Acta Antiqua Academiae Scientiarum Hungaricae 63 (2023) 4, 575–594
DOI: 10.1556/068.2024.00123

AKADÉMIAI KIADÓ

Gender confusions and other linguistic changes
A provisional description of Vulgar Latin Phenomena

BÉLA SZLOVICSÁK[1,2*]

[1] HUN-REN Hungarian Research Centre for Linguistics, Institute of Historical Linguistics and Uralic Studies, Research Group for Latin Historical Linguistics and Dialectology, Budapest, Hungary

[2] ELTE Eötvös Loránd University, Doctoral School of Linguistics, Ancient Studies Doctoral Programme, Budapest, Hungary

RESEARCH ARTICLE

Received: November 12, 2023 • Revised manuscript received: February 21, 2024 • Accepted: February 22, 2024
Published online: May 30, 2024

ABSTRACT

The main goal of this paper is to provide a preliminary examination of the interaction between the Vulgar Latin grammatical gender system and other levels of linguistic change, such as phonological confusions. To achieve this description conditional inference trees and random forests were fitted to our data which enabled a more thorough understanding of these interactions than would be possible to notice without statistical methods.

KEYWORDS

Vulgar Latin, dialectology, neuter, grammatical gender, LLDB database, inscriptions, conditional inference trees, random forests, computational linguistics

INTRODUCTION: PROBLEM STATEMENT, RESEARCH HISTORY AND METHODOLOGY

The transformation of the grammatical gender system during the Vulgar Latin period is a particularly interesting area from a linguistic point of view. The sources that evidence this

* Corresponding author. E-mail: szlovicsak.bela@nytud.hun-ren.hu

change are not particularly frequent, but they can provide sufficient information about the exact course of the transformation. The aim of this paper is, as the title suggests, to examine and control for a variable that may influence our understanding of this transformation. I will also present and evaluate a possible statistical model for the analysis of these variables.

There have been several recent works on the transformation of the grammatical gender system. Among these, the most important is a monograph written by Loporcaro,[1] in which he examines the gender system in detail, starting with Latin and thoroughly inspecting some of the Romance languages. Of particular interest here are those Romance dialects he presents where in some (possibly altered) form the neuter has been preserved,[2] as this gender has, in fact, largely disappeared from most Romance languages.[3] Moreover, I have also examined this transformation in a previous paper.[4] However, in contrast to Loporcaro's monograph, I have investigated the transformation of the gender system using the inscriptional material and the *Computerized Historical Linguistic Database of the Latin Inscriptions of the Imperial Age* (hereafter LLDB database).[5] In the present paper, I follow my earlier direction and investigate the grammatical gender changes using the data obtained from the LLDB database. However, in contrast to my previous work, I have sought to eliminate one confounding variable, which will allow for a more confident interpretation of the data and an examination of the significance of these other effects.

The inscriptional material and the LLDB database are particularly useful for this research in several ways. On the one hand, the appropriate temporal and spatial distinction of the inscriptions allows the observation of dialectological differences and linguistic changes. In addition, the fact that the same data can be recorded with not only one, but with two different codes is a particularly useful useful asset of this database. That is, if a gender confusion can also be interpreted as a consonant confusion for example, this fact is recorded.[6] It should be stressed, however, that the decision whether a given interpretation is entered into the code or into the alternative code field does not imply a preference.[7] The alternative code has the same validity as the main code,[8] it was just placed second. Consequently, for ease of data handling, I treated the gender confusion codes under analysis as the main codes and other possible interpretations as alternative codes, no matter what order they appeared in the LLDB database data sheets.

[1] Loporcaro (2018).

[2] Loporcaro (2018) 12–14.

[3] Loporcaro (2018) 63–70.

[4] Szlovicsák (2022).

[5] See: https://lldb.elte.hu/en. The data in this paper reflect the status of the LLDB database as of 20/11/2022.

[6] E.g.: LLDB-110320: VEX[IL]LVM ARGENT | INSIGNEM = *vexillum argento insigne*. The code for this data is *masc. pro neutr.*, the Alternative code is *-ø > -m*. In this paper phonological "errors" recorded in the LLDB database are referenced as "confusions", even if the given "error" results in the disappearance of a phoneme.

[7] See the LLDB Guidelines for Data Collection: II/1.2 https://lldb.elte.hu/admin/doc_guidelines.php (last accessed 15/02/2024).

[8] I use the term *main code* interchangeably with the term *code*, especially when I want to emphasize its contrast with the Alternative code. This is particularly important when it comes specifically to gender confusion, which I always treat as the main code.

Alternative codes can make the interpretation and validity of data significantly more difficult. Suppose that, as in the example in Note 6, all confusions between the masculine and neuter gender could be explained as consonant confusions. Then, if alternative codes were not included in our analysis, it might appear as if the distribution of gender confusions were directly related to, for example, spatial or temporal differences, while it is clear (in this example) that this is not the case, but that spatial differences in consonant confusions would account for the differences in gender confusions, not the spatial differences themselves. A model without alternative codes in this case would not be able to present the real explanation for the differences. The possible ways in which alternative codes might affect gender confusion and the effect they might have on our picture of it will be discussed in more detail later. The aim of this paper is therefore to control the effect of alternative codes using statistical tools and, by doing so, to test whether my previous results hold up or taking alternative codes into account makes the effect of temporal and spatial differences on gender confusions negligible. The utility of statistical tools for the validation of different observations was already shown by Papini for the methods of Herman.[9] This paper therefore seeks to achieve similar results on the interaction between gender confusions and other linguistic changes.

Within gender confusion, I distinguish between two possible interpretations, to one I refer to the hypercorrect reading, to the other as the non-hypercorrect reading.[10] In the present paper I will primarily focus on the non-hypercorrect reading and only present the hypercorrect reading and its possible problems for the sake of comparison with previous results.[11] Thus, I distinguish between three levels within the Main code: *Masc/Fem*, *Fem/Masc* > *Neutr* and *Neutr* > *Fem/Masc*, where the first level contains those confusions, where masculine nouns are incorrectly used as feminine or *vice versa*, the second level contains non-neuter words that are used in neuter, while the third level serves as the main evidence for the disappearance of the neuter, where neuter words are used in other genders.[12] In the case of the hypercorrectly interpreted Main code, the first level is identical to the first one here, but the other two are modified to *Masc/Neutr* and *Fem/Neutr* levels, where in the first case the neuter is confused with the masculine, while in the second case it is confused with the feminine.[13]

[9] Papini (2022) 374.

