

# Variation and change in Latin BE-periphrases: empirical and methodological considerations<sup>1</sup>

LIEVEN DANCKAERT  
*Ghent University*

## 1 Introduction: documenting the course of syntactic change

### 1.1 Two types of BE-periphrases in Latin

The empirical focus of this paper is the diachronic development of Latin periphrastic verb forms involving a BE-auxiliary and a past participle. As is well known, the transition from Latin to the Romance languages witnesses a shift from periphrases involving an auxiliary built on the *infectum* stem (a-sentences), towards patterns where participles are coupled with perfective auxiliaries (b-sentences). Both can occur with passive (1) and with deponent (2) past participles:

- (1) a. si forte aliquid laesum erit  
‘if by chance something shall have been damaged.’ (= Cass. Fel. 29 Rose (29.18 Fraisse))  
b. postquam eruptio fuert facta  
‘after an outburst will have taken place.’ (= Cass. Fel. 21 Rose (21.9 Fraisse))
- (2) a. quo usus est Galenus ad uniuersas tusses et dyspnias  
‘which Galenus used in all cases of coughing and shortness of breath.’ (= Cass. Fel. 41 Rose (41.7 Fraisse))  
b. et sanguinem detrahes quantum causae sufficere fuert arbitratus  
‘and draw as much blood as you will think suffices for this case.’ (= Cass. Fel. 21 Rose (21.4 Fraisse))

Throughout this paper, I will refer to the old forms as E-periphrases (or simply E), and to the new ones as F-periphrases. The basic patterns are summarized in Tables 1 and 2. In the *perfectum*, the forms in the middle column are gradually replaced by the forms in the rightmost column:

	Old <i>perfectum</i> : E/S-stem	New <i>perfectum</i> : F-stem
<b>Infinitive (perfect)</b>	<i>amatus esse</i>	<i>amatus fuisse</i>
<b>Indicative (perfect)</b>	<i>amatus sum</i>	<i>amatus fui</i>
<b>Indicative (pluperfect)</b>	<i>amatus eram</i>	<i>amatus fueram</i>
<b>Indicative (future perfect)</b>	<i>amatus ero</i>	<i>amatus fuero</i>
<b>Subjunctive (perfect)</b>	<i>amatus sim</i>	<i>amatus fuerim</i>
<b>Subjunctive (pluperfect)</b>	<i>amatus essem</i>	<i>amatus fuisset</i>

Table 1: E- and F-periphrases in passives (*perfectum*).

Later, a similar shift took place in the *infectum*, where the old synthetic forms (E) are replaced by analytic expressions (F):

---

<sup>1</sup> The research reported on in this paper was funded by the ‘Fonds voor Wetenschappelijk Onderzoek’, postdoctoral grant FWO13/PDO/024.

	Old synthetic <i>infectum</i> : E	New analytic <i>infectum</i> : F
<b>Infinitive (present)</b>	<i>amari</i>	<i>amatus esse</i>
<b>Indicative (present)</b>	<i>amor</i>	<i>amatus sum</i>
<b>Indicative (imperfect)</b>	<i>amabar</i>	<i>amatus eram</i>
<b>Indicative (future)</b>	<i>amabor</i>	<i>amatus ero</i>
<b>Subjunctive (present)</b>	<i>amer</i>	<i>amatus sim</i>
<b>Subjunctive (imperfect)</b>	<i>amarer</i>	<i>amatus essem</i>

Table 2: E- and F-periphrases in passives (*infectum*).

Aspects of these two developments were discussed earlier by among others Leumann (1921), Muller (1924), Herman (2002), Clackson and Horrocks (2007) and de Melo (2012b). See also Burton (this volume).

## 1.2 Aims of this paper

A first descriptive goal of this paper is to document the diachronic trajectory of above-mentioned evolutions by means of corpus data. More particularly, I will present a number of novel empirical observations that shed some light on (i) the origins of the F-paradigm (section 3), as well as (ii) on the way this pattern spread from the *perfectum* towards the *infectum* (sections 4 and 5). Concerning the first point, I will show that the development of the F-periphrases is closely related to the Latin tense system (section 3). My data reveal that in the future perfect the spread of the F-pattern is almost entirely completed by the end of the sixth century (see especially section 3.2.2). Importantly, the fact that in the future perfect all stages of the gradual evolution from E to F are documented in the corpus can be construed as an argument against the idea that the development from Early to Late Latin is essentially discontinuous, in the sense that features characteristic of Late Latin are present in very early authors (typically Plautus) but absent from the classical language. Secondly, as to the later development of the F-paradigm, I will suggest that the Latin passive perfects of the type *amatus est* are not the predecessors of Romance present tense analytic passives like Italian *sono amato* and French *je suis aimé* (section 4). Evidence for this claim comes from word order facts. In the fifth section of the paper, I will offer some additional discussion of word order in BE-periphrases, which corroborates the conclusion arrived at in section 4.

A second goal is to make a case for the use of corpus data — coupled with the appropriate statistical methods — when studying phenomena of the grammar of Latin which are subject to (synchronic or diachronic) variation. As this is not standard practice in the field of Latin linguistics, I will first briefly elaborate on why one might want to adopt such a methodology, and how this can be done.

