse distribution of 'faults' must display linguistic differences. See Adamik 2012 in detail.

search in every field of the data forms with an unlimited combination of the search criteria. This means that with the help of the 'Add search group' system you can set more than one search criterion for each data field. You can connect search groups and combine criteria with the operators AND/OR. Thus, contrary to the Simple Search module, any number of criteria may be set for each field of data forms. For example, you may set Code for Vocalismus AND/OR Consonantismus etc. as code types and á > E AND/OR á: > e etc. as codes, restricting this search to the Localization (Province) as Pannonia Inferior AND/OR Pannonia Superior, at the same time setting dates from the year 101 to the year 300 AND/OR from the year 151 to the year 200, adding various other criteria, such as AND/OR prose AND/OR stone etc. Moreover, not only can additional criteria be included in any given search, but you may also exclude criteria (by choosing 'not equal' in the search window of the relevant field or subfield). If you want to exclude criteria, all criteria to be excluded (set either in the same search group or in separate search groups) should always be connected by AND and never by OR (e.g. Fortasse recte? Equal No), see Figure 2.
There is specific search help available connected to each subheading of the data form (e.g. to various types of search concerning the subheadings Period of the heading Date). If you are a registered user, you can even save your queries by giving them a name in the field 'Saved query name' and clicking 'Save this search query'. By clicking 'Saved queries', you can go to your list of saved queries and insert them again in the Extended Search module (where you will be able to save it again or delete it).

6. The charting module

The charting module connected to the search modules enables us to investigate and visualize inner structural properties of the searched data sets displayed as separate records arranged under the relevant chart. Each record containing more data rows can be split into separate records (by clicking 'UnJoin data rows'). For technical reasons the chart itself will, however, only be displayed if it has less than 50 records, i.e. unjoined data rows; if there are more than 50 data rows (records) to be displayed, the number of records below the chart must be reduced to less than 50 records until the chart above appears, either by switching off unnecessary records or by contracting related codes in the same record and switching off cleared records via clicking 'don't chart this record'.

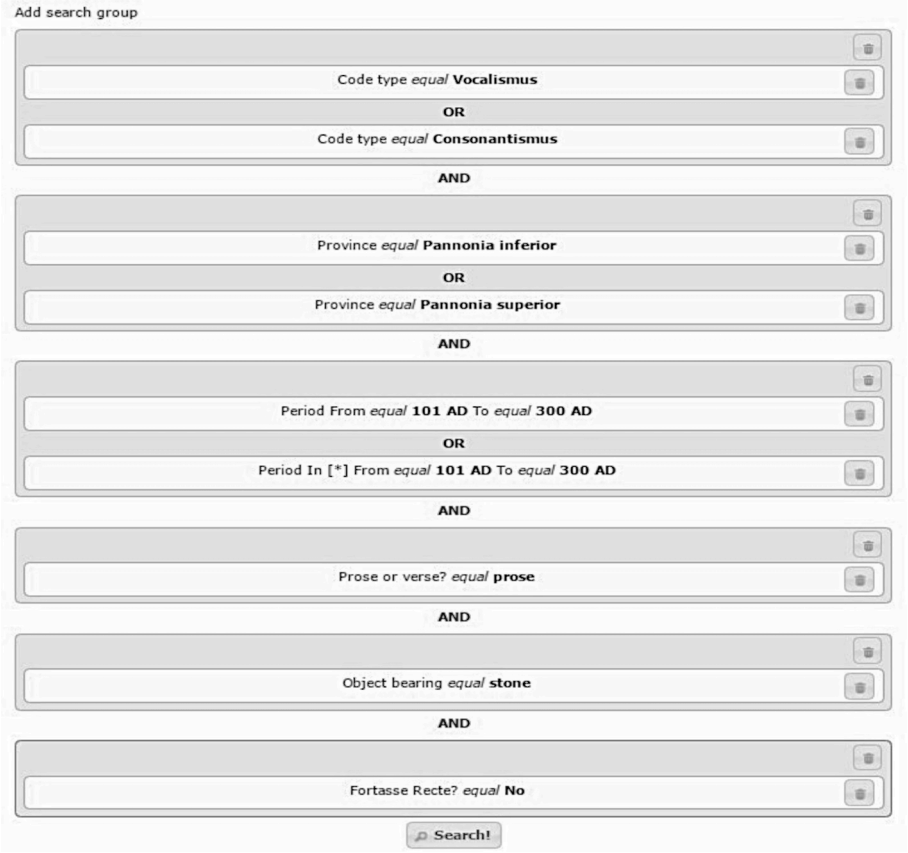
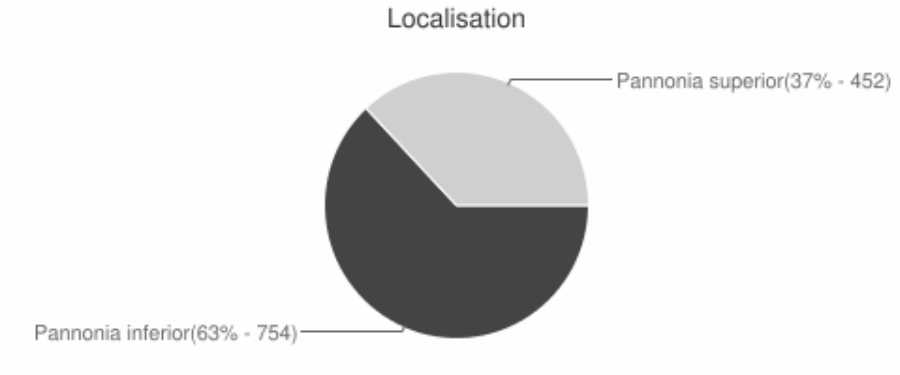