[10] For example, LLDB-111677: HOC TITVLVM = *hunc titulum*. Main code: *neutr. pro masc.* since the masculine form *titulus* is accompanied by the neuter *accusative* (or *nominative)* form of the demonstrative pronoun. Data of this type (and similarly data coded *neutr. pro fem.*) can be interpreted both hypercorrectly and non-hypercorrectly. The hypercorrect reading is that the neuter stands for the masculine gender because the distinction between the two genders is weak, which is evidenced by the fact, that the neuter disappeared during/after the Vulgar Latin period (a reading also criticized by Loporcaro [2018] 12–14). By contrast, a non-hypercorrect reading accepts that this type of gender confusion (i.e., *neuter* instead of another gender) could have resulted from a genuinely existing persistence of this gender, where the neuter could in some cases, contrary to what was expected, incorporate new nouns, and not disappear. This second reading is particularly important in light of the fact that Loporcaro has presented several dialects where the distinction of the neuter has been preserved in some way. Loporcaro (2018) 60.

[11] See Szlovicsák (2022) 423 where I followed the hypercorrect interpretation.

[12] Based on the LLDB codes: *masc/fem* = *fem. pro masc.*, *masc. pro fem.* and *fem. per communi.*; *fem/masc* > *neutr* = *neutr. pro masc.* and *neutr. pro fem.*; *neutr* > *fem/masc* = *masc. pro neutr.* and *fem. pro neutr.*

[13] Thus, according to the codes used in the LLDB database: *masc/fem* = *fem. pro masc.*, *masc. pro fem.* and *fem. pro communi.*; *masc/neutr* = *neutr. pro masc.* and *masc. pro neutr.*; *fem/neutr* = *neutr. pro fem.* and *fem. pro neutr.*

POSSIBLE EFFECTS OF ALTERNATIVE CODES

Not all data in the LLDB database have alternative interpretations, but for those that do have one, the question of which interpretation is more likely and whether an interpretation predicts another is always present. The present paper attempts to investigate how well the distribution of alternative codes describes the distribution of gender confusions. Does the transformation of the gender system form an independent direction within the transformation of the grammatical gender system or is it completely determined by transformations at other levels of the system. Indeed, if the findings would suggest that only alternative codes really determine the distribution of gender confusions, this would be a relevant result, but it would significantly nuance what I have shown earlier,[14] that a correlation was observed between the spatial distribution of the data and the distribution of gender confusions. For then it would become clear that the spatial differences and the temporal variation were merely due to and determined by the transformation of other grammatical subsystems.

Among the Alternative codes of gender confusions there is a high frequency of different phonetic confusions. See for example the following error: TABVLAS () PICTA = *tabulas pictas* (LLDB-15616). This data has two interpretations. First, one can think of this error as resulting from the pronunciation uncertainty of the word-final *-s*, the importance of which has been shown by Paulus.[15] This interpretation is also listed as the Main code in the LLDB database. In addition, however, we can think of this data as resulting from the transformation of the gender system and showing either that the neuter is weak, or even that it is replacing the feminine in the case of some nouns. Beyond phonological transformations, the transformation of the Classical Latin case system might have also had an effect on the gender system. Consider the following error: EX VOTA = *ex votis / voto* (LLDB-1177). This might be an example of gender confusion, where the neuter plural of a word becomes feminine singular. However, another explanation for this error might be the transformation of the case system. As this data might be the result of the confusion between the accusative and the ablative. A third type of Alternative codes also need to be taken into consideration. There are cases where gender confusions might be results of syntactical uncertainties. In the case of the following error: FILIABVS | SVIS STERCORIAE () IOVINO (|) LVCIO () INDVLGENTISSIMIS = *filiis suis Stercoriae () Iovino () Lucio () indulgentissimis* (LLDB-31750) the gender confusion could also be explained by an uncertainty in agreement.

In any case, it is not at all clear whether gender confusion is a clear feature of these errors or whether they can simply be explained by some sort of other linguistic transformation. In many cases, this question is of the utmost importance: is it always the case that the various levels of the Latin language system are independent of one another and that their rearrangements do not result in other transformations? In the case of significant interaction, it would never be enough to look at individual sub-systems. If we want to know the exact characteristics of a given level, we have to take into account other linguistic levels that might have an impact on it.

[14] Szlovicsák (2022) 431–434.

[15] Paulus (2020) 125.

Adamik[16] has already addressed this issue in connection with the disappearance of the world final *-m*. In this connection, he,[17] after carefully examining the various contexts of the confusions, concluded that in the context of this process of transformation, in some well-defined syntactic situations the uncertainty of the word final *-m* cannot be ruled out as being explained by case confusions, while in other cases it is clearly a matter of phonetic transformation. For the time being, I have used statistical tools to investigate whether, in general, the independence of confusions can be ruled out, i.e. whether the transformation of the grammatical gender system is entirely due to other linguistic changes, or whether there are factors in this transformation that can be explained by spatial and temporal differences alone.

It is therefore not exactly clear what the impact of the Alternative codes is on the previous observations and therefore needs exploration. The possible effects of this variable on gender confusions can be visualised using the three graphs below (Figs 1–3). The letters here correspond to Provinces, Dates, i.e., the presumed date of origin of the data, Alternative codes, and Gender confusions i.e., the Main codes, while the arrows illustrate the possible relationships between them. Thus, if there is an arrow between two variables in a potential model, it symbolizes that one has an effect on the other. The fact that spatial and temporal differences affect different phonological and morphological changes has been shown, among others by Paulus[18] in the context of word final *-s*. Hence, there is certainly a correlation between these three variables, spatial and temporal differences, and phonological changes, which is reflected in all three potential models.[19] The question I am investigating is how gender confusion is incorporated into this picture.

These graphs are also of great statistical importance, as they help us formulate statements about the data and the relationship between the variables, which can be used to turn our preconceptions about the data into scientifically verifiable statements. They can also be used to visualise causal relationships in a straightforward way, especially when working with relatively few variables. Cinelli, Forney, and Pearl[20] provide a very illustrative introduction to their precise use. The key point for the present paper regarding their treatment is that including a variable in a model "absorbs" the effect of variables that affect the dependent variable through that variable. Thus, for example, in the first model (Fig. 1), the inclusion of the Alternative codes "blocks" the effect of the other two variables on gender confusion, since they only affect the dependent variable through the Alternative codes.

Figure 1 shows that spatial and temporal differences have no direct effect on gender confusion. In this case, the previously observed correlations only appear because of their effect on Alternative codes. If Alternative codes are taken into account here, the effect of spatial and temporal differences on gender confusions disappear. In this case, it would be sufficient to consider Alternative codes to characterise the distribution of gender confusions with high

[16] Adamik (2019).

[17] Adamik (2019) 107–108.

[18] Paulus (2020) 141–143.

[19] With the grouping of Alternative Codes used here (see below), these correlations cannot be easily detected, so it will be important to achieve a suitable refinement in the future. Achieving this is a significant challenge, but even without it the present paper can answer the question of whether Alternative codes have an effect on gender confusions.

[20] Cinelli et al. (2020).