## 2 Latin in diachrony: a ‘variation and change’ perspective

### 2.1 The study of linguistic variation

At a methodological level, the present paper is to be situated in the ‘variation and change’ approach to grammar, as initiated by Labov (1966) and Weinreich, Herzog and Labov (1968) (see also Kroch (1989) and Pintzuk (2003) for further background on variationist approaches to language change). An important property of this paradigm is its strong emphasis on the use of empirical data, which very often are quantitative in nature and therefore require appropriate statistical treatment before they can be interpreted. Most work in this tradition has focused on sociolinguistic correlates of synchronic and diachronic variation. I hasten to add that taking

into account sociolinguistic factors goes well beyond the essentially descriptive aim of this paper. However, what I do want to show is how a corpus study can be useful in documenting patterns of variation and change in Latin. For additional discussion on this topic (with particular reference to word order in BE-periphrases), the reader is referred to Brookes (2014).

A crucial observation is that the distribution of two or more competing linguistic forms (whether allophones, lexical items, word order patterns or whatever) is typically not correlated with a single discriminating factor, but rather the result of the complex interplay of a number of variables. Consider for instance the following case study which is to some extent comparable to the Latin data which are the focus of this paper. In Dutch, past participles in compound tenses with BE (not illustrated here) and HAVE can either precede (3a) or follow (3b) the auxiliary:

- (3) a. dat Jan geslapen heeft  
       that Jan slept has  
       b. dat Jan heeft geslapen  
           that Jan has slept  
           ‘that Jan has slept’

In the literature, it has been shown that this alternation is subject to both diachronic (Coussé 2008) and synchronic (de Sutter 2005) variation. Crucially, the distribution of these two minimally different variants is governed by multiple factors, many of which can — in a broad sense — be called ‘usage-based’: they include factors such as register, information structure and prosody. On the basis of a corpus study where dialectal and register variation were controlled for by only considering Southern Dutch journalistic prose, de Sutter (2009: 229) lists nine factors that can be shown to influence the choice between the order ‘past participle - auxiliary’ and the order ‘auxiliary - past participle’ in Dutch embedded clauses:

1.	Distance between previous clause accent and participial accent (# of accented syllables)
2.	Distance between following clause accent and participial accent (# of accented syllables)
3.	Morphological structure of participle [ $\pm$ separable]
4.	Presence vs. absence of extraposed constituent
5.	Length of the ‘middle field’
6.	Definiteness of the last preverbal constituent
7.	Inherence of the last preverbal constituent
8.	Type of finite verb [ $\pm$ copular]
9.	Syntactic persistence [ $\pm$ previous VPAux]

Table 3: Factors influencing the alternation between the orders ‘past participle - HAVE’ and ‘HAVE- past participle’ in Southern Dutch journalistic prose.

In order to arrive at detailed results such as the ones summarized in Table 3, a number of methodological guidelines have to be respected. First, whenever one builds an argumentation based on quantitative (corpus) data, one has to work with a corpus which is sufficiently large. Drawing conclusions on the basis of small samples (one author for a synchronic stage, a handful of texts for the whole of Latinity), will almost certainly yield inaccurate or even wrong results, as one cannot conclude with any confidence which factor(s) (time, genre, any syntactic factor, lexical frequency, etc) a given observed effect (difference in ‘raw’ counts) is correlated with. Second, in order to assess the relative weight of a number of significant factors (say the extent to which they (dis)favour one word order pattern over the other), one

has to be equipped with a set of tools which are sufficiently powerful to distinguish meaningful patterns from accidental quirks in the data. The statistical tool that I will use in this paper is the one which is very often used in sociolinguistics, namely logistic regression (the nature of which will be explained in due course). Here too the ‘small sample’ issue arises: running this kind of analysis on small data sets is simply impossible. In any event, whenever one makes use of quantitative arguments, one should never ever draw conclusions on the basis of raw frequencies (absolute counts or percentages) only. I take this last point to be entirely uncontroversial.

Note finally that the above discussion is not to be interpreted as a plea for reducing historical (socio)linguistics to number crunching. However, what I do claim is that when one does not adopt a methodology similar to the one defended here, one risks ending up explaining non-existing data.

## 2.2 A Latin prose corpus

A full description of the Latin text corpus that I will be using in this paper is provided in Table 4:

	<b>Author (work(s))/text</b>	<b>Period</b>	<b># words</b>	<b>Source</b>
1.	Plautus (comedies)	ca. 210 BC	165607	Brepolis + lexicon*
2.	Terence (comedies)	ca. 165 BC	49939	Brepolis + lexicon*
3.	Cato ( <i>De Agricultura</i> )	ca. 160 BC	16026	Hyperbase
4.	Cicero (selection of speeches)	ca. 60 BC	471530	Hyperbase
5.	Caesar ( <i>De bello ciuili</i> , <i>De bello Gallico</i> 1-7)	ca. 50 BC	79058	Hyperbase
6.	Varro ( <i>Res rustica</i> ; <i>De lingua Latina</i> )	45 BC	75619	Brepolis
7.	Sallust ( <i>Bellum Iugurthinum</i> ; <i>De coniuratione Catilinae</i> )	ca. 40 BC	32360	Hyperbase
8.	Hyginus ( <i>De astronomia</i> )	ca. 20 BC	22288	Brepolis
9.	Vitruvius ( <i>De architectura</i> )	0 AD	58630	Brepolis
10.	Livy ( <i>Ab urbe condita</i> ; for passive BE-periphrases only books 1-2, 21-22, 31-32 and 41)	ca. 5 AD	514370 (105794)	Brepolis
11.	Celsus ( <i>De medicina</i> )	ca. 30 AD	104017	Brepolis
12.	Seneca ( <i>Epistulae ad Lucilium</i> , <i>Consolationes</i> , <i>Dialogi</i> , <i>Apocolocyntosis</i> )	ca. 50 AD	247708	Hyperbase
13.	Columella ( <i>De Agricultura</i> )	ca. 55 AD	109177	Brepolis
14.	Petronius ( <i>Satyrice</i> )	ca. 60 AD	31093	Hyperbase
15.	Frontinus ( <i>Strategemata</i> , <i>De aquaeductu urbis Romae</i> )	ca. 90 AD	30391	Brepolis
16.	Quintilian ( <i>Institutio Oratoria</i> )	95 AD	174237	Brepolis
17.	Pliny the Younger ( <i>Epistulae</i> , <i>Panegyricus</i> )	95-110 AD	85073	Brep./Hyp.
18.	Tacitus ( <i>Germania</i> , <i>Dialogus de oratoribus</i> , <i>Agricola</i> , <i>Historiae</i> , <i>Annales</i> )	ca. 110 AD	165345	Hyperbase
19.	Suetonius ( <i>Vitae Caesarum</i> )	119-120 AD	72000	Brepolis
20.	Gaius ( <i>Institutiones</i> )	ca. 170 AD	43676	Brepolis