Fig. 2.: Including and excluding criteria in the Extended Search Module

The order of the records can be changed manually by clicking on the arrows (↓ or ↑) at the head of the record and the movement of the records will automatically be followed by the related section of the chart. New records can be created by clicking 'Add record'. The data rows of a record can be moved and placed in another record by clicking on the related data row (i.e. a line in a record) and at the same time dragging it over to another record. The records can be named or renamed freely, and the change will immediately be reflected in the relevant section of the chart. Please see the

following figures with the charts and relevant records of a search concerning the localisation of a data set:[6]



Add record

Don't chart this record!

UnJoin data rows

Pannonia inferior

sum of record: 754

- Ad Statuas: 1
- Alisca: 5
- Alta Ripa: 1
- Altinum: 2
- Aquincum: 534
- Gorsium: 20
- Intercisa: 95
- Lugio: 1
- Lussonium: 7
- Matrica: 10
- Sopianae: 29
- Tricciana: 2
- Ulcisia: 41
- Vetus Salina: 6

[6] The charts referred to in this survey represent the status of the Database on 20/04/2016.

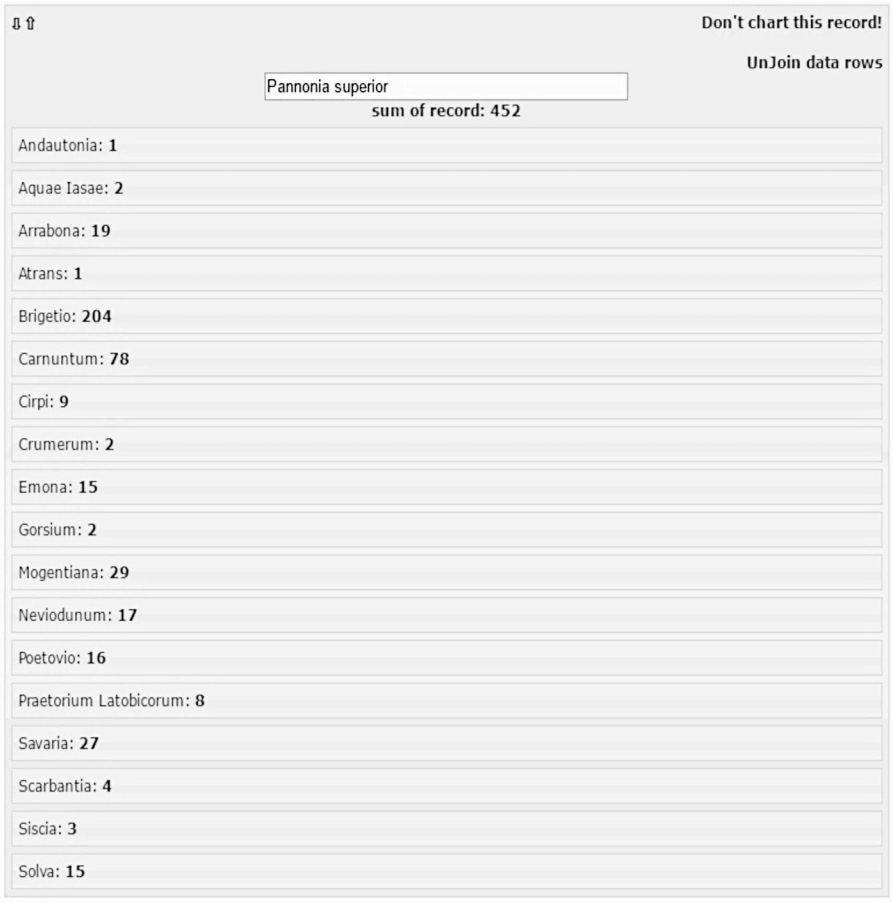


Fig. 3.: Chart on Localisation with relevant records displayed under it