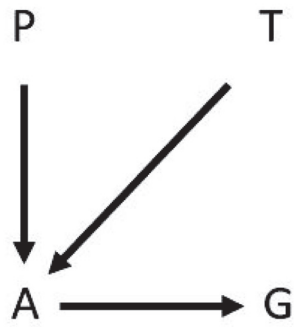
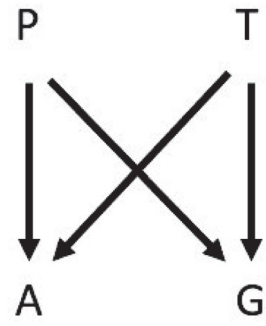
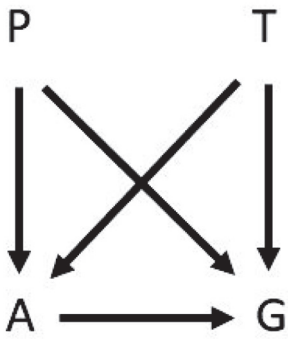
Possible models for the impact of Alternative codes	
	**Figure 1** P: Province T: Date A: Alternative code G: Gender confusion X→ Y: variable X has an effect on variable Y
	**Figure 2** P: Province T: Date A: Alternative code G: Gender confusion X→ Y: variable X has an effect on variable Y
	**Figure 3** P: Province T: Date A: Alternative code G: Gender confusion X→ Y: variable X has an effect on variable Y

Fig. 1–3. Possible models for the impact of alternative codes

confidence. In the case of the scenario shown in Fig. 2, the Alternative codes have no effect on the Main codes, that is, they do not really allow us to describe gender confusions. If this were the case, the previous results would be fully true since the inclusion of Alternative codes would not change anything. Figure 3 outlines the possibility that all three variables influence gender confusion. Provinces and Date even affect these confusions indirectly, through the Alternative

codes. If this is the case, the previous results may remain partially valid, but may need to be changed, while the Alternative codes may also affect gender confusions. This would therefore allow us to observe the direct effect of Provinces and Dating on gender confusions.

We can rule out the second of these possibilities if we observe a correlation between Alternative codes and gender confusion. That is, if it becomes clear that there is a relationship between these two variables. The statistical model then helps us choose between the first and the third option. It examines whether the effect of Provinces and Dating disappears when the Alternative codes are taken into account. If the effect of the provinces disappears when the Alternative code is considered, the first option should be adopted. If, however, the effect of the regional and temporal differences is not negligible when these are considered, the third option describes best the real processes. In this paper I will show why this third case holds.

THE DATA UNDER CONSIDERATION

The data examined in this paper was obtained from the LLDB database by exporting all data sheets whose Main or Alternative code was a gender confusion[21] and then manipulating the data using the statistical programming language R[22] and the integrated development environment RStudio. I also used this program and some of its libraries to generate the charts, and for statistical analyses.[23] I have grouped the Main code, Alternative code and the Province variables of the data extracted from the LLDB database. The resulting areas are Africa, Gaul and Germania, Hispania, Southern Italy, Northern Italy, Illyricum, and Rome. All other areas were excluded from the analysis. After limiting the period under study to the 1st–7th centuries AD, 643 observations, that is, 643 records containing gender confusions were left. Regarding dating, it should be pointed out that in many cases no exact date can be given, only an interval. For these intervals, I have treated the arithmetic mean of the interval cut-off points as the date of the data, similar to the latest dating function of the LLDB database.[24] However, in contrast to the approach used previously,[25] I treated the dates of the data as continuous variables rather than categorical, which allowed the use of more complex statistical models.[26] As this type of dating differs from the procedure followed so far by researchers using the LLDB database,[27] I have also included a variable which contained the Dating as a categorical variable, with the same periods

[21]See n. 11.

[22]R Core Team (2022).

[23]*R* version 4.2.2, *RStudio* 2022.12.0+353 "Elsbeth Geranium" version, packages used: *tidyverse, readxl, writexl, broom, dagitty, rethinking, partykit*.

[24]See the *Period [A]* function: https://lldb.elte.hu/admin/search_2.php (last accessed 15/02/2024).

[25]Szlovicsák (2022) 423–426.

[26]A categorical variable would be included here if the Date variable were divided into different periods. If we treat the Date as a continuous variable, we assume that it can take any value between 1 and 700.

[27]See, for example, Paulus (2020) 129–130, where linguistic change is also examined in two periods (Earlier period: 1st–3rd centuries AD, Later period: 4th–7th centuries AD). On reservations against examining dating as a continuous variable, see for example. Van de Velde–Petré (2020) 346–347.

as I used earlier.[28] To distinguish between these two dating variables, the continuous one will be called Year, while the categorical one will be called Period.

I split the Alternative code into four levels. These are *Phonologia, Nominalia, Syntactica etc.*[29] and *None*.[30] The last one is self-explanatory, representing the situation where there is no alternative interpretation of the gender confusions. The remaining three levels correspond to the different linguistic levels that can explain gender confusions. Among these, we can expect results especially for the *Phonologia* and *Nominalia* levels, as the role and the way of interaction of these different linguistic levels in the restructuring of Vulgar Latin is often raised in the literature, for example in the summary works of Väänänen[31] and Herman,[32] where we find some reflections on these issues. In addition, Löfstedt[33] has shown the possible stability of the neuter precisely in the context of the transformation of the case system, using a much more concentrated source material than the one considered in the present paper.[34] The broad categorization of the Alternative codes was done to enable the easier use of statistical tools and help with the interpretation of the data. Even though the statistical tools (conditional inference trees and

[28] Szlovicsák (2022) 423–426.

[29] The code *Syntactica etc.* has become outdated, as the current Code used in the LLDB database is *Syntactica et lexica*. As the difference is not significant in the case of this paper, I kept the previous name.

[30] The *Nominalia* and *Syntactica etc.* categories have been used because they are themselves broad coding categories occurring in the LLDB database. I also included codes of the type *Errores non grammatici* in the *Syntactica etc.* level, as the two categories had a rather small number of elements, but did not differ in their proportions (i.e., the distribution of gender confusions within these codes did not show any difference). In the *Phonologia* category, I combined gender confusions that fall into either the *Vocalismus* or *Consonantismus* category in the LLDB database, i.e., these are vowel and consonant changes. The reason for the merging here was both the small number of elements, and also the similar behaviour. Furthermore, each category contains the following errors (I list only those that occurred in the examined data), according to their codes used in the LLDB database: *Phonologia = -s > ø, -s > ø elisa, -m >-ø, -ø > -m, -m > ø elisa, i > E, í > E, i: > E, í: > E, ae > I, e > I, áe > I, ú > O, u (+ voc) > ø, a > E, a/á > O, commutationes vocalium variae, c > Q / QV / CV, n (+ cons.) > ø. Syntactica etc. = ablativus absolutus pro participio coniuncto vel appositione, hypercorrectio, permixtio syntagmatum, varia ad congruentiam nomin. et adiect. pertinentia, onomastica (nomina grammatice conspicua), graecismus, sing. pro plur., plur. pro sing., litterae omissae, litterae perperam incisae, litterae superfluae, abbreviatio insolita. Nominalia = nom. pro acc., nom./abl. pro gen., nom./acc. pro abl., acc. pro nom., acc. pro gen., commutatio vel permixtio casuum aliorum, dat./abl. pro gen., gen. pro nom., decl. I per II, decl. I pro III, decl. II pro I, decl. III pro I, decl. III pro II, commut. in formatione pronominum, praep. > casus sine praep., cett. ad usum pron. pertinentia, commut. in decl. pron. hic, commut. in decl. pron. ille, commut. in decl. pron. relat.* Also included here (in the *Nominalia* category) are the gender confusion codes that were included in the Alternative codes. As some errors could be interpreted as two different kinds of gender confusions, for example LLDB-28296: GRADA D S D = *gradus de suo dederunt,* Main code in the LLDB database: *neutr. pro masc.*, and its Alternative code is *fem. pro masc* (in this case *-m > -ø* too). In such cases I did not change the order of the codes, so I kept the order chosen by the data collector, even if it was not particularly meaningful decision.