21.	Tertullian ( <i>Aduersus Marcionem</i> 1-5)	ca. 210 AD	83601	Brepolis
22.	Cyprian ( <i>Epistulae</i> 1-76, except 4, 57, 61, 64, 67, 70 et 72)	ca. 255 AD	54151	Brepolis
23.	<i>Scriptores Historiae Augustae</i>	ca. 320 AD?	109199	Brepolis
24.	Palladius ( <i>De agricultura</i> , including the <i>De ueterinaria medicina liber</i> )	ca. 350	50119	Brepolis
25.	<i>Itinerarium Egeriae</i>	385 AD	17552	Brepolis
26.	Hieronymus ( <i>Epistulae</i> 1-30, 60-80)	ca. 390 AD	81391	Brepolis
27.	Augustine (Selection of 80 <i>Sermones</i> : 1-10 (without 4A), 101-113A, 201-223E, 301-313A (Verbraken 1976))	ca. 400	127118	Brepolis
28.	Vulgate (Old Testament: <i>Genesis</i> , <i>Samuel</i> 1-2, <i>Isaias</i> ; New Testament: <i>Marcus</i> , <i>Matthaeus</i> , <i>Lucas</i> , <i>Iohannes</i> , <i>Actus</i> , <i>Apocalypsis</i> )	ca. 405 AD	83580 (OT); 83844 (NT)	Brepolis
29.	<i>Gesta Conlationis Carthaginiensis</i>	411 AD	55002	Brepolis
30.	Vegetius ( <i>Epitoma rei militaris</i> , <i>Mulomedicina</i> )	ca. 420 AD	73428	Brepolis
31.	Cassius Felix ( <i>De medicina</i> )	447 AD	29673	Brepolis
32.	Victor Vitensis ( <i>Historia persecutionis Africanae prouinciae</i> )	ca. 490 AD	19777	Brepolis
33.	Pompeius Maurus ( <i>Commentum Artis Donati</i> )	ca. 500 AD	79364	Brepolis
34.	Caesarius of Arles ( <i>Sermones</i> 1-80)	ca. 520 AD	91753	Brepolis
35.	Anthimus ( <i>De obseruatione ciborum</i> )	ca. 535 AD	4479	Brepolis
36.	Iordanes ( <i>Getica</i> , <i>Romana</i> )	ca. 550 AD	38039	Brepolis
37.	Gregory of Tours ( <i>Historiae</i> )	ca. 590 AD	122003	Brepolis

Table 4: Description of the corpus used (ca. 210 BC - 590 AD). \*: for Plautus and Terence the relevant lexica (Lodge 1904 and McGlynn 1963-67 respectively) were also consulted.

As indicated in the rightmost column, the texts that I have used were either drawn from the online database available at [www.brepolis.net](http://www.brepolis.net), or from the morphosyntactically annotated corpus available on the CD-ROM 'Hyperbase - Latin'. For reasons of space, I cannot here provide a full list of the text editions used, but I refer to the information available on the Brepolis website as well as to Brunet and Mellet (n.d.) for full details. Similar considerations prevent me from justifying why I included the texts I did, and why I did not take into account others.

### 2.3 Further methodological remarks

As is well known, Latin BE-periphrases are often ambiguous between a genuine verbal reading and an adjectival one (see for instance de Melo 2012b: 88-9).<sup>2</sup> In the present context, only truly verbal passives are of interest. Unfortunately, systematically distinguishing between these two broad classes of passives in a corpus study is a very difficult task. To control for this problem, I have adopted the following methodology (which is essentially the same as in the Hyperbase corpus, as well as in specialized *indices uerborum*, where a distinction is made between *esse* as a copula (*ui copulatiua*) and as an auxiliary (*ui auxiliari*),

<sup>2</sup> Distinguishing verbal from (various types of adjectival) passives is actually quite difficult: see Gehrke (2013, to appear) for relevant discussion and references to further literature.

such as Lodge (1904) and McGlynn (1963-67)). First of all, I excluded all clear-cut adjectives, such as comparatives and superlatives of past participles, all past participles appearing with the negative prefix *iN-*, as well as *-to* adjectives which do not have any verbal counterpart (like *barbatus* ‘bearded’). In addition, I left out a number of cases where there was no strictly objective criterion to classify a given expression as an adjectival passive, but where the interpretation left little room for doubt (for instance, the expression *syllaba producta est* clearly means ‘the syllable is long’, not ‘the syllable was lengthened’).