The chart itself can be named (‘title of the diagram’) and its settings can be modified in the given ways (size|color|shape of the diagram). The original layout of the diagram can be reset by reloading the related webpage or by using CTRL + F5. The chart can be downloaded to the user’s computer (right click on the mouse and 'Save Picture as'), while registered users can save it along with the related records into the registered user’s folder in the system. After that, the diagram saved along with the related records can be reloaded from the ‘saved diagrams’ folder by clicking on its title; then the

reloaded diagram (with the same data content) can be edited and saved again under another name etc.
The basic searches which can be displayed on charts are as follows:
'Code 1': for considering the main code of the selected data forms. If more code-types (e.g. Vocalismus OR Consonantismus) are chosen, the distribution will show each code-type (e.g. Vocalismus, Consonantismus), see figure 4; if only one code-type (e.g. Vocalismus) is chosen, the distribution will consider each code (e.g. á > E, á: > E etc.), while respecting the technical peculiarities explained above.[7]

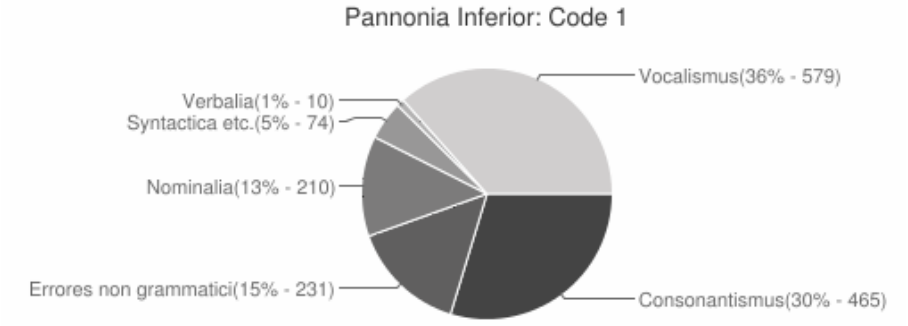


Fig. 4.: Data according to Code 1 (main code types)

'Code 1a': charts the search results just like 'code 1' but by merging similar code-types, i.e. Vocalismus and Consonantismus are integrated as Phonologica, and Nominalia, Verbalia and Syntactica are integrated as Morphosyntactica.