[31] Väänänen (1981) 103.

[32] Herman (2000) 65.

[33] Löfstedt (1961) 226–227.

[34] These scholars however don't deal with the questions of the spatial differences of these effects and the reliability of these claims. Papini (2022) 351–360 has already shown in a special case that applying statistical tools to the results of Herman (or other scholars) can help us prove their validity and see if some of those claims need to be revisited.

random forests)[35] used in this paper could be used with more variables.[36] However, my main goal was to examine, whether we need to deal with the Alternative Codes, and for this decision a broad categorization is enough. The distribution of Alternative codes with the categories mentioned is shown in Chart 1.

The data from this bar graph can be used to determine whether we need to deal with the Alternative codes. If the impact of Alternative codes on gender confusion becomes evident here, the model of the data shown in Fig. 2 can be dismissed and the treatment of Alternative codes becomes necessary. This relationship can indeed be seen from Chart 1. As, for example, the distribution of gender confusions with no Alternative code, i.e., the data marked *None* on the chart, is roughly even, while the *Phonologia* level is dominated by confusions indicating the possible stability of the neuter, and the *Nominalia* level is dominated by confusions indicating the disappearance of the neuter, although not in the same proportion as the phonological confusions indicate its' potential stability. It is also worth highlighting that the category of *Syntactica etc.* has the highest proportion of confusions between the masculine and feminine, and the lowest number of data, which makes this category somewhat problematic even after the broad categorization. As a counterpart to Chart 1, Chart 2 shows the relationship between the hypercorrectly interpreted gender confusions and the Alternative codes.

In this case, the distribution of data without an Alternative code is not nearly as even as it was in Chart 1. In the case of the present chart, a considerable proportion of the confusions between feminine and neuter are found among the data with *Phonologia* Alternative code,

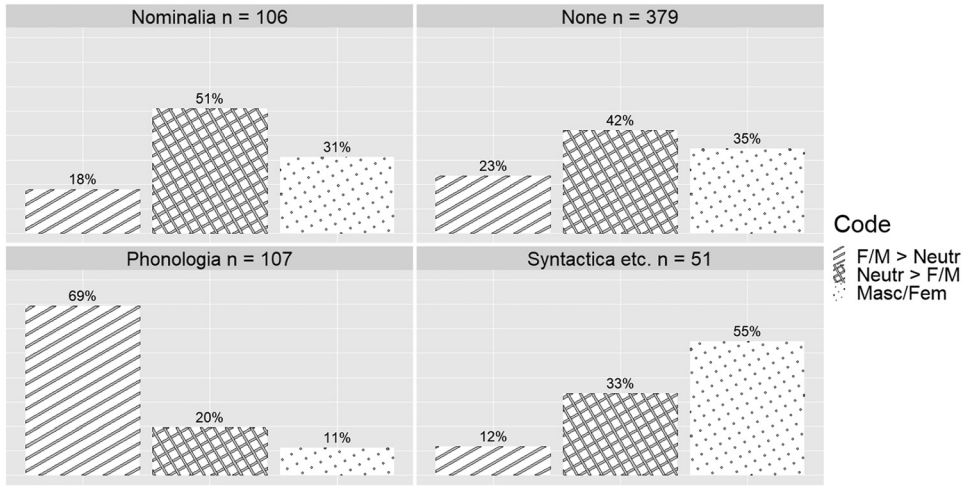


Chart 1. The effect of different Alternative codes on Gender Confusions Non Hypercorrect reading

[35]For an explanation of these methods, see below and also Levshina (2015) 291–299.

[36]Strobl et al. (2009) 340.

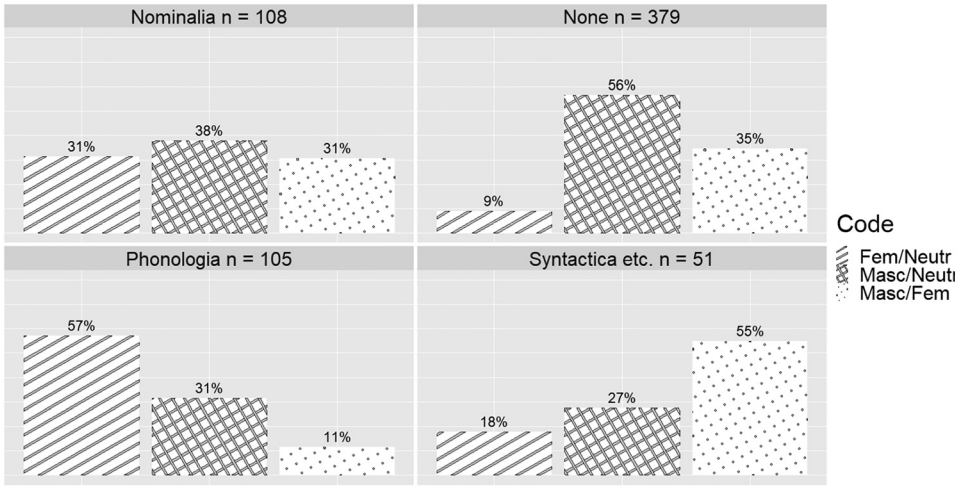


Chart 2. The effect of different Alternative codes on Gender Confusions Hypercorrect reading

suggesting that other grammatical transformations and in particular phonological confusions have played a particularly significant role in the formation of feminine-neuter confusions. It is also striking that the distribution of gender confusions with *Nominalia* Alternative code is fairly even, i.e., the distribution of the hypercorrect data was not particularly affected by these other confusions, whereas for the non-hypercorrect data a difference in the distribution of these confusions was noticeable (see Chart 1). The token numbers are of course not different from those seen previously and therefore the proportion of confusions between masculine and feminine has not changed either, so this remains the confusion with the highest proportion within the *Syntactica etc.* level. For this reason, this gender confusion is a suitable benchmark, because when we change the grouping of the neuter confusions, these confusions do not change.