In any event, as a result of this conservative way of counting, a fair number of adjectival passives have found their way into the sample, which inevitably gives rise to some noise in the data. Crucially however, adjectival passives appear in both the E- and the F-pattern, so this potentially confounding factor can be expected to add ‘noise’ to both groups. In addition, there is no *a priori* reason to assume that the availability of adjectival passives is in any sense subject to diachronic variation. In sum, we can hope that this problem is — at least to some extent — orthogonal to the main point at issue, viz. the transition from E to F.

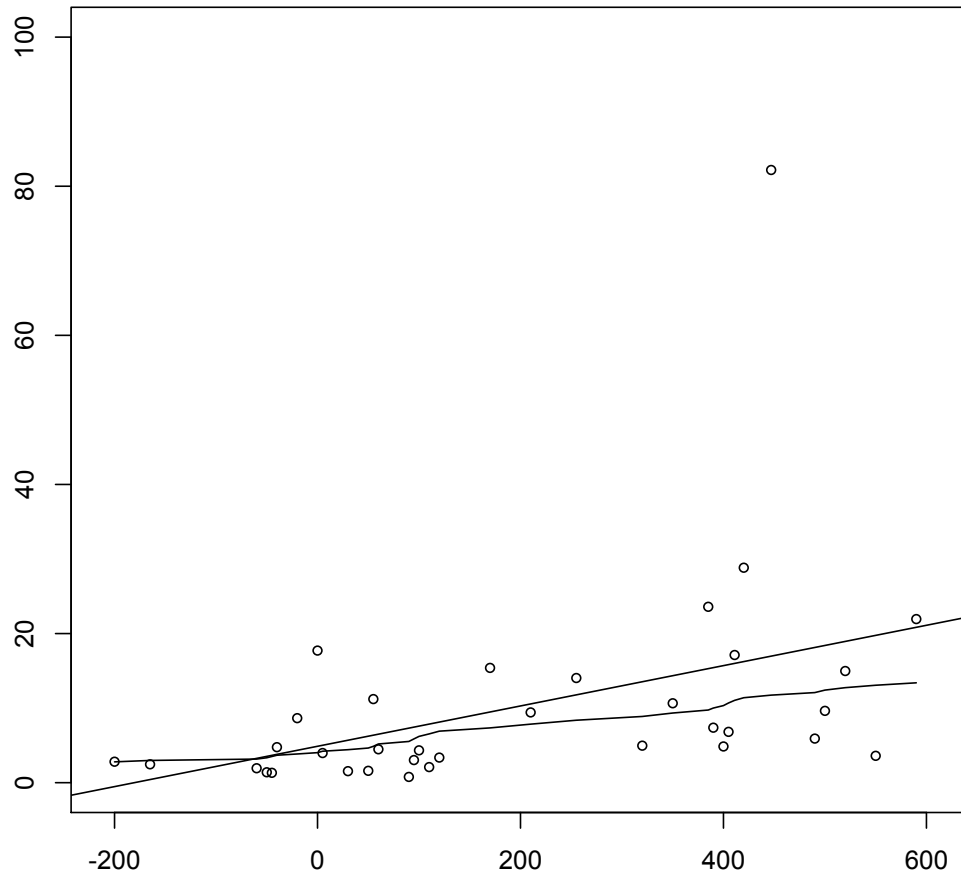
Finally, to avoid inaccurate average values estimated on the basis of small samples (cf. the word count per author/text provided in Table 4), I only took into account the data for a given author or text if that text contained at least 10 passive and 10 deponent BE-periphrases (a decision which excludes Cato and Anthimus from the corpus).

All data analysis was carried out in the statistical environment of R 3.1.0.