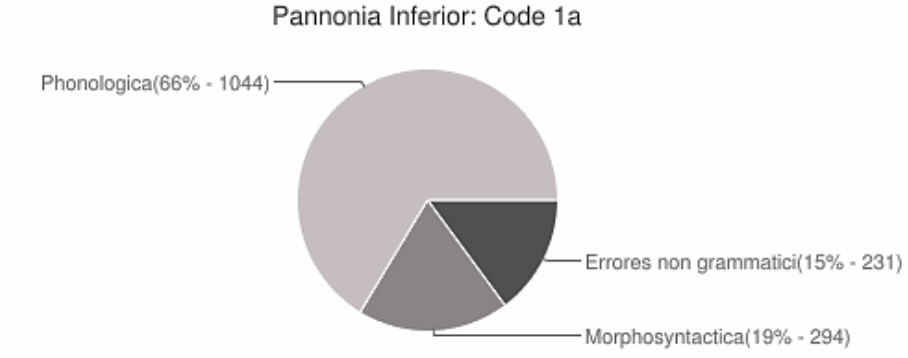


Fig. 5.: Data according to Code 1a (main code types, merged)

[7] The code system see under http://lldb.elte.hu/admin/abbrev_codes.php

‘Code 2’: for considering the main and alternative codes of the selected data forms, beyond the main codes (e.g. o: > V) also the same corresponding alternative codes (e.g. o: > V) are displayed with their figures integrated in the relevant record and chart. The module enables us to work with specific codes regardless if they are recorded as main or alternative codes. It works only for different types of main and alternative codes, e.g. where the main code is Decl. IV pro II, and the alt. code is o: > V or vice versa. For data forms where the main and the alternative code are of the same code-type, e.g. both belong to Nominalia, just like in the case of the main code nom.|acc. pro abl. with the alt. code dat.|abl. pro acc., the data form will be displayed only with its main code considered in order to avoid involving the same data form twice. For the general rules of displaying and editing charts and records, see above.

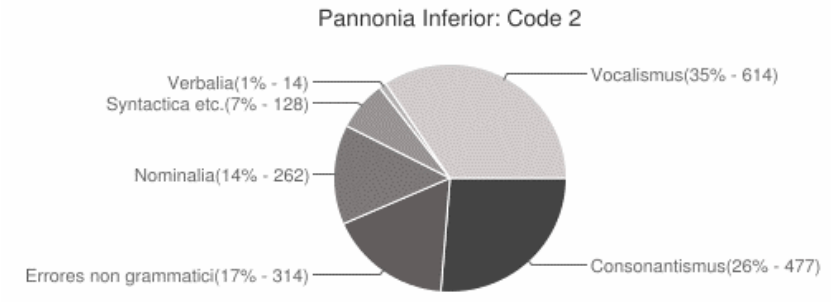


Fig. 6.: Data according to Code 2 (codes, by type)

‘Code 2a’: charts the search results just like ‘code 2’ but by merging code-types, i.e. Vocalismus and Consonantismus are integrated as Phonologica, and Nominalia, Verbalia and Syntactica are integrated as Morphosyntactica.

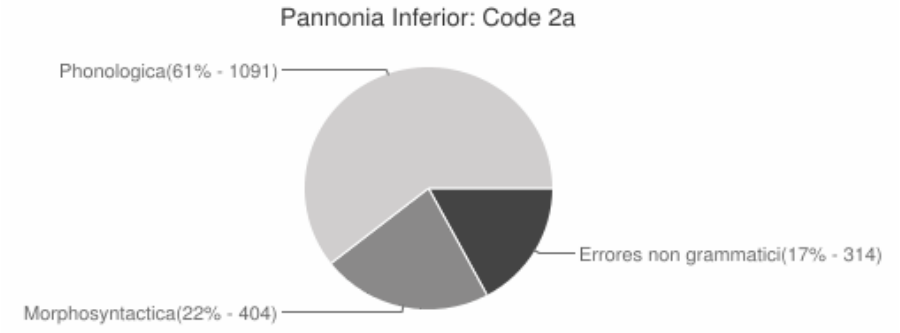


Fig. 7.: Data according to Code 2a (codes, by merged type)

‘Localisation’: for considering places. If more provinces are chosen, the distribution will show single provinces; if one single province is chosen, the distribution will show cities, while respecting the technical peculiarities described above.