For the gender confusions examined in this paper, there was no strong correlation between Alternative codes and spatial differences. This partly simplifies the interpretation of the data. Beyond this, however, it is still possible that there is an interaction between these two variables, i.e. there may be a difference in the spatial distribution of the Main code within a level of the Alternative code variable, i.e. the possible effect of the alternative code on the Main code may vary from area to area.[37] After examining the data (using Fisher's exact tests),[38] it was found that there is indeed an interaction between Alternative codes and Provinces, but that in no case did it

[37]See Agresti (2013) 53.

[38]This was used to decide whether the variables in the charts created along each level of the Alternative code (out of which only Chart 3 is depicted in this paper), were independent of each other or not. So, if there is an interaction, we can observe that the distribution of the Main codes per area changes when the Alternative code is changed. Alternatively, symmetrically, the effect of the Alternative codes on the distribution of the Main code varies by province. This can be checked by examining whether there is a correlation between the spatial distribution and the distribution of the Main codes, while being narrowed down to a given Alternative code.

occur at all levels. This means that their interaction is more intricate than one would expect. And this type of interaction makes the use of random forests very adequate as even a general linear model (or mixed-effect model) could not grasp an interaction of this type.[39] I have illustrated one of the Alternative codes concerned using Chart 3, which although does not cover all the Alternative codes, illustrates the nature of the interaction and the problems involved. This also makes it easy to see the differences between areas.

In Chart 3 one can see that effect of the *Phonologia* Alternative code was quite different area-by-area. As in some provinces the hypercorrectly read gender confusions were dominated by confusions between the feminine and the neuter, while in others the most prominent confusion was the one between masculine and neuter. However, the distribution of these areas is not the same as observed in my previous study.[40] That is, the *Phonologia* Alternative code indeed affects the distribution of gender confusions. This is further evidenced by the fact that a correlation holds here between Code and Province, checked by Fisher's exact test, the *P*-value obtained was 0.0015.[41] It can therefore be claimed with high certainty that a correlation occurred at this level. In addition to this, an interaction also occurred at the *Nominalia* level, with a *P*-value of 0.00015 obtained with a Fisher's exact test.[42] In other words, the source of spatial differences under the hypercorrect interpretation is not entirely the property of the gender confusions themselves but

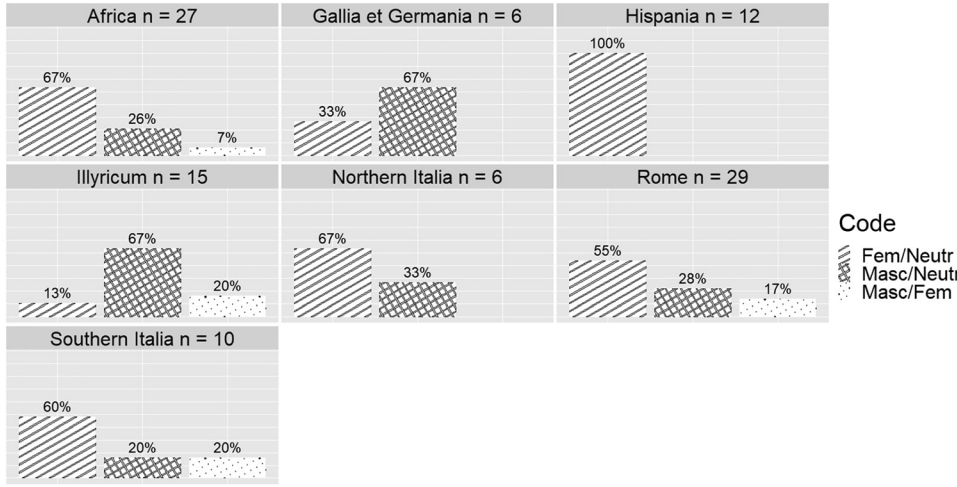


Chart 3. The spatial distribution of Gender Confusions with 'Phonologia' Alternative Code Hypercorrect reading

[39] Tagliamonte–Baayen (2012) 161.

[40] Szlovicsák (2022) 14–15.

[41] In this paper, I consider the *P*-values to be significant in all cases when they are less than 0.05.

[42] The other *P*-values obtained with the test were *Synctactica etc* = 0.211, *None* = 0.326.

is related to possible alternative interpretations and is markedly influenced by other levels of linguistic variation. In the case of the non-hypercorrect reading, the interaction between Province and Alterative Code was not observable at most levels of the Alternative code. It was only present at the *None* level,[43] i.e., the level of *purely* gender confusions. This does not, of course, exclude the possibility that other levels of linguistic change have also historically determined the development of these confusions, but in the case of those the spatial differences were probably independent of them.

THE STATISTICAL MODEL

As stated, for the data presented here, two kinds of models were used. Conditional inference trees,[44] and one random forest was fitted on the hypercorrect and the non-hypercorrect data each. The advantage of presenting both models, is that trees are more understandable, while random forests are more accurate and can provide a reliable way to compute variable importance.[45] Figs 4 and 5 show the conditional inference trees computed for both the hypercorrect and the non-hypercorrect Codes.

Both figures help illustrate the way a conditional inference tree operates. At each step (node) an algorithm finds the variable with the lowest *P*-value, i.e., the variable that has the strongest correlation with the dependent variable (here the Codes). Then the algorithm finds a point in which it separates the chosen variable into two categories in a way that maximizes the difference between the two categories. Once all *P*-values are higher than a predefined value (in our case 0.05), the algorithm stops. So, for example in Fig. 4 at Node 1, we can see that Alternative code was the chosen independent variable, as it had the highest correlation with the Main codes. Then it was separated into two categories: *Phonologia* and every other level. This results in a confident prediction. If a given gender confusion has the Alternative code *Phonologia*, then it will be of the type *F/M* > *Neutr* with high likelihood. The bar plots at the bottom nodes indicate the distribution of the Codes of our sample corresponding to the independent variable levels given by higher nodes. In the case of Node 11 we do not know anything about other independent variables, we only know that the data here has *Phonologia* as its' Alternative code. In the case of Node 4 on the other hand, we also have a restriction on the Province from which this data originates. However, from the bar plot it is also clear that the algorithm was not able to purify the distribution and the prediction of *Masc/Fem* as the Code of these items is not quite confident. The main goal to achieve here would be to have one highly likely level of the Code in each bottom node with a low level of uncertainty. But achieving this was not quite possible with conditional inference trees given our variables. Which is why I will also present the use of random forests.

[43]In the case of the non-hypercorrect reading, the other *P*-values were: *None* < 0.001, *Synctactica etc* = 0.151, *Phonologia* = 0.107, *Nominalia* = 0.154

[44]The word "conditional" here refers to the fact that these trees are made using *P*-values computed for a conditional probability on possible permutations of the data. This makes the types of trees used here more robust and less sensitive to variables with a high number of missing values. See Hothorn et al. (2006) 663–668.