### **3 The origins of the F-paradigm: a description**

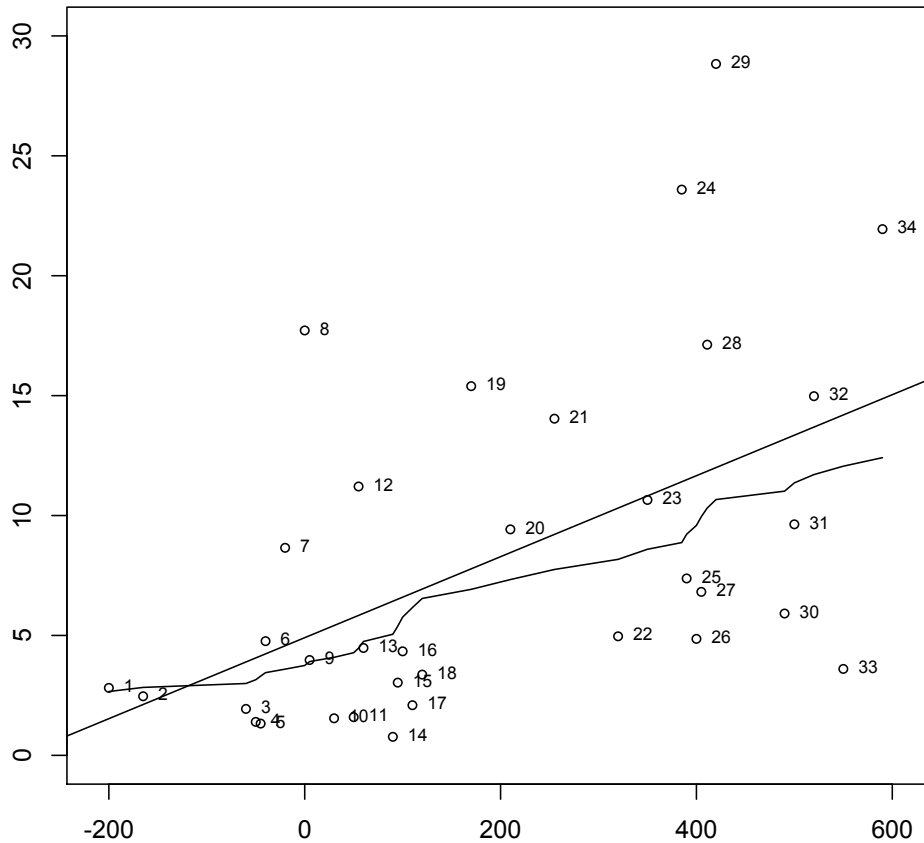
#### *3.1 The descriptives*

Although F-periphrases are well represented from the earliest text in the corpus onwards, there seems to be a real — albeit mild — increase in frequency as time goes by. The overall diachronic development of the F-periphrases is diagrammed in Graph 1, where each data point represents the relative frequency of the F-pattern (compared to the E-pattern) for a given author or text. In addition, I have plotted two so-called regression lines: a straight one, which tries to optimally characterize the overall trend in the data by means of a single straight line, and a ‘smoothed’ one, which is sensitive to more local trends. We immediately see that both of these lines go up, but that there is one data point which is clearly different from all others in exhibiting an extremely high proportional frequency of the F-pattern:



Graph 1: The proportional frequency of F-periphrases in diachrony.

Note first of all that the one text in which F-periphrases outnumber E-periphrases is Cassius Felix' *De medicina*: I will postpone a discussion of why this author behaves so differently from any other Latin author, late or otherwise, until the next section. For now, suffice it to say that even without the 'leverage' induced by this highly influential data point, we can still observe a rise in frequency of the F-pattern in our corpus, at least in absolute terms. Graph 2, where Cassius Felix is not taken into account, shows that without this one outlying observation, we still obtain a regression line with a positive slope. Additional details as to the identity of the individual data points are provided by means of case labels:



Graph 2: The proportional frequency of F-periphrases, without Cassius Felix. Case labels: 1= Plautus, 2= Terence, 3= Cicero, 4= Caesar, 5= Varro, 6= Sallust, 7= Hyginus, 8= Vitruvius, 9= Livy, 10= Celsus, 11= Seneca, 12= Columella, 13= Petronius, 14= Frontinus, 15= Quintilian, 16= Pliny, 17= Tacitus, 18= Suetonius, 19= Gaius, 20= Tertullian, 21= Cyprian, 22= *Historia Augusta*, 23= Palladius, 24= *Itinerarium Egeriae*, 25= Jerome, 26= Augustine, 27= Vulgate, 28= *Gesta Conlationis Carthaginiensis*, 29= Vegetius, 30= Victor Vitensis, 31= Pompeius Maurus, 32= Caesarius of Arles, 33= Iordanes, 34= Gregory of Tours.

In the light of what we have said earlier about multivariate data distributions, we should at this point try to find out with which factors the shift from E- to F-periphrases is correlated (on top of the factor ‘Time’, that is). In the literature on the topic, a number of candidates have been mentioned, most notably the tense and voice (passive vs. deponent) of the periphrastic expression involved, as well as the type of clause (main vs. embedded) in which the pattern occurs (see Leumann 1921 and de Melo 2012b). In the following section, I will briefly demonstrate how one can empirically assess whether or not a given factor is in any meaningful way correlated with the distribution of the two types of BE-periphrases.

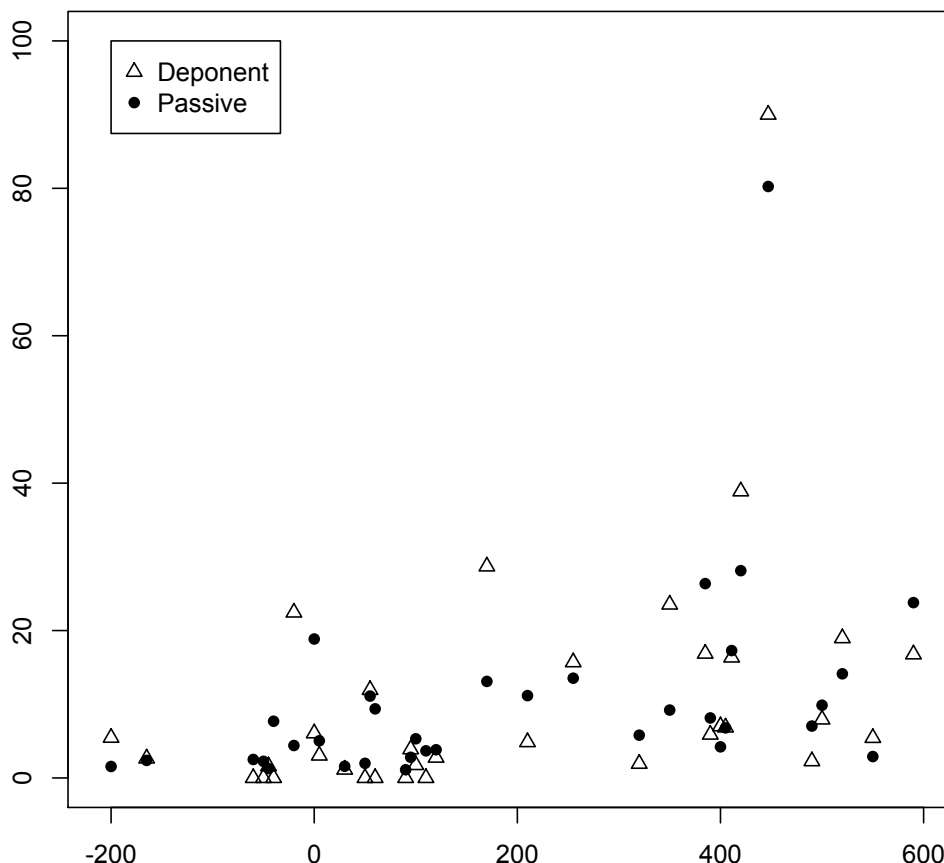
### 3.2 Factors influencing the choice of F over E

