



Béla Adamik\*

# Romanization and Latinization of the Roman Empire in the light of data in the Computerized Historical Linguistic Database of Latin Inscriptions of the Imperial Age

<https://doi.org/10.1515/joll-2021-2016>

Published online June 7, 2021

**Abstract:** The present study demonstrates that the process of linguistic Romanization, i.e. Latinization of the Roman Empire, is traceable by the data of the Computerized Historical Linguistic Database of Latin Inscriptions of the Imperial Age (LLDB). A multi-level analysis of linguistic and non-linguistic data in the LLDB has shown that Latinization, i.e. the spread of spoken or vulgar Latin, became more and more intensive over time in all concerned provinces (i.e. Lusitania, Gallia Narbonensis, Venetia et Histria, Dalmatia, Moesia, Pannonia, and Britannia), although to a varying degree in each. What is more, in many aspects of the investigation, it was possible to find differences between the selected provinces of the Roman Empire corresponding mostly to the future Romance (both negative and positive) outcomes of the respective areas. All in all, the analysis of data of the LLDB database can contribute to solving the complex problem of Latinization, and is a lot more appropriate for this purpose than a simple comparative analysis of epigraphic corpora of the selected provinces.

**Keywords:** database; inscriptions; Latinization; Romanization; vulgar Latin

## 1 Introduction

For contributing to the discussion on the challenging and still somehow controversial problem of the linguistic Romanization, i.e. Latinization<sup>1</sup> of the Roman

---

<sup>1</sup> For the concept of Romanization, see Woolf (2006 [2001]) and Spickermann (2006 [2001]); for that of Latinization, see Binder (2006 [1999]). For the relationship between Romanization and Latinization, see Raupach (1996) and Adamik (2006).

---

**\*Corresponding author: Béla Adamik,** Lendület ('Momentum') Research Group for Computational Latin Dialectology, ELKH Research Centre for Linguistics, Benczúr utca 33, Budapest, 1068, Hungary; and Department of Latin, Eötvös Loránd University, Múzeum körút 4/F, Budapest, 1088, Hungary, E-mail: adamik.bela@nytud.hu

Empire, I am going to involve the data of the Computerized Historical Linguistic Database of Latin Inscriptions of the Imperial Age (“the Database” or “LLDB”<sup>2</sup> henceforth). The aim of this project 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 different territories of the Roman Empire (Illyricum, Gallia, Britannia, Germania, Hispania, Italia, Africa, Roma and the eastern provinces). This will allow for a more thorough study of the diatopic changes and variation of the Latin language of the Imperial Age in a wider sense, and for a multilayer visualization of the structures discovered concerning linguistic geography.<sup>3</sup>

In this paper, nevertheless, I will not be dealing with actual dialectological problems – however vital and interesting they are – but with more general problems related to the linguistic Romanization and Latinization of the Roman Empire. By that I mean the spread of spoken or vulgar Latin (preceded and triggered by mostly spontaneous language shift<sup>4</sup> through the agency of the vernacular populations) both territorially and chronologically.<sup>5</sup> These aspects will be addressed in light of data in the Database. The presentation of the problem and the research questions of my paper are as follows: Is the process of Latinization of the Empire traceable by the data of such a linguistic database? If yes, can any differences be

---

**2** Concerning its basic concepts and main features, the Computerized Historical Linguistic Database of Latin Inscriptions of the Imperial Age has to be regarded as a direct continuation of the failed Late Latin Data Base (LLDB) founded by József Herman, which is reflected in the website address <http://lldb.elte.hu/> and the abbreviated prefix LLDB standing before the serial number of each data form (such as LLDB-7188, etc., cited in Section 3 below) in the Database, see Adamik (2016: 14–15).

**3** The project is 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 previously governed by the Hungarian Academy of Sciences (MTA), now under the umbrella of the Eötvös Loránd Research Network (ELKH). For a general description of the Database, see Adamik (2009, 2016); for the methodology behind the Database, see Adamik (2012).

**4** In accordance with the concept of Self-Romanization as formulated by Spickermann (2006 [2001]) which “describes a dynamic process also implying that socially significant groups in Roman provinces have the desire to adopt the Latin language and the culture, lifestyles and religious practices of the Romans. [...] The term therefore allows a distinction [...] between an active policy pursued by the Romans and an uncontrolled development process leading to the formation of specific provincial identities” (Spickermann 2006 [2001]). Cf. also Herman (2000): “The change to using Latin was thus the result of an apparently spontaneous process, of the pressure of many straightforward practical needs, and also, in many cases, of the cultural prestige that Latin had” (Herman 2000: 10).

**5** Cf. Herman (2000): “the linguistic Romanization of these provinces meant the gradual adoption of the Latin language by the native population” (Herman 2000: 10).

evidenced between the provinces of the Roman Empire? If there are differences to be revealed between the provinces, is there any correlation between these differences and the future Romance continuation or discontinuation in the respective provinces, that is, whether Romance languages developed there or not? Can such a linguistic database contribute to this complex problem?

## 2 Methodology

To answer these difficult questions we shall use the Database, which has some unique features. First of all, its entries consist of deviations of all sorts from “Classical Latin” or what is called the “classical norm,”<sup>6</sup> irrespective of whether they are due to linguistic or extra-linguistic (i.e. graphic or technical, etc.) factors. LLDB data collectors record not only data of linguistic but also of technical origin; moreover, we also differentiate between actual linguistic data and purely orthographic phenomena. We have accumulated a huge number of data forms in the Database (more than 100,000),<sup>7</sup> which means we can work with exact, high figures and accurate rates for every relevant aspect related to the linguistic Romanization and Latinization of the Roman Empire. This means we can analyze the spread of spoken or vulgar Latin – as opposed to previous work on the subject, which instead described the general conditions and course of this process, without a detailed analysis of the specific Latin linguistic data.<sup>8</sup>

As for the territorial delimitations of my current analysis, the following restrictions must be taken into consideration. First, although the Database contains data from the entire territory of the Empire, now only the western or the so-called Latin part of the Empire is considered, while the eastern or so-called Greek part is excluded.<sup>9</sup> The reason for this is that the Greek part had quite different

---

6 The term “Classical Latin” can be used in a double meaning; firstly, for the concept of the standard variety of Latin and, secondly, for a phase in the history of Latin (ca. 120 BC – ca. 250 AD); in the same way the term “Vulgar Latin” can be used in a double sense, too; firstly, for the concept of the substandard variety of Latin as an opposite concept to “Classical Latin” and then for the language period subsequent to Classical Latin (ca. 250 – ca. 600); for details, see Adamik (2015: 648).

7 The data forms referred to in this survey represent the status of the Database on 1 October 2020.

8 This applies, for example, just as much to the section “Faktoren und Arten der Romanisierung und Latinisierung” of Raupach (1996: 11–14) as to the chapter “Die Latinisierung Hispaniens” of Garcia y Bellido (1972).