[45]Strobl et al. (2009) 333, 335. Tagliamonte–Baayen (2012) 163.

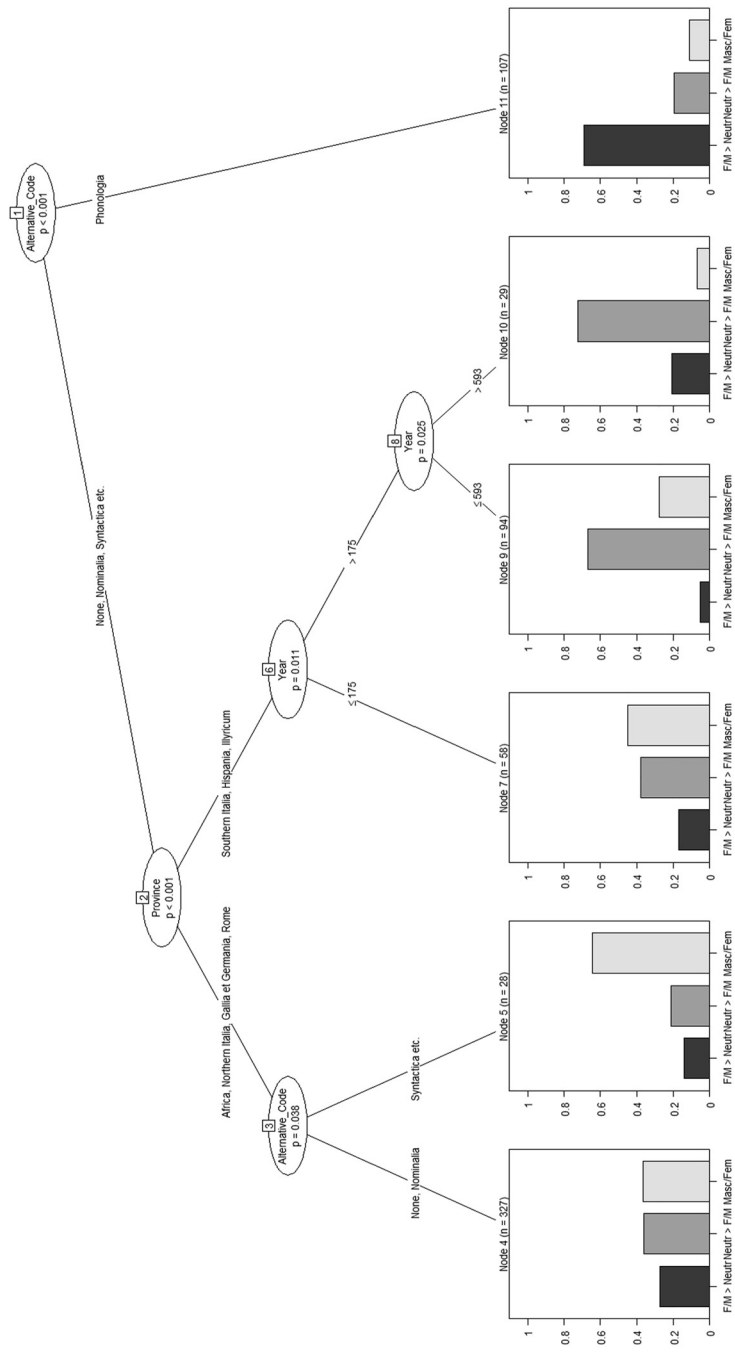


Fig. 4. Conditional inference tree based on the non-hypercorrect codes

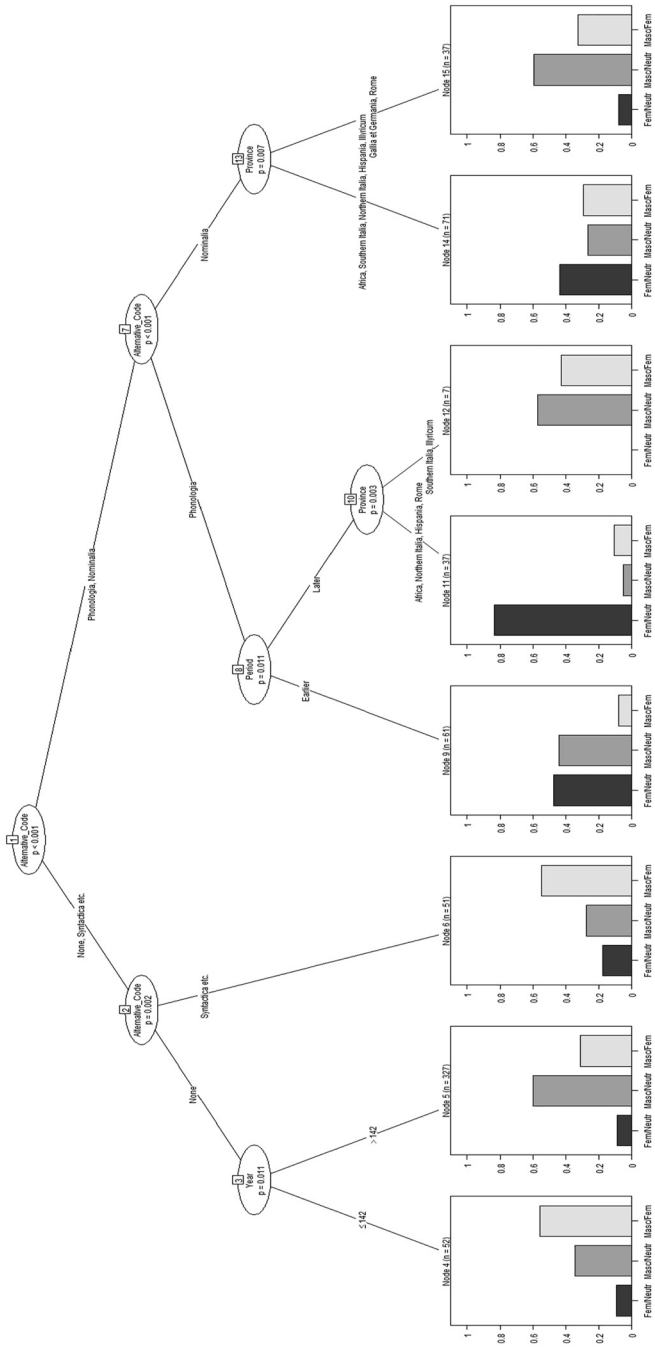


Fig. 5. Conditional inference tree based on the hypercorrect codes

Nonetheless the image of non-hypercorrect Codes we get from Fig. 4 is quite interesting. Node 11 illustrates that no matter the spatial distribution, *Phonologia* Alternative codes are in themselves powerful predictors of some Main codes. From this it becomes apparent that a universal interaction might not be adequate for this data.[46] As neither Year, nor Province, nor Period appear on the right-hand side of our tree. And, Year only interacts with some of the Provinces, but not all of them. We have seen this already between Alternative codes and Provinces (Chart 3). It is also striking that in the case of Fig. 4 the Period variable did not even appear. This might suggest that if we take Year into account the need for Period disappears. And, that the cut-off point of Period might not be the best choice for our data. As in the case of *Southern Italia, Hispania* and *Illyricum* Year is divided into three periods (CE 1–175, 176–593, 593–700).