In order to evaluate to what extent a (level of a) given covariate favours the use of one competing variant over the other, one has to split up the entire data set according to the levels of the independent variable under consideration. In what follows, I will proceed in two stages: I will first look at the ‘raw’ data, which I will visualize by means of a graph (tables with counts and/or percentages being rather difficult to parse). Next, for reasons elaborated on above, I will check whether an observed difference is statistically significant or not. For this, I will build a so-called logistic regression, which models the influence of one or more independent variables (in this case ‘Voice’, ‘Tense’, ‘Clause type’, ‘Time’ etc.) on a binary outcome variable (which for us is the factor ‘E vs. F’). I will start by looking at the factor ‘Voice’.



### 3.2.1 The distinction between deponents and passives

It was observed at least as early as Brix (1901: 36-7) that in Plautus, F-periphrases occur more frequently with deponent verbs than with passives. More recently, in de Melo (2012b) the distinction between deponents and passives was taken to be of diachronic importance, in the sense that the facts from Plautus are considered to be indicative of a more general tendency for Early Latin F-periphrases to be deponent rather than passive. However, it is doubtful whether this last view is correct. If we look at the role of this factor in the entire corpus, it turns out that despite some (fully random) variation, passives and deponents behave more or less the same, at all stages of the language:



Graph 3: Relative frequency of deponent and passive F-periphrases (in percentages) over time.

As can be observed, the triangle representing Plautus' deponent F-periphrases is indeed situated higher than the dot providing data on Plautine F-passives. Proportionally, as the two data points are located at the very low end of the scale, the difference between the two average values is certainly striking.<sup>3</sup> However, I do not think that we are at this point in any position to conclude that the observed difference (7 passive vs. 19 deponent F-periphrases in de Melo's sample, 16 vs. 25 according to my apparently more conservative counting) is an effect of time, and thus that these data shed special light on the origins of the F-periphrases. First and foremost, it is simply not warranted to extrapolate findings from a single author (Plautus) to an entire language system (Early Latin), as there is no reason whatsoever to

<sup>3</sup> To be precise, 16 out of 983 Plautine passives are of the F-type (1.63%), compared to 25 out of 457 deponents (5.47%). This difference is highly statistically significant (Fisher's exact test (two-tailed),  $p = .0001$ ).

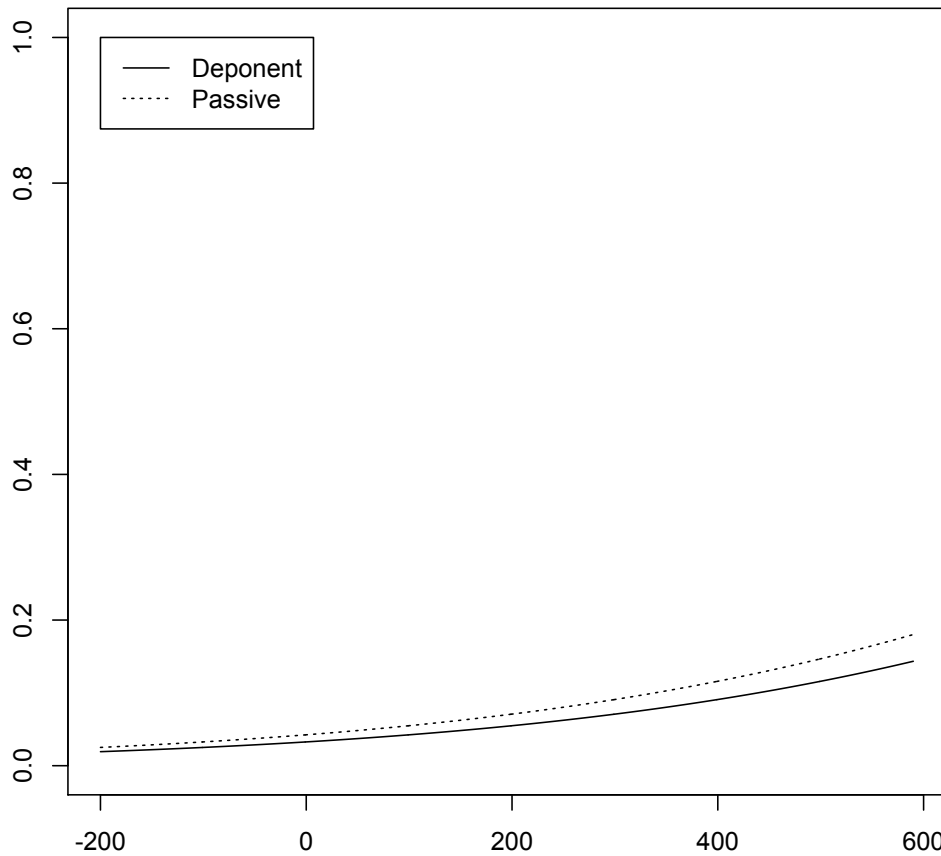
assume that tendencies observed in Plautus are representative of the whole of Early Latin.<sup>4</sup> Second, given what we have said earlier about multivariate data distributions, it might be worthwhile to consider whether any other factors are correlated with the high incidence of deponent F-periphrases in Plautus: for instance, the observed difference might well be an effect of the frequency of individual lexical items (such as *perfectum* forms of *obliviscor* ‘forget’, which occur 6 times in the F-pattern), or of any other synchronic variable (completely) unrelated with the factor ‘Time’.