9 The Roman Empire was divided in two parts in terms of official language use, one Latin and one Greek. The two parts were divided by an official linguistic frontier, which on the Balkan peninsula corresponded to the provincial borders between the “Latin” provinces of Dalmatia and Moesia in the north and the “Greek” provinces of Macedonia and Thracia in the south; in Africa, the provincial border between the “Latin” province Africa Proconsularis in the west and the “Greek” province Cyrenaica in the east. Cf. Adamik (2006: 27).

sociolinguistic peculiarities that were unfavorable for a linguistic Romanization, i.e. Latinization.<sup>10</sup> Secondly, from the vast territory of the western or Latin part of the empire, I selected only a few provinces, which are all well processed in the database project, that is, their epigraphic material is almost fully entered in the Database, so we have a considerable amount of data available from there. Regarding the Romance continuation or discontinuation, for areas characterized by a persistent Romance continuation I selected (i) Lusitania, (ii) Gallia Narbonensis and (iii) Venetia et Histria; then, for areas characterized by an extinct Romance continuation I selected (iv) Dalmatia;<sup>11</sup> and, finally, for areas of no Romance continuation, I selected (v) Moesia, (vi) Pannonia, and (vii) Britannia, where no Romance language evolved due to the known historical and sociolinguistic circumstances.<sup>12</sup> In order to see the changes over time, I divided the relevant material into two periods: an early one from the 1st through the 3rd century, and a later one from the 4th through the 7th century (see Figures 1a and 1b).<sup>13</sup>

For the next step, we divided the data of both periods of each province into the following four categories: *Phonologica* and *Morphosyntactica*, representing the “linguistic” data (abbreviated as L in the chart titles of Figures 1a and 1b), and *Orthographica* and *Errores non grammatici*, representing the “non-linguistic” data (abbreviated as NL in the chart titles of Figures 1a and 1b). For each category, two (or in one case, three) illustrative examples are given directly below.

---

**10** Cf. Adamik (2006: 24–29) and see Herman (2000): “Under these circumstances we can understand why the Latinization of the provinces was a slow process, taking several centuries in almost every area; and also why it did not happen at all in the eastern provinces such as Egypt and Asia Minor. Here Greek was firmly in place before the Roman conquest, which meant that these areas already had available a language fully fitted for all practical communication needs and whose cultural prestige was far higher than that of Latin” (Herman 2000: 11).

**11** As for Dalmatia, see Herman (2000): “Dalmatian Romance was spoken throughout the Middle Ages in the coastal cities, and on the island of Veglia it lasted until the nineteenth century; the last native speaker died in 1898” (Herman 2000: 12–13).

**12** As for Pannonia and Moesia, see Adamik (2003). Moesia here means Moesia Inferior and Moesia Superior, and Pannonia means Pannonia Inferior and Pannonia Superior.

**13** Accordingly, we have excluded data forms with a date unclassifiable in the current periodization with a break at 300 AD, e.g. those dated with a time span of 201–400, i.e. to the 3rd–4th centuries. However, data forms without a datation are included in the early period, since as non-Christian inscriptions they actually belong predominantly to the early period (while undated Christian inscriptions are automatically dated to the later period). Moreover, I excluded items imported from other provinces (labelled “Imported from” in the Database), such as military diplomas or imported pottery (e.g. vasa Arretina). I also excluded pottery as such (labelled “pottery” in the Database), since most of these are imported anyway, and it is very difficult to link each item to a single province. Finally, I also excluded data forms which might be regarded as correct and were therefore labelled “*fortasse recte*” in the Database.

### 3 Analysis of linguistic and non-linguistic data

The linguistic data are illustrated by the following examples:<sup>14</sup>

a. *Phonologica*:

1. LLDB-7188: (voc.)-v-(voc.) > B, MILIT|ABIT = *militavit*, RIU 2, 559, 1–2 = CIL 3, 11026, 1–2, Pannonia superior, Brigetio, 301–350 AD.
2. LLDB-11849: é: > I, POSVIRVNT = *posuerunt*, InscrAqu 3, 2,910, 6 = CIL 5, 1,680, 6, Venetia et Histria / Regio X, Aquileia, 365 AD.

b. *Morphosyntactica*:

1. LLDB-25589: acc. pro abl., CVM FILIOS = *cum filiis*, ILGN 102, 3 = CIL 12, 5,419, 3, Gallia Narbonensis, ?, 313–700 AD.
2. LLDB-17177: commut. in form. perfecti activi verbi ponendi, POSIT = *posuit*, Conrad 268, 7 = AE 1972, 523, 7, Moesia inferior, Tropaeum Traiani, 151–200 AD.

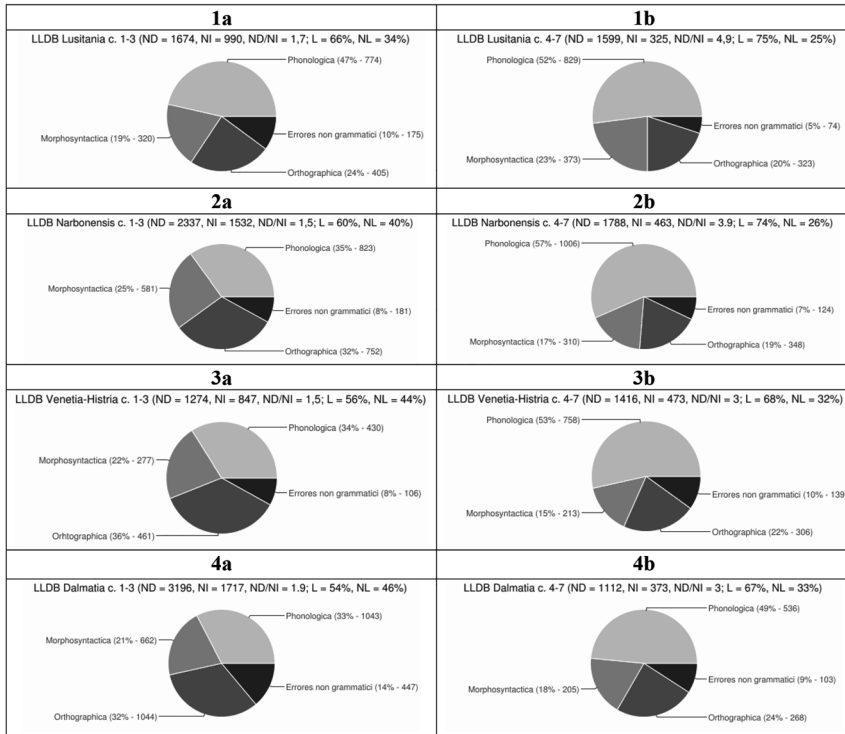
The non-linguistic data are illustrated by the following examples:

c. *Orthographica*:<sup>15</sup>

1. LLDB-30272: H > ø, IC = *hic*, CIIC 1, 327, 1 = IBC 34, 1, Britannia, by Gobannium, 501–600 AD.
2. LLDB-30361: x > SX / CS / XS / XSS / XX, VICSIT = *vixit*, CIL 3, 2,151, 6, Dalmatia, Salona, 151–230 AD.
3. LLDB-60209: ae > E, SVE| = *suae*, AE 2013, 795, 7, Lusitania, by Norba, 190–310 AD.