In a similar way one can also examine the conditional inference tree fitted to the Hypercorrect data shown in Fig. 5. Here we find even more intricate interactions between the variables, as for example Province influences Alternative code on both the *Phonologia* and the *Nominalia* levels, however the grouping of Provinces differs based on the Alternative code. What this is means is that there is no uniform effect of the Provinces on the Codes. How an Alternative Code effects the Code might differ Province-by-Province. In the present case if we have data from *Rome* with *Nominalia* Alternative code, we will expect to find *Masc/Neutr* as the Main code. However, with data from the same city, but with *Phonologia* Alternative code (and from the *Later* Period) we will expect to find *Fem/Neutr* as the Main code. Contrary to this, in the case of Africa, we will expect to find *Fem/Neutr* confusions in both these cases. This suggests that we cannot describe this data with universal interactions and easily interpretable properties, even more notable. It is also worth noting that in the case of the Hypercorrect data, the Period variable became significant enough to include in the tree under consideration. Therefore, it is not immediately evident that this is a worse predictor than Year. To check this, we will need to use random forest methods. To also illustrate a weakness of Conditional Inference Trees, I have included the confusion matrices for the trees fitted on both the Non-Hypercorrect and the Hypercorrect data, seen in Tables 1 and 2.

Table 1. Confusion matrix of the tree-based predictions for the non-hypercorrect Codes

		Predicted values		
		F/M > Neutr	Neutr > F/M	Masc/Fem
True values	F/M > Neutr	74	11	103
	Neutr > F/M	21	84	146
	Masc/Fem	12	28	164

[46]Strictly speaking conditional inference trees are not very likely to find other types of interactions and are more likely to represent the data with these complex interactions. However, as we have seen already that the variables under consideration are not totally independent, the use of trees is adequate and does not misrepresent the data. See Strobl et al. (2009) 329–330.

Table 2. Confusion matrix of the tree-based predictions for the hypercorrect codes

		Predicted values		
		Fem/Neutr	Masc/Neutr	Masc/Fem
True values	Fem/Neutr	91	32	14
	Masc/Neutr	48	222	32
	Masc/Fem	30	117	57

Confusion matrices help us understand the way classification models operate, and the errors they make while trying to classify the data. In the case of Table 1, both the rows and the columns represent the Main codes. However, in the rows we can see the true values of the data, while in the columns we can see the predicted values. Therefore, the numbers in the cells of this table show how many times a given true value was predicted to be each of the values. So, in the first row 74 shows the number of times our tree was able to correctly classify *F/M > Neutr* Codes, 11 times it predicted the data with this Code to be *Neutr > F/M* and 103 times to be *Masc/Fem*. The sum of counts in the main diagonal shows the number of correct predictions, if we divide this by the number of data points (643), we get the proportion of correct predictions.

From Table 1 we can see that this model can predict the *Masc/Fem* Code quite well, however it struggles with the Main codes of interest, which are mostly misclassified. The reason behind this comes from Node 4. A great amount of data ($n = 327$) is concentrated here without any real purity. The number of data points with *Masc/Fem* Code is only marginally larger than the amount of *Neutr > F/M* Codes. On the upside however, if in a given case the model predicts one of the Main codes concerning the neuter to occur, we can quite confidently accept that prediction. Regardless it is worth noting that classification trees will always remain uncertain to a degree, and this is where random forests can prove to be even more useful.[47]

Table 2 shows the predictions for the Hypercorrect data in an analogous way. Most notable is the fact that this model can distinguish better between the Main codes. However, in this case, it was the *Masc/Fem* Codes that were mostly misclassified. This probably happened due to node 5, containing a large amount of data ($n = 327$), with a relatively high proportion of *Masc/Fem* Codes. The high rate of error in these cases is due to the high variability of inference trees,[48] which however can be used to our advantage with a random forest model.

Now that we have seen the workings of Conditional Inference Trees, we can turn to random forests. As the name suggests, a random forest is a collection of trees, which were all fit to a subsample of our data. These trees then "vote" for each data point when it comes to prediction. This way the variance of the trees, which can occur from even small modifications, helps the forest take all the variables into account and see how much they help in making correct predictions. Based on this a variable importance can be calculated which helps us decide relative usefulness of the different variables and whether some variables are unnecessary.[49] The main

[47] Strobl et al. (2009) 341.

[48] Strobl et al. (2009) 331.

[49] Tagliamonte–Baayen (2012) 162.

downside of random forests is that they cannot be as easily visualized as an inference tree. To further illustrate them, I created a confusion matrix for the two random forests each (Tables 3 and 4), which help us see the improvement in classification compared to the conditional inference trees.

From Table 3 we can see that compared to the inference tree on the same data, a random forest can predict the Main codes with much higher accuracy. Here in each row the true value is also the most predicted one, meaning that we can trust this model to recognize from which type of gender confusion a given data point is coming from. This property of random forests can also be seen in the case of the Hypercorrect data in Table 4.

These Codes however seem to be harder to deal with resulting in a high rate of misclassification in the case of the *Masc/Fem* level. Suggesting that somehow the variables used here cannot provide enough information to distinguish between the *Masc/Fem* and the *Masc/Neutr* level. Regardless this unexplained variation, we can still use the random forests to calculate the variable importance measures.

Variable importance helps one decide which variables contribute the most to the accuracy of the predictions and which ones do not. In this paper I used the conditional variable importance to understand the variables, as introduced by Strobl and her colleagues.[50] This measure is conditional in the sense that, when calculating variable importance for a given variable other variables are taken into account, and therefore this method can enable the detection of variables that are only significant conditionally and also root out variables that are unimportant given other variables.[51] Having run these it became evident that regardless of the interpretation of the

Table 3. Confusion matrix of the random forest-based predictions for the non-hypercorrect codes

		Predictions		
		F/M > Neutr	Neutr > F/M	Masc/Fem
True values	F/M > Neutr	118	29	41
	Neutr > F/M	32	174	45
	Masc/Fem	28	64	112

Table 4. Confusion matrix of the random forest-based predictions for the Hypercorrect Codes

		Predictions		
		Fem/Neutr	Masc/Neutr	Masc/Fem
True values	Fem/Neutr	79	39	19
	Masc/Neutr	16	258	28
	Masc/Fem	17	108	79

[50] Strobl et al. (2008), for the computation I used the *partykit* R package.

[51] Strobl et al. (2009) 337.