As Jim Adams (p.c.) points out to me, the data from Plautus might be related to the fact that in Early Latin (semi-)deponent past participles often appear to have a stative meaning (e.g. *confisus* ‘trusting’ rather than ‘having trusted’, or *uectus* ‘travelling’ rather than ‘having travelled’), a reading which does not seem available in later periods (for critical discussion of stative (‘present’) readings of deponent past participles in Cicero, see Laughton 1964: 13). Such a ‘stative deponent periphrasis’ (my terminology) would yield a structure more akin to an adjectival than to a genuinely verbal passive, and as such, would constitute a likely environment for F-periphrases to originate, since changing the tense of the auxiliary would be the only way to express temporal anteriority. However, in order for this set of facts to lend support to the view that F-periphrases originated with deponents and not with passives, it would have to be demonstrated that in Early Latin there is a stage where stative (adjectival) participles — which are arguably the historical source of all genuinely verbal past participles in Latin, deponent and passive alike — were more frequently used in deponent than in passive environments. As far as I can tell, no evidence to this effect is at present available: there only seem to be indications that stative readings of (some, cf. Laughton 1964: 13) deponent *-to* adjectives were lost over time. As mentioned earlier, the availability (and incidence) of adjectival passives might very well be diachronically stable.

When we then look at the plotted coefficients of a regression model predicting the E/F alternation, with the factors ‘Voice’, ‘Time’ and an interaction term between these two as independent variables, it is confirmed that the role of the voice distinction is very close to that of a complete null effect (the passives scoring a bit higher than the deponents, even in Early Latin). In Graph 4, the values on the Y-axis are to be interpreted as probabilities on a scale from 0 to 1, where 0.5 is chance level (more technically, they are converted log odds: see section 3.2.3 on this last notion):

---

<sup>4</sup> The scarcity of additional comparative Early Latin material is of course very unfortunate (for instance, all 66 BE-periphrases from Cato that I extracted from the LASLA database are of the E-type), but there is obviously nothing one can do to remedy this state of affairs.



Graph 4: The effect of Voice (deponent vs. passive) on the likelihood of a BE-periphrasis to be of the F-type (probabilities over time).

As we will see in section 3.2.3, this slight difference between two types of voice does not turn out to be significant in a richer regression model.

To sum up, my own corpus findings suggest that the distinction between deponents and passives is not in any meaningful way related to the diachronic development of the F-paradigm. Moreover, given what we know now, an agnostic stance as to why Plautus produced more deponent than passive F-periphrases is perhaps the only defensible one.