<sup>14</sup> For resolving abbreviations of inscriptional corpora used in the following examples taken from the LLDB Database, see [http://lldb.elte.hu/admin/abbrev\\_bibl.php](http://lldb.elte.hu/admin/abbrev_bibl.php).

<sup>15</sup> The following codes were included in the purely orthographic phenomena (these codes can be easily separated by the corresponding element of the chart module, i.e. “Code 1 O” or “Code 1a O” in the Extended Search module of LLDB): litterae Graecae, g > C, qu > CV, H > ø, aspiratio vitiosa, ch > C, ph > P, th > T, PH ~ F, c > K, k > C, x > SX / CS / XS / XSS / XX, i (= /j/) > II, áe > E, é > AE, é: > AE, ae > E, e > AE, e: > AE, ae / áe > AI, ae / áe > AEI, ai > AEI / AE, i: > II, e: > EE, a: > AA, o: > OO, u: > VV (the colon or “:” after a vowel indicates that the vowel is long; the accent above a vowel as in e.g. “é” indicates that the vowel is stressed). Purely orthographic phenomena here include not only spelling patterns based on different (substandard) practices (which were to be avoided in standard orthography), such as CV instead of QV, CS instead of X, or geminating vowels as VV to denote long *u*, but also which testify to linguistic changes that have already taken place (are no longer active), such as not writing H (H > ø), writing it in the wrong place (*aspiratio vitiosa*), or confusing AE and E. Cf. Herman (2000): “... in the Republican period, that is, even before the Empire; the laryngeal aspirate /h/ was dropped, in all positions in a word” (Herman 2000: 38). The linguistic change (i.e. the monophthongization of *ae* to open *e*) underlying the fluctuation between AE and E had already taken place at the beginning of the Empire, “probably in the first century AD,” see Herman (2000: 31).

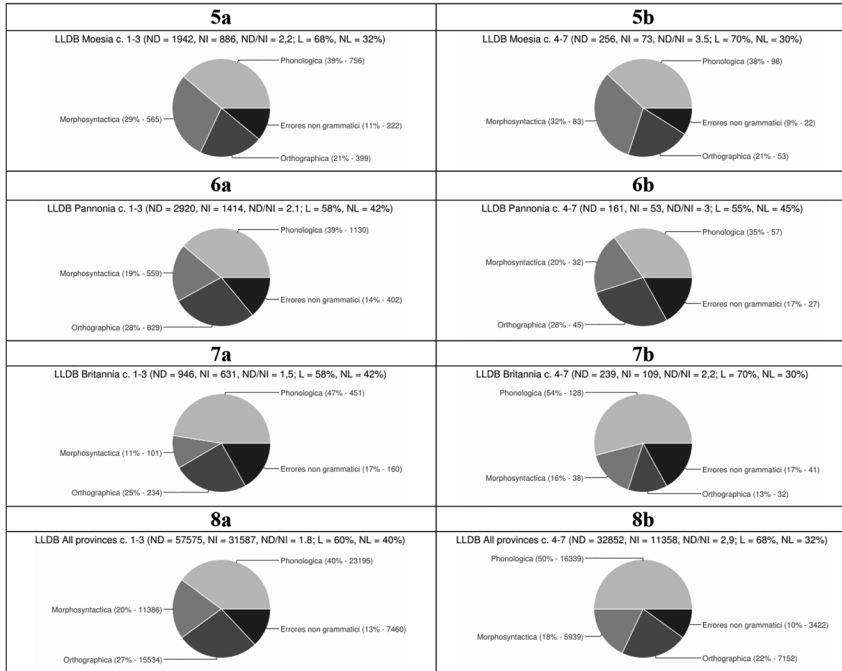


**Figure 1a:** Charts 1a–4b.

d *Errores non grammatici:*

1. LLDB-20923: litterae perperam incisae, EILIAE = *filiae*, InscrIt 10, 1, 559, 1, Venetia et Histria / Regio X, Pola, 313–600 AD.
2. LLDB-31256: syllaba perperam adiuncta (dittographia), AGVSTA|STAS = *Augustas*, Badajoz 41, 9–10 = HEp 7, 154, 9–10, Lusitania, by Emerita Augusta, AD 574 AD.

In Figures 1a and 1b, I present data from each selected province in this fourfold distribution. In the first column you can see data from the early period (i.e. from the 1st through the 3rd century), and in the second column you can see data from the later period (i.e. from the 4th through the 7th century). In the heading of each chart of Figures 1a and 1b, after the abbreviated name of the Database (“LLDB”), you can see the name of the province, the time span of the charted data (e.g. Chart 1a in Figure 1a has “Lusitania c. 1–3 AD”), and between brackets the total number of data forms charted (ND, e.g. 1,674), the number of inscriptions with misspellings



**Figure 1b:** Charts 5a–8b.

(NI, e.g. 990) since one inscription can often yield multiple data forms, and, accordingly, the average number of misspellings or deviations per inscription which we are going to refer to as Data density ( $ND/NI = 1.7$ ). These figures are followed by the rate of linguistic data (L, e.g. 66 %) and the rate of non-linguistic data (NL 34 %). In the last charts of Figure 1b, 8a–8b, you can see displayed these phenomena for all provinces considered, with the average numerical values to be taken as a point of reference. This way the charts in Figures 1a and 1b, all generated by the charting module of the LLDB Database, represent the basic data sets for our interpretation to be explained right now.<sup>16</sup>

In Figures 2a–2f you can see charts interpreting and summarizing the data displayed in Figures 1a and 1b. They will help us compare and rank the selected provinces with respect to the chosen categories, discover the tendencies of changes between the early and the later period, and draw our conclusions with

<sup>16</sup> To learn more about the search and charting modules of the LLDB Database (cf. [http://lldb.elte.hu/admin/search\\_2.php](http://lldb.elte.hu/admin/search_2.php)), see Adamik (2016).

regard to Latinization, so we can see if there is any correlation between the data and the fate of the local Romance languages.

### 3.1 Analysis according to the number of inscriptions with misspellings

Let us start with Charts 1a–1c of Figure 2a, which display the number of registered inscriptions with misspellings, that is, containing any kind of deviations.