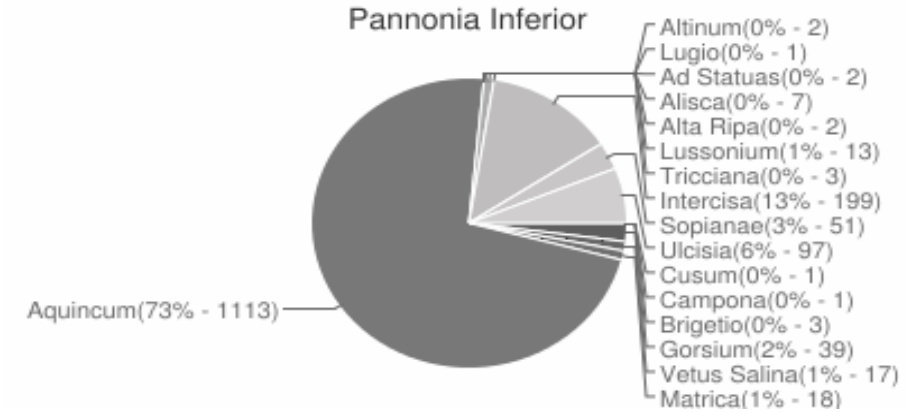


Fig. 8.: Data according to Localisation

‘Date 1’: for considering centuries. The selected data forms are automatically categorized in centuries. Data forms with a date longer than one year are inserted by averaging their date as follows: a data form with a date 151–200 A.D. will be categorized into the 2nd century A.D. based on its average of 176 A.D.; a data form with a date 251–350 will be categorized in the 4th century A.D. based on its average of 301 A.D., and so forth.

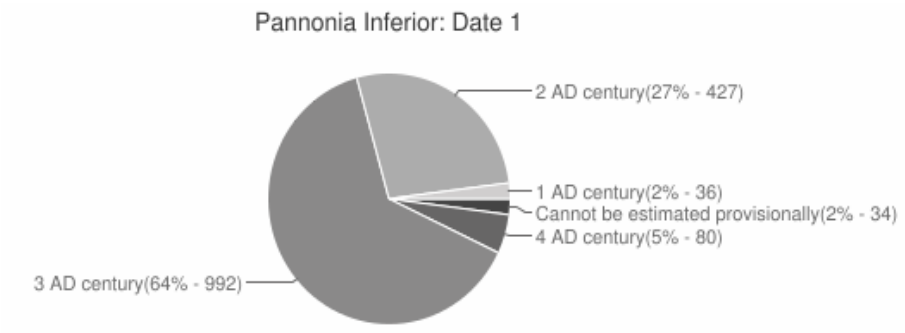


Fig. 9.: Data according to Date 1 (centuries)

'Date 2': for considering centuries. The selected data forms are automatically categorized in centuries (e.g. the 2nd century A.D.: 101–200, 151–175, 155 etc.). Data forms with a date expanding a century (e.g. 101–250, 98–110, 1–300 etc.) are inserted in a separate record entitled 'other', where they can be edited and rearranged (e.g. put in new records) manually. Undated data forms are inserted in a separate record entitled 'cannot be estimated provisionally'.

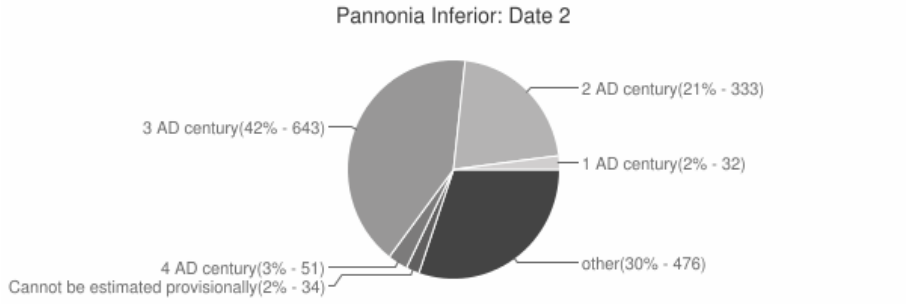


Fig. 10.: Data according Date 2 (centuries and other)