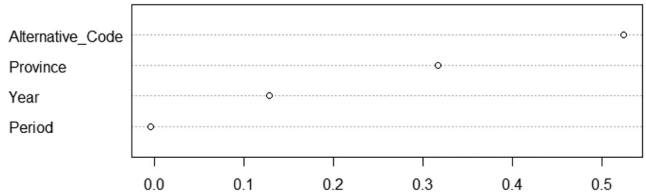


Chart 4. Conditional variable importance from random forest

Main codes, the order of the variables was the same.[52] Therefore, I only include one chart (Chart 4) to show the results. Regardless the fact that the chart below has numeric values for variable importance, we should focus only on the order of the variables, as the number here is only meaningful if it is close to or less than zero, which would deem the given variable quite unimportant.[53] From Chart 4 it is clear then that the most important variable out of these is Alternative code, followed by Province and then Year. Period is not only the least important variable among them, but the value computed for it is less than zero, therefore it is not a meaningful variable. As expected, Year is a better predictor in this case and the extra information we gain from treating it as a continuous predictor is non-negligible. This chart also suggests that not only did Alternative Codes influence the distribution of Codes, but they also had a larger influence on them than either temporal or spatial variation. Making it clear that this variable should be considered any time we try to deal with gender confusions, as other levels of language change had clearly influenced this transformation. And therefore, it becomes evident that among the models of Figs 1–3, only Fig. 3 represent our data accurately. The transformation of the grammatical gender system of Vulgar Latin was influenced greatly by other levels of language change but it has also shown spatial and temporal variation, the effect of these two variables does not disappear when considering Alternative codes. Therefore it can be stated with high certainty that the transformation of the grammatical gender system during the Vulgar Latin period was highly influenced by other levels of linguistic change, while having a specific dialectic variation which was not the pure result of these other transformations.

CONCLUSION

Using inference trees and random forests proved quite useful in our case. With their help we were able to visualize and better understand the complex intercations between our predictor variables, as their effects could differ greatly based on what other variable levels are co-occurring with them. It was also possible to show that considering Alternative codes is not a dismissible

[52]In both cases the number of trees (*ntree* = 1,000) and the number of preselected splitting variables was the same (*mtry* = 2). With these parameters the variable importance remained stable in both cases. For the importance of these numbers see Strobl et al. (2008).

[53]The reason behind this is that variable importance is computed using a non-deterministic process and therefore the exact numbers could differ given another starting point for the randomization process.

issue and needs thorough consideration as they have a large effect on gender confusions. On the other hand, by using random forests I was also able to show that treating Dating as a categorical variable is not a negligible issue as it resulted in considerable information loss, which can be avoided by treating this variable as continuous.

These results are only preliminary in the sense that we were not able provide a careful linguistic explanation for this phenomenon. Previous linguists[54] have already made it clear that these types of interactions between levels of language change are ever present in Vulgar Latin. This paper aimed to shed light on the presence of this interaction in case of gender confusions and to examine the significance of this interaction using statistics. What remains therefore is to try to explain these phenomena and provide a more thorough description of the ways Alternative codes influence gender confusions. To do this, it seems that the best way forward is a more refined division of the Alternative codes to see which levels result in which gender confusion exactly.

Nevertheless, the goal of this paper to provide a preliminary description of the effect of other linguistic changes on gender confusions was achieved. By using conditional inference trees and random forests I was able to provide a more detailed description of this data than previously possible. What remains is no small task either, but the progress made here will greatly aid future research in this area.

ACKNOWLEDGEMENTS/FUNDING INFORMATION

Supported by the ÚNKP-23-3-I-ELTE-732 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. The present paper was prepared within the framework of the HORIZON-ERC-2022-ADG project no. 101098102 entitled Digital Latin Dialectology (DiLaDi): Tracing Linguistic Variation in the Light of Ancient and Early Medieval Sources and of the NKFIH (National Research, Development and Innovation Office) project no. K 135359 entitled Computerized Historical Linguistic Database of Latin Inscriptions of the Imperial Age (see: http://lldb.elte.hu/). I am most grateful to Béla Adamik for all his invaluable help and to Alessandro Papini for suggesting the use of Inference Trees.

BIBLIOGRAPHY

LLDB Database: *Computerized Historical Linguistic Database of the Latin Inscriptions of the Imperial Age*. Available at: http://lldb.elte.hu/ (Accessed 15 February 2024).

[54]E.g., Väänänen (1981) 103; Herman (2000) 65; Löfstedt (1961) 226–227.

Adamik, B. (2019). On the Loss of Final -m: Phonological or Morphosyntactic Change? *Acta Antiqua Academiae Scientiarum Hungaricae*, 59: 97–108.

Agresti, A. (2013). *Categorical Data Analysis*. 3rd ed. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken.

Cinelli, C. – Forney, A. – Pearl, J. (2020). A Crash Course in Good and Bad Controls. Available at: https://doi.org/10.2139/ssrn.3689437 (Accessed 15 February 2024).

Herman, J. (2000). *Vulgar Latin.* University Park, Pennsylvania.

Hothorn, T. – Hornik, K. – Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3): 651–674.

Levshina, N. (2015). *How to do Linguistics with R. Data exploration and statistical analysis.* John Benjamins Publishing Company, Amsterdam–Philadelphia.

Loporcaro, M. (2018). *Gender from Latin to Romance. History, Geography, Typology*. Oxford Studies in Diachronic and Historical Linguistics 27. Oxford University Press, Oxford.

Löfstedt, B. (1961). *Studien über die Sprache der langobardischen Gesetze. Beitrag zur frühmittelalterlichen Latinität*. Almqvist & Wiksell, Stockholm.

Paulus, N. (2020). A study on the weakening of the word final -s compared to -m in the epigraphic corpus. *Acta Classica Universitatis Scientiarum Debreceniensis*, 56: 125–143.

Papini, A. (2022). *Ipsa Latinitas et regionibus cotidie mutetur et tempore:* Some methodological considerations on the use of Herman's quantitative method. *Listy filologické*, 145(3–4): 343–378.

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/. (Accessed 15 February 2024).

Strobl, C. – Boulesteix, A.L. – Kneib, T. – Augustin, T. – Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(307). Available at: https://doi.org/10.1186/1471-2105-9-307 (Accessed 15 February 2024).

Strobl, C. – Malley, J. – Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods*, 14(4): 323–348.

Szlovicsák, B. (2022). Preliminary Examination of the Latin Neuter on Inscriptions. *Acta Antiqua Academiae Scientiarum Hungaricae*, 62(4): 419–434.

Tagliamonte, S.A. – Baayen, R.H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2): 135–178.

Väänänen, V. (1981). *Introduction au latin vulgaire*. 3rd ed. Klincksieck, Paris.

Van de Velde, F. – Petré, P. (2020). Historical Linguistics. In: Adolphs, S. – Knight, D. (eds.), *The Routledge Handbook of English Language and Digital Humanities*. Routledge, London, pp. 328–352.