From the distribution in the early period, displayed in Figure 2a, Chart 1a, we cannot infer the future Romance developments at all, since e.g. Pannonia with no Romance continuation takes the third place in the leading trio together with Gallia Narbonensis with Romance continuation and with Dalmatia with an extinct Romance continuation, while Lusitania and Venetia et Histria, both with a Romance continuation, are positioned in the group lagging behind, together with Moesia and Britannia, both with no Romance continuation. As opposed to the early period, the distribution in the later period, displayed in Figure 2a, Chart 1b, projects the future Romance developments in an obviously much more faithful way, with a clear split into two prominent groups, one with quite a huge number of registered inscriptions (ca. 300–500) including Gallia Narbonensis, Venetia et Histria, Dalmatia, and Lusitania, all of them with a Romance continuation; and one underdeveloped group with a low number of registered inscriptions (ca. 50–100)

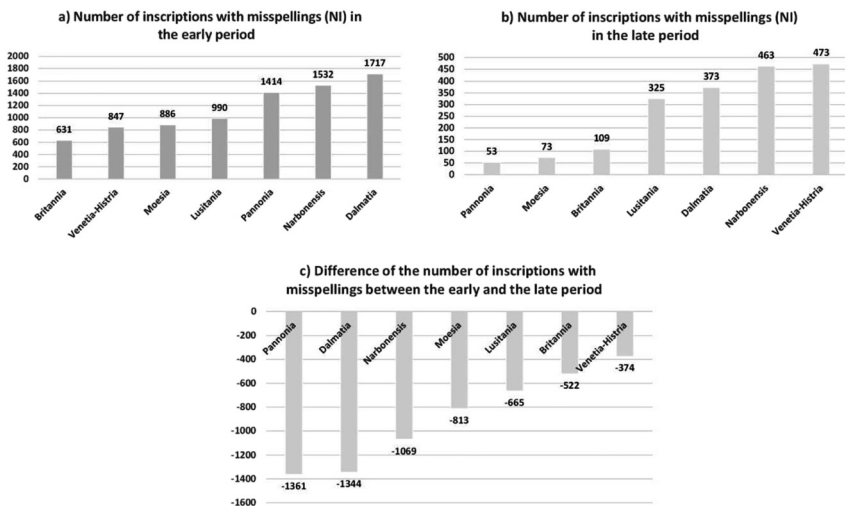


Figure 2a: Charts 1a–1c.



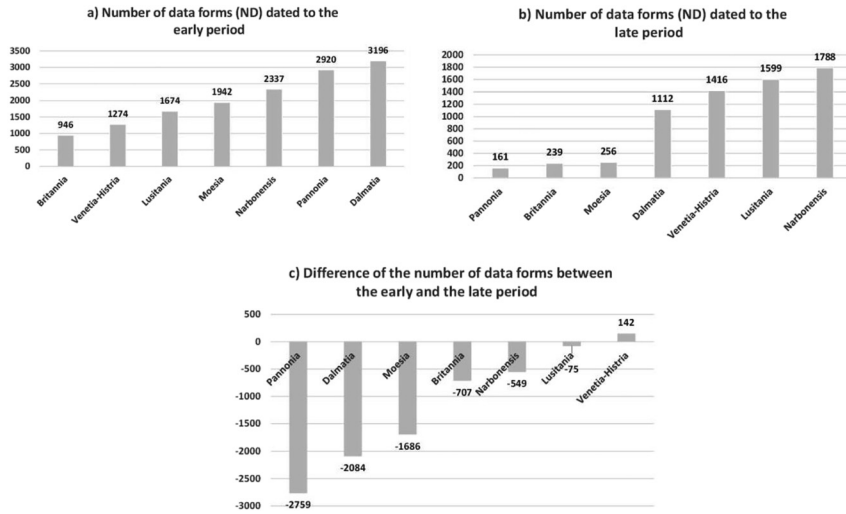


Figure 2b: Charts 2a–2c.

including Britannia, Moesia, and Pannonia, all with no Romance continuation. Only the position of Dalmatia does not show any correlation with the later fate of the local Romance language: with an extinct Romance continuation, we expected it somewhere halfway between the two groups with a number of ca. 200 entered inscriptions. Figure 2a, Chart 1c, displaying the difference in the number of inscriptions with misspellings between the early and the later period, does not help much regarding this issue, since even though Dalmatia holds the second worst position (with a decrease of 1,344 inscriptions) preceding Pannonia in the worst position (with a decrease of 1,361 inscriptions), Gallia Narbonensis got in a similar – that is, the third worst – position (with a decrease of 1,069 inscriptions), while Britannia (with a decrease of 522 inscriptions) got in the leading trio together with Lusitania and Venetia et Histria (with a decrease of 665 and 374 inscriptions, respectively).

### 3.2 Analysis according to the number of data forms

The data are better understood if we regard the distribution of the number of data forms instead of the number of inscriptions with misspellings in the selected provinces, as displayed in Figure 2b, Charts 2a–2c.

While in the early period (cf. Figure 2b, Chart 2a), unsurprisingly, there is no correlation to be detected between the figures for each province and any future

Romance development, in the later period a clear correlation is visible (see Figure 2b, Chart 2b). In the underdeveloped group we see Moesia, Britannia, and Pannonia, all with no Romance continuation (with figures between 100 and 300), and in the well-developed group we see Venetia et Histria, Lusitania, and Gallia Narbonensis, all with a persistent Romance continuation (with figures between 1,400 and 1,800), while Dalmatia with an extinct Romance continuation takes a sort of middle position, a bit lagging behind the developed group (with 1,112 data forms). In this case, Figure 2b, Chart 2c also reflects the future Romance developments in the difference of number of data forms between the early and the later period: the three provinces with a persistent Romance continuation are in the first three positions (Gallia Narbonensis with a decrease of only 549 items, Lusitania with a decrease of 75 items, and Venetia et Histria with an increase of 142 items). The two provinces with no Romance continuation are in the last and third from last positions (Pannonia with a decrease of 2,759 items and Moesia with a decrease of 1,686 items). At the same time, quite “disturbingly,” Dalmatia has a position in the middle of the latter group, with a huge decrease in data forms (2,084 items), while, by contrast (with a decrease of 707 items), Britannia gets very close to the developed group. However, Britannia’s distinguished position regarding both the decrease in the number of inscriptions with misspellings (Figure 2a, Chart 1c) and in the number of data forms (Figure 2b, Chart 2c) is rather illusory and misleading, as it can instead be explained by a superficial Latinization of the province in the early Imperial Age, resulting in low figures (actually the lowest figures in comparison to Charts 1a and 1b) both in the number of inscriptions with misspellings and of data forms.<sup>17</sup> In this case, the distributional patterns displayed in Figure 2b, Chart 2b match the distribution in Romance the best.