'Bibl 1': for considering the abbreviated title of the corpus displayed in the main bibliography of the data forms (e.g. as for Pannonia Inferior: TitAqu, RIU, RIU-S etc.).[8]

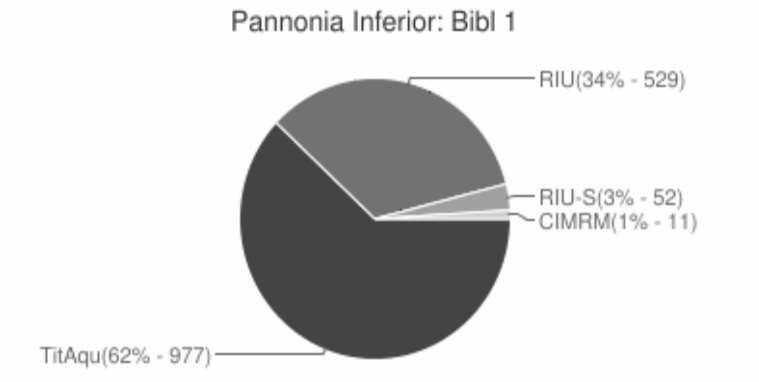


Fig. 11.: Data according Bibl 1(main bibliography)

[8] In the main bibliography, the collection of inscriptions is referred to where the datum was taken from. For the abbreviations of corpora, see http://lldb.elte.hu/ admin/abbrev_bibl.php.

'Bibl 2': for considering the abbreviated title of the corpus displayed in the equivalent bibliography of the data forms (e.g. as for Pannonia Inferior: CIL, AE etc. etc.).[9]

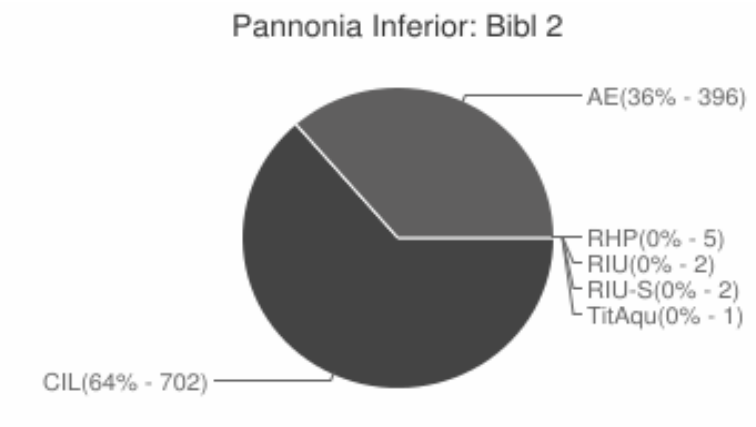


Fig. 12.: Data according Bibl 2 (equivalent bibliography)

The modules of the Database presented in the above chapters, i.e. the search modules and the charting module are definitely suitable for revealing and visualizing the inner structures of data sets concerning various criteria. Accordingly, our Database can well become a very successful tool for dialectological research based on Latin inscriptions, as has been proved on several occasions at an international level.

7. The future of this Project

After a successful application in 2015 for the Momentum (Lendület) program of the Hungarian Academy of Sciences, the Momentum (Lendület) Research Group for Computational Latin Dialectology was founded on the 1st July 2015 in the Research Institute for Linguistics of the Hungarian Academy of Sciences for the time span 2015–2020.[10] The main task of this Research Group for these five years is to intensify and possibly complete the research carried out in the project "Computerized Historical Linguistic

[9] The fields of the subheading for equivalent bibliography are filled in if the inscription in question was previously published in another corpus (e.g. CIL). Referring to equivalent bibliography is necessary mainly in order to avoid recording the same datum twice.
[10] See http://www.nytud.hu/depts/fu/indexlendulet.html.