### 3.3 Analysis according to data density

Now we turn to Figure 2c, Charts 3a–3c, which display data density in the two periods and the changes in data density between the early and the later period.

As I have mentioned earlier, data density here refers to the average number of misspellings or deviations per inscription. For example, in early Dalmatia we have

---

<sup>17</sup> This is generally also true for the British Latin inscription material, cf. Adams (2013): “The British corpus for its part is limited in extent, and partly composed by outsiders to Britain and cannot be used to establish much in the way of ‘British Latin’” (Adams 2013: 190) and “the inscriptions of Britain [...], which are generally well spelt and also less numerous than those from some other parts of the Empire” (Adams 2013: 151), cf. Adams (2007: 577–623) on Latin in Britain as well; cf. also Smith (1983: 935–938), esp.: “we have no real V. L. texts for Britain: all that we have is vulgar intrusions into would-be correct C. L.” (Smith 1983: 936).

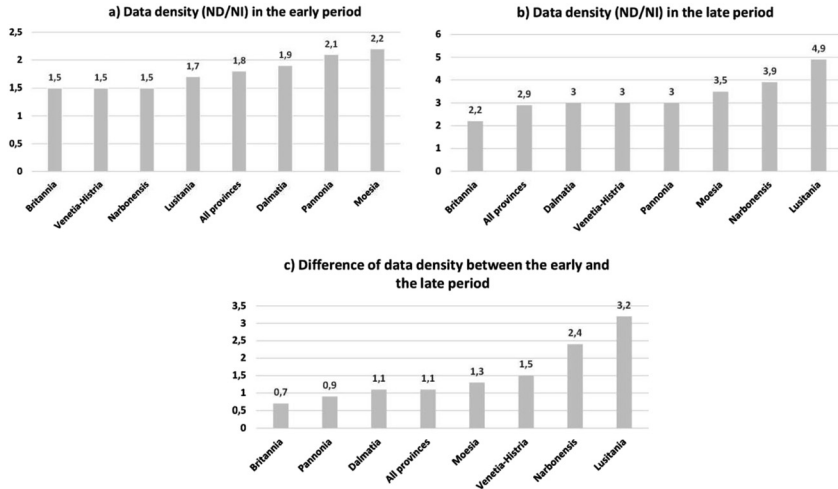


Figure 2c: Charts 3a–3c.

3,196 data forms (ND, cf. Chart 4 of Figure 1a and Chart 2a in Figure 2b), recorded from 1,717 inscriptions (NI, cf. Chart 4 in Figure 1a and Chart 1a in Figure 2a). If we divide the 3,196 data forms by the 1,717 inscriptions, we get a data density of 1.9 (rounded off to the first decimal place, see Chart 3a in Figure 2c). While in the early period (cf. Chart 3a in Figure 2c) again there is no correlation to be detected between the data of the provinces and the future Romance developments, in the later period a partial correlation is to be noticed (cf. Chart 3b in Figure 2c), even if not so clear as in Chart 2b (Figure 2b). For better comparison, in Charts 3a–3c (Figure 2c) we introduced an imperial average figure for “all provinces” (in Chart 3b in Figure 2c shown as  $32,852/11,358 = 2.9$ , cf. Chart 8b in Figure 1b). If we measure the provincial data density figures against this imperial average figure (2.9), in the later period we can see that Britannia is set clearly below it (with 2.2 by 0.7), and Moesia clearly above it (with 3.5 by 0.6), while Dalmatia, Venetia et Histria, and Pannonia all show roughly the imperial average value (with 3, with a deviation of 0.1 %). This way only Britannia (2.2), Gallia Narbonensis (3.9), and Lusitania (4.9) seem to be in their appropriate places on the scale of data density in the later period. However, if we move on to the changes in data density between the early and the later period (Chart 3c in Figure 2c), we notice two significant phenomena. First, the data density increased both on the imperial average (1.1) and in all provinces concerned. If we consider the pattern of growth rates, we see a clear correlation between this pattern and the future Romance linguistic geographical distribution. Britannia with no Romance continuation has the lowest growth rate (0.7), then Pannonia with no Romance continuation has the second lowest growth rate (0.9), and Dalmatia with

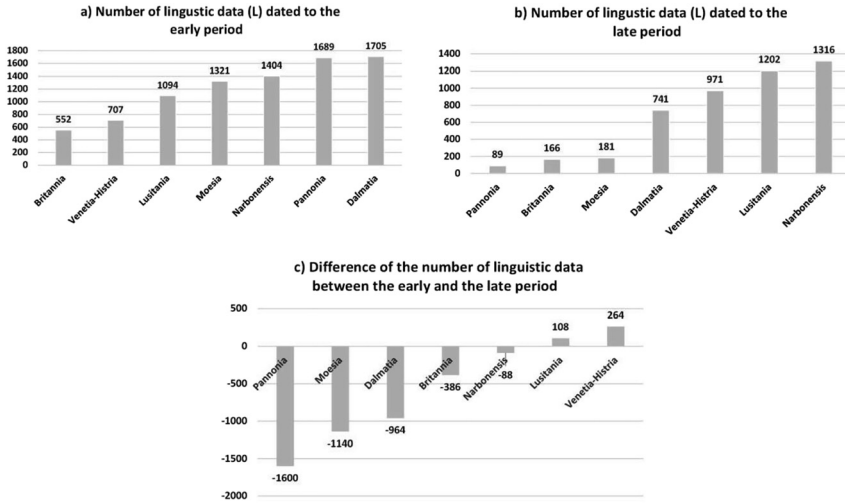


Figure 2d: Charts 4a–4c.

an extinct Romance continuation has the third lowest growth rate (1.1), corresponding to the imperial average (1.1), followed by Moesia with no Romance continuation and with a growth rate slightly over the imperial average (1.3), while the top trio consists of Venetia et Histria (1.5), Gallia Narbonensis (2.4), and Lusitania (3.2), in all of which we have a persistent Romance continuation. Here only the position of Moesia would require an explanation – which at this point could only be hypothetical and tentative, so we are not discussing it in detail right now. We can only risk a short explanation: later Moesia might have been the temporary settlement area of the Romanized population that left Roman Dacia after its abandonment and evacuation in the last third of the 3rd century, who were later forced to leave the territory of Moesia as well due to the waves of migration and later the Slavic tribes in the 5th and 6th centuries.<sup>18</sup> This increase in data density in later Moesia might reflect this temporary increase in a Latinized population which later vanished from there.

### 3.4 Analysis according to the number of linguistic data

Now let us turn to Charts 4a–4c and 5a–5c in Figures 2d and 2e, which exclusively deal with the purely linguistic data, excluding orthographic and extra-linguistic data as non-linguistic ones.

<sup>18</sup> For the complex and controversial problem connected to the prehistory of the speakers of the Rumanian language, see Adamik (2003: 680).

In Chart 4a we can immediately notice that in the early period there is again no correlation to be detected between data of the provinces and future Romance developments. In the later period, however (cf. Chart 4b), there is a clear correlation. What is more, both Charts 4a and 4b have a very similar or nearly identical pattern and the same ranking order of provinces as Charts 2a and 2b displaying both the linguistic and non-linguistic data. In both Charts 2b and 4b, in the top positions we find Gallia Narbonensis, Lusitania, and Venetia et Histria; in the middle, Dalmatia; and at the bottom, Moesia, Britannia, and Pannonia. At first glance this could mean that separating linguistic and non-linguistic data was not too profitable, but, as we shall see, such a conclusion is not justified at all. The advantage of such a separation is clearly visible if we go to Chart 4c, displaying the changes in the number of purely linguistic data between the early and the later period, and compare it with Chart 2c, which includes the non-linguistic data as well. Firstly, regarding data growth, the prominence of Venetia et Histria (+264) and Lusitania (+108) as leading terrains of Latinization is much more evident in Chart 4c than in Chart 2c (Venetia et Histria +142, Lusitania –75). Secondly, Dalmatia (–965) with an extinct Romance continuation precedes Moesia (–1,140) with no Romance continuation in Chart 4c, while in Chart 2c Dalmatia (–2,085) lags behind Moesia (–1,686). Through the analysis excluding the non-linguistic phenomena (as in Chart 4c) we were able to identify a more conspicuous concordance with the Romance outcomes than through an analysis including non-linguistic phenomena as well (as in Chart 2c).

### 3.5 Analysis according to the rates of linguistic data

A further benefit of the method of excluding non-linguistic phenomena can be seen in Charts 5a–5c, displaying the rates of linguistic data as measured against non-linguistic phenomena.

Here not so much the separate percentages by province are interesting and expressive (displayed in Charts 5a and 5b), but their comparison displayed in Chart 5c, where we can see the changes in the rates of linguistic data between the early and the later period. It is of course a welcome sight that all provinces having some kind of Romance continuation are in the top positions: Gallia Narbonensis (+14), Dalmatia (+13 %), Venetia et Histria (+12 %), and Lusitania (+9 %), even if Britannia with no Romance continuation seems to join them (+12%) as an odd one out (in need of a special explanation). It is even more interesting to find Pannonia in the last position, as the only province where the rate of linguistic data did not increase but instead decreased (–3 %). This means that while linguistic Romanization, i.e. Latinization, the spread of vulgar Latin became more and more

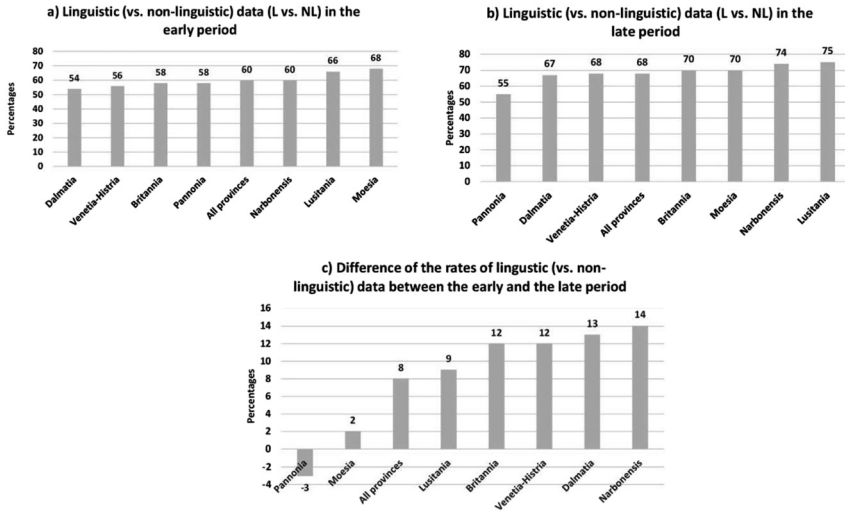


Figure 2e: Charts 5a–5c.

intensive and advanced in nearly all selected provinces and also on an imperial average (+8 %), in Pannonia it started to recede and weaken, paving the way to a complete evanescence of any Latin-speaking communities from Pannonia, hindering the potential birth of a Romance language there.<sup>19</sup>

### 3.6 The inadequacy of a comparative analysis of epigraphic corpora

Before we draw our final conclusions, a short detour seems to be necessary to the question why a simple comparative analysis of epigraphic corpora of selected provinces (by including all inscriptions, that is, also correct inscriptions without any misspellings) is not sufficient for answering our initial questions. Firstly, in terms of chronological classification, we have no reliable figures for the actual number of inscriptions in either the early or the later periods in any of the selected provinces. We only know how many inscriptions survived, but not all surviving inscriptions are properly dated – which would be vital for our analysis. Only inscriptions from a few provinces are fully registered and dated in epigraphic databases; in our case, this means only Pannonia, Dalmatia and Moesia, fully

<sup>19</sup> On the problem of de-Romanization in Pannonia, see Adamik (2003: 679–680, “Entromanisierung in Pannonien”).

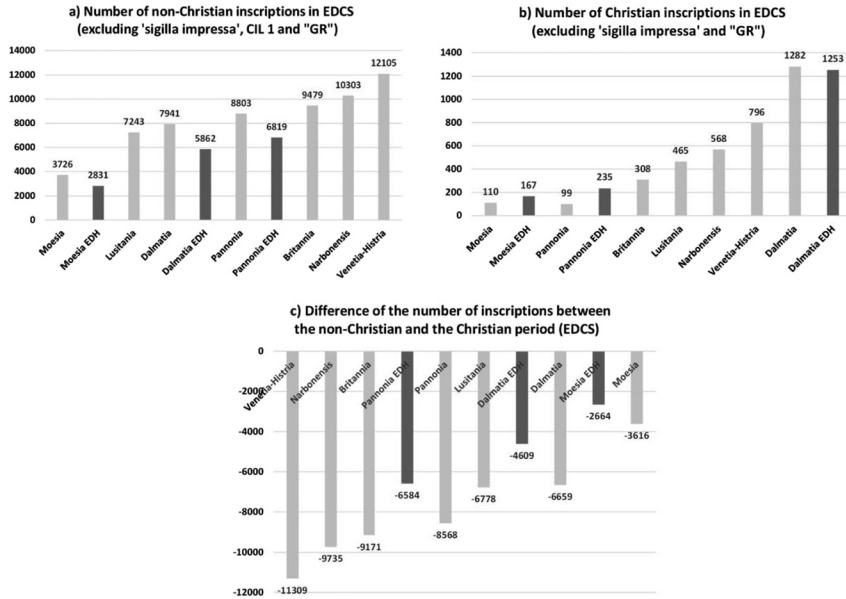


Figure 2f: Charts 6a–6c.