Database of Latin Inscriptions of the Imperial Age" (formerly supported by OTKA, the Hungarian Scientific Research Fund). In the framework of the present project, the territorial scope of the database has been expanded to those areas of the Roman Empire which are richest in Latin inscriptions (Rome, South-West Italy and Africa), from where at least 30,000 pieces of data will be collected and recorded by a team of more than 20 data collectors during these 5 years. The first 6,000 data forms (mainly from the city of Rome) have already been recorded in the Database (see next chart), and this recording work will hopefully continue with the same intensity in the coming years of the project.

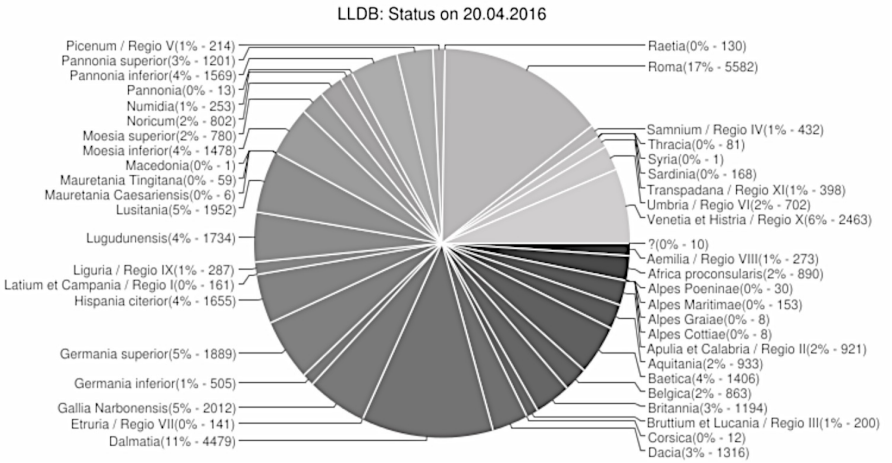


Fig. 13.: Data collected by provinces

It is also intended that all the recorded data be linguistically and dialectologically evaluated, and the results of this evaluation will be presented and published each year by the fellows of the research group, both at a national and an international level. With the indispensable contribution of the members of the wider and wider international team of data collectors,[11] the research group consisting of the fellows Béla Adamik, Andrea Barta and Attila Gonda will hopefully be able to fulfil all these plans.

[11] I am indebted to all of them (cf. http://lldb.elte.hu/team.php) for their help and active contribution.

To sum up, the significance of the research project lies with the fact that the Database will be the only international tool which, thanks to the theoretically well founded background, can resolve the practical problems of Latin dialectology. This will allow for a better understanding of the processes which led to the development of Romance languages and determined the linguistic, ethnic, and even cultural features of medieval and modern Europe.

Bibliography

Adamik 2009 = Adamik, B.: In Memoriam József Herman: von der Late Latin Data Base bis zur Computerized Historical Linguistic Database of Latin Inscriptions of the Imperial Age. *AAntHung* 49, 11–22 (cf. http://lldb.elte.hu/documents.php).

Adamik 2012 = Adamik, B.: In Search of the Regional Diversification of Latin: Some Methodological Considerations in Employing the Inscriptional Evidence. In: Biville, Fr. – Lhommé, M.-K. – Vallat, D. (edd.): Latin vulgaire – latin tardif IX.: Actes du IXe colloque *international sur le latin vulgaire et tardif, Lyon, 6 – 9 septembre 2009*. Lyon 123–139 (cf. http://lldb.elte.hu/documents.php).

Herman 1991 = Herman, J.: *Late Latin Data Base: Guidelines for Data Collection*. Research Institute for Linguistics of the Hungarian Academy of Sciences; manuscript Budapest. (cf. its revised and extended version at http://lldb.elte.hu/admin/ doc_guidelines.php)