processed in the Epigraphic Database Heidelberg, which we can consider the golden standard. As for Venetia et Histria, we can use the data in Epigraphic Database Roma, but this province is not yet fully processed in the EDR. As for Lusitania, unfortunately we cannot use the Hispania Epigraphica Online database, since its search module does not have a query possibility either for datation or for religion (the latter is vital for differentiating Christian and non-Christian inscriptions). For Gallia Narbonensis, we have nothing but the data of the Epigraphik-Datenbank Clauss / Slaby (EDCS).<sup>20</sup> As a result, what we could do for every province concerned here was to use the most comprehensive database of EDCS, which only allows a modest chronological distinction of inscriptions (as most of the available dates are still waiting to be uploaded into this database) but at least has a query option for Christian and non-Christian inscriptions. Since non-Christian inscriptions can mostly be equated with *pre*-Christian inscriptions, a division between non-Christian and Christian inscriptions can be roughly equated

<sup>20</sup> EDCS = Epigraphik-Datenbank Clauss / Slaby (<http://db.edcs.eu/epigr/>), EDH = Epigraphische Datenbank Heidelberg (<http://edh-www.adw.uni-heidelberg.de/home>), EDR = Epigraphic Database Rome ([http://www.edr-edr.it/English/index\\_en.php](http://www.edr-edr.it/English/index_en.php)), HEpOnl = Hispania Epigraphica Online (<http://eda-bea.es/>).

with a division between inscriptions dated to the early and the later period. Accordingly, Charts 6a–6c in Figure 2f display the distribution of non-Christian (i.e. mostly pre-Christian) and Christian inscriptions, and the changes (i.e. decreases) in the number of inscriptions between the non-Christian (i.e. pre-Christian) and the Christian period as shown in the EDCS.<sup>21</sup> To supplement these data, I also added figures from the EDH for the provinces fully processed there.<sup>22</sup>

As you can see, however much Chart 6b seems to reflect the future Romance situation, it fails entirely in the leading position, where you can see a high prominence of later Dalmatia with an extinct Romance continuation. This ranking fails to express that the territory of Roman Dalmatia was far from having the most flourishing Romance population. Precaution is also necessary while using this kind of data of the EDCS. While regarding later Dalmatia the figure for Christian inscriptions in EDCS (1,282) and that of inscriptions dated to the late period (i.e. 301–700 AD) in EDH (1,253) are essentially identical (2.3 % difference), the same cannot be said of data concerning later Moesia and Pannonia: Pannonia EDCS 99 (104 by including “sigilla impressa”) vs. EDH 235 (57.9 % difference), and Moesia EDCS 110 (114 by including “sigilla impressa”) vs. EDH 167 (34.1 % difference).

The other reservation about involving all inscriptions regardless of whether they are correct or not, that is, whether they do or do not have misspellings, is rather theoretical. The presence of correct inscriptions attests the presence of the Latin language, but not necessarily the presence of Latin *speakers* (meaning people who spoke Latin as a mother tongue or as a first or main language after a language shift from the vernacular language). Correct inscriptions may be erected by people who did not speak Latin at all, if they copied formulas.<sup>23</sup> However, inscriptions containing linguistic errors suggest the presence of people who actually spoke the language to some degree. This way only incorrect inscriptions can give positive evidence of the presence of local speakers of Latin.

---

**21** Excluding “*inscriptiones christianae*”, “*sigilla impressa*” and Greek inscriptions (“GR”) by EDCS’ search engine and then manually removing republican inscriptions (i.e. those published in CIL 1) from the data set. I used EDCS and EDH data as of 3 November 2020.

**22** In the case of EDH, which indicates a datation for most inscriptions, I used the chronological setting of 27 BC – 300 AD and 301–700 AD to search for Latin inscriptions in the selected provinces using its advanced search module (<https://edh-www.adw.uni-heidelberg.de/inschrift/erweiterteSuche>).

**23** This statement is otherwise true of inscriptions that contain only non-linguistic errors.



## 4 Conclusions

To sum up, we can try and answer the questions asked in the introduction of my paper in a mostly positive way. The process of Latinization of the Empire is traceable by the data of our Database. According to our data, the linguistic Romanization, i.e. Latinization, the spread of spoken or vulgar Latin, became variably more and more intensive in all concerned provinces, since data density increased in the inscriptional sources, even if to different degrees (cf. Chart 3c in Figure 2c). What is more, in many investigational aspects i.e. regarding the patterns of distribution in number of data of all kinds (Chart 2b), of purely linguistic data (Chart 4b), and of rate changes in data density (Chart 3c), we were able to find differences between the selected provinces of the Roman Empire corresponding mostly to the future Romance outcomes of the respective areas (regardless of whether that outcome was positive or negative).

The picture was then successfully refined by an analysis excluding non-linguistic phenomena (4c) that produced a more evident and explicit concordance with Romance outcomes than that of including the non-linguistic phenomena (e.g. Chart 4c). Thanks to this refined method of excluding non-linguistic phenomena from the survey (as used in Charts 4a–4c and 5a–5c of Figures 2d and 2e), we were able to determine (i) the actual position of Dalmatia corresponding better to the Romance outcomes (cf. Chart 4c compared to 2c), and (ii) that in late Pannonia the process of Latinization became slower, since here the rate of linguistic vs. non-linguistic data decreased (by 3 %). This finding helps us better understand the situation in late Pannonia characterized by an explicit decline of Latinization in the 4th–5th centuries, hindering the potential birth of a Romance language there. Accordingly, it is evidently pointless to use such (now obsolete) concepts as “Pannono-Romanisch,” which Reichenkron originally came up with to create a link between the Western Romance Rhaeto-Romance and the Eastern Romance Rumanian language.<sup>24</sup>

---

<sup>24</sup> See Reichenkron (1959), who also has reservations: “Viel wichtiger ist die Verbindung zwischen Rumänen und Rätoromanen, also wieder die Frage der Grenze der West- und Ostromania. Das Verbindungsglied kann nur in Pannonien, also dem Pannono-Romanischen, gelegen haben, das aber sicher, sofern es überhaupt Zeit hatte, sich zu der Vorstufe einer ostromanischen Sprache zu entwickeln, schon bald im Slaventum und danach im Madjarentum aufgegangen ist. Ob diese ostromanische Schicht in Pannonien überhaupt einmal zu fassen sein wird, ist noch sehr die Frage” (Reichenkron 1959: 170). Cf. also Mihăescu (1993): “Accepter de prendre en considération une langue romane de Pannonie (‘Pannono-Romanisch’) et escompter sa restitution ne saurait être qu’un *pium desiderium* en raison de la carence des moyens d’information à ce sujet” (Mihăescu 1993: 153).

All in all, the analysis of data in our Database can contribute to solving the complex problem of Latinization, and is a lot more appropriate for this purpose than a simple comparative analysis of epigraphic corpora of the selected provinces.

**Acknowledgments:** The present paper was prepared within the framework of the NKFIH (National Research, Development and Innovation Office) projects no. K 124170 and 135359 entitled “Computerized Historical Linguistic Database of Latin Inscriptions of the Imperial Age” (see: <http://lldb.elte.hu/>) and of the project “Lendület (‘Momentum’) Research Group for Computational Latin Dialectology” (Research Institute for Linguistics). I wish to express my gratitude to Silvia Tantimonaco for her valuable comments and Zsuzsanna Sarkadi for her help in the revision of the English text.

## References

- Adamik, Béla. 2003. Die verlorene Romanität in Mösien, Thrakien und Pannonien. In Gerhard Ernst, Martin-Dietrich Gleßgen, Christian Schmitt & Wolfgang Schweickard (eds.), *Romanische Sprachgeschichte. Ein internationales Handbuch zur Geschichte der romanischen Sprachen 1*, (Handbücher zur Sprach- und Kommunikationswissenschaft 23[1]), 675–683. New York & Berlin: Walter de Gruyter.
- Adamik, Béla. 2006. Offizielles Kommunikationssystem und Romanisierung. In Carmen Arias Abellan (ed.), *Latin vulgaire – latin tardif VII.: Actes du VII<sup>ème</sup> Colloque international sur le latin vulgaire et tardif, Séville, 2–6 septembre 2003*, 17–29. Sevilla: Universidad de Sevilla.
- Adamik, Béla. 2009. In memoriam József Herman: von der Late Latin Data Base bis zur Computerized Historical Linguistic Data Base of Latin Inscriptions of the Imperial Age. *Acta Antiqua Academiae Scientiarum Hungaricae* 49. 11–22.
- Adamik, Béla. 2012. In search of the regional diversification of Latin: Some methodological considerations in employing the inscriptional evidence. In Frédérique Biville, Marie-Karine Lhommé & Daniel Vallat (eds.), *Latin vulgaire – latin tardif IX.: Actes du IX<sup>e</sup> colloque international sur le latin vulgaire et tardif, Lyon, 6–9 septembre 2009*, 123–139. Lyon: Maison de l’Orient e la Méditerranée.
- Adamik, Béla. 2015. The periodization of Latin: An old question revisited. In Gerd V. M. Haverling (ed.), *Latin linguistics in the early 21st century: Acts of the 16th International Colloquium on Latin Linguistics, Uppsala, June 6th–11th, 2011*, 640–652. Uppsala: Uppsala Universitet.
- Adamik, Béla. 2016. Computerized Historical Linguistic Database of the Latin Inscriptions of the Imperial Age: Search and charting modules. In Ádám Szabó (ed.), *From Polites to Magos: Studia György Németh sexagenario dedicata*, 13–27. Budapest & Debrecen: University of Debrecen.
- Adams, James N. 2007. *The regional diversification of Latin, 200 BC–AD 600*. Cambridge: Cambridge University Press.

- Adams, James N. 2013. *Social variation and the Latin language*. Cambridge: Cambridge University Press.
- Binder, Vera. 2006 [1999]. Latinization. In Hubert Cancik, Helmuth Schneider, Christine F. Salazar, Manfred Landfester & Francis G. Gentry (eds.), *Brill's New Pauly, Antiquity volumes*. Brill: Leiden & Boston. [https://doi.org/10.1163/1574-9347\\_bnp\\_e632280](https://doi.org/10.1163/1574-9347_bnp_e632280) (accessed 29 January 2019) = Binder, Vera. 1999. Latinisierung. In Hubert Cancik & Helmuth Schneider (eds.), *Der Neue Pauly: Enzyklopädie der Antike* 6, 1175–1174. Metzler: Stuttgart & Weimar.
- García y Bellido, Antonio. 1972. Die Latinisierung Hispaniens. In Hildegard Temporini (ed.), *Aufstieg und Niedergang der römischen Welt* 1, 1, 462–500. Berlin & New York: De Gruyter.
- Herman, József. 2000. *Vulgar Latin*. translated by R. Wright. University Park, PA: The Pennsylvania State University Press.
- Mihăescu, Haralambie. 1993. *La romanité dans le Sud-Est de l'Europe*. București: Editura Academiei Române.
- Raupach, Manfred. 1996. Expansion und Rückzug des Lateins. In Günter Holtus, Michael Metzeltin & Christian Schmitt (eds.), *Lexikon der Romanistischen Linguistik* 2, 1. *Latein und Romanisch: Historisch-vergleichende Grammatik der romanischen Sprachen*, 5–19. Niemeyer: Tübingen.
- Reichenkron, Günter. 1959. Das Ostromanische. In Balduin Saria (ed.), *Völker und Kulturen Südosteuropas: Kulturhistorische Beiträge*, 153–172. München: Südosteuropa-Verlagsgesellschaft.
- Smith, Colin. 1983. Vulgar Latin in Roman Britain: Epigraphic and other evidence. In Wolfgang Haase (ed.), *Aufstieg und Niedergang der römischen Welt* 2, 29, 2, 893–948. Berlin & New York: De Gruyter.
- Spickermann, Wolfgang. 2006 [2001]. Self-Romanization. In Hubert Cancik, Helmuth Schneider, Christine F. Salazar, Manfred Landfester & Francis G. Gentry (eds.), *Brill's New Pauly, Antiquity volumes*. Brill: Leiden & Boston. [https://doi.org/10.1163/1574-9347\\_bnp\\_e1024520](https://doi.org/10.1163/1574-9347_bnp_e1024520) (accessed 29 January 2019) = Spickermann, Wolfgang. 2001. Romanisation. In Hubert Cancik & Helmuth Schneider (eds.), *Der Neue Pauly: Enzyklopädie der Antike* 10, 1121–1122. Metzler: Stuttgart & Weimar.
- Wolf, Greg. 2006 [2001]. Romanization. In Hubert Cancik, Helmuth Schneider, Christine F. Salazar, Manfred Landfester & Francis G. Gentry (eds.), *Brill's New Pauly, Antiquity volumes*. Brill: Leiden & Boston. [https://doi.org/10.1163/1574-9347\\_bnp\\_e1024530](https://doi.org/10.1163/1574-9347_bnp_e1024530) (accessed 29 January 2019) = Wolf, Greg. 2001. Romanisierung. In Hubert Cancik & Helmuth Schneider (eds.), *Der Neue Pauly: Enzyklopädie der Antike* 10, 1122–1127. Metzler: Stuttgart & Weimar.