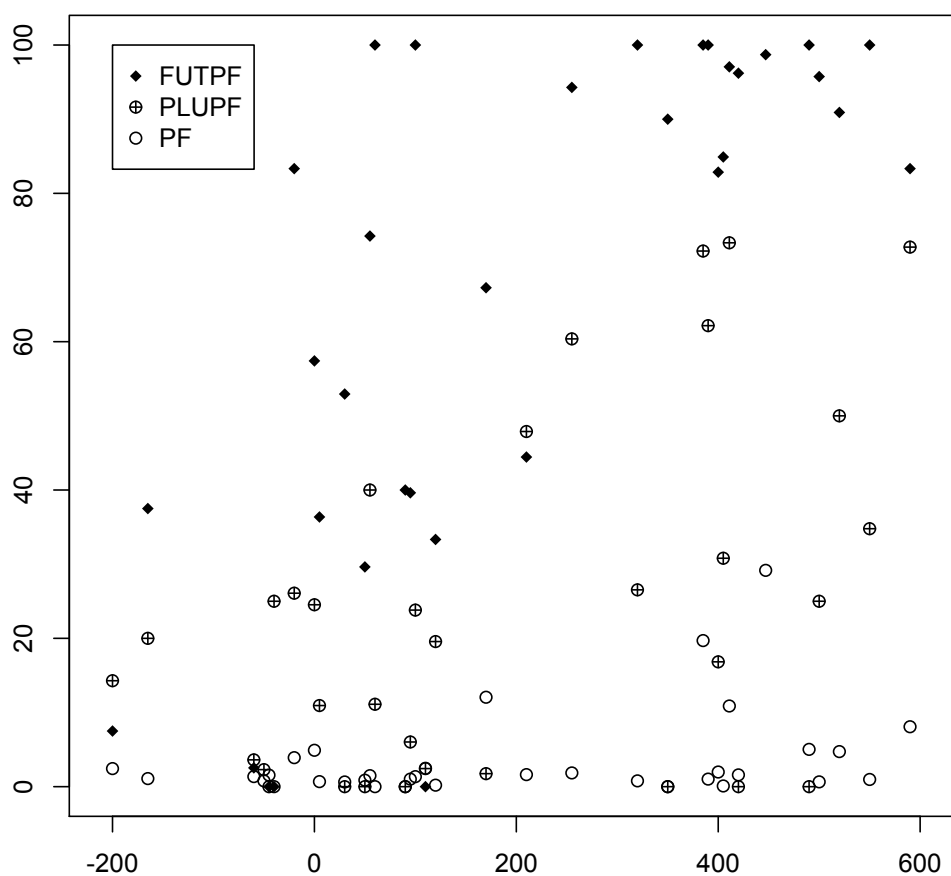
### 3.2.2 The role of the Tense system

I now turn to a second, and much more interesting predictor, namely Tense. The role of this factor has occasionally been pointed out (see de Melo 2012b and references cited there), but no full description of the relevant facts has ever been offered.

Before we look at the data, a short methodological parenthesis is in order. Undoubtedly the most problematic issue when trying to characterize BE-periphrases as ‘perfect’, ‘pluperfect’ or ‘future perfect’ is distinguishing future perfects from perfect subjunctives in the F-paradigm, which can only be told apart on morphological grounds in the first person singular. However, I do think that systematically annotating every single token is to be preferred over lumping the two categories, which (at least for Late Latin) is the option taken in de Melo (2012b). First of all, many cases are fairly unproblematic: perfect subjunctives appear in indirect questions and adverbial clauses referring to past events, and future perfects in clauses embedded under plain future tenses and imperatives. Problems of disambiguation are mainly restricted to certain *si* and *cum*-clauses with unclear (or — perhaps more accurately — underspecified) time reference, of the type ‘when(ever) X, one should Y’, which one often finds in technical prose. When judging such cases, I took into account

whether or not a given author (frequently) uses subjunctives in comparable non-perfective environments. Second, it is unlikely that the two categories were indeed conflated at the level of the (Late Latin) language system, as such a diachronic development would involve the merger of non-adjacent cells in an inflectional paradigm: elsewhere in the system, subjunctives are still distinct from indicatives, and at least in the first person singular, future perfects remain distinct from perfects (and it is obviously not the case that all Late Latin perfects can be interpreted as future perfects). Instead, I take it that the morphological quasi-identity of the two categories is accidental, and that treating them as distinct is the only linguistically defensible option. In actual practice, this inevitably gives rise to some unwelcome noise in the annotated data. However, as we will see shortly, there is every reason to assume that tense distinctions are crucially related to the rise of the F-pattern, which is why I think this is a price one should be willing to pay.

This being said, consider the data in Graph 5:



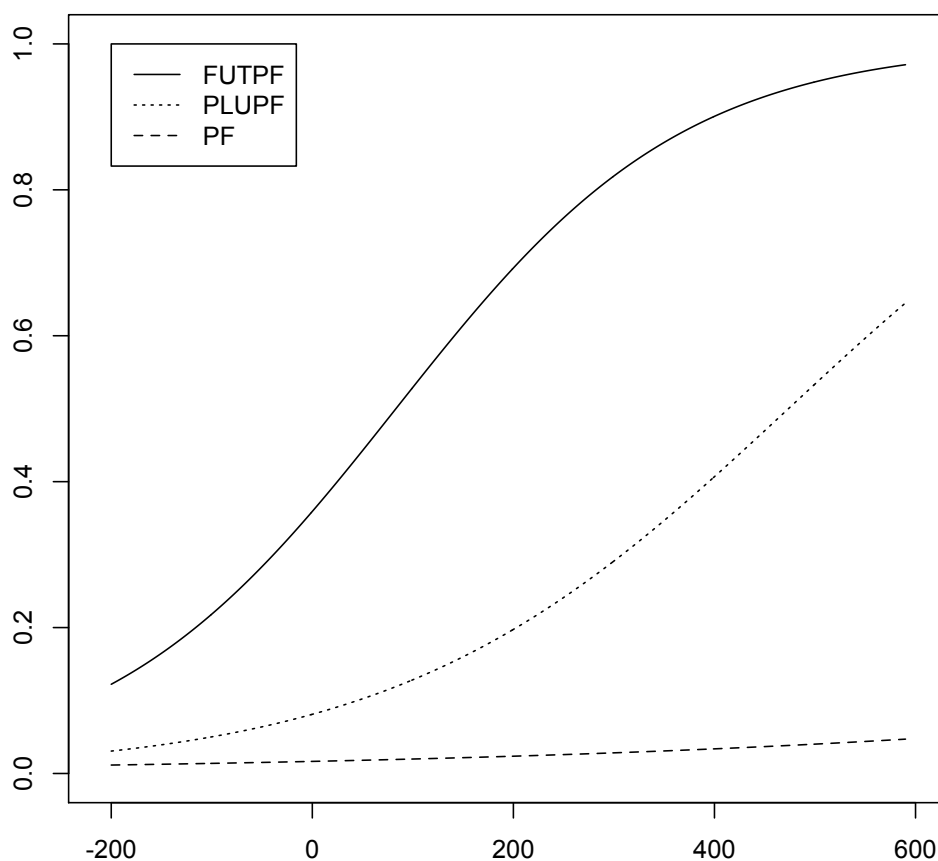
Graph 5: Relative frequency of F-periphrases across three types of Tense (in percentages, over time).

Just like Graph 3, this plot shows the relative frequency of F-periphrases in a number of discrete conditions. We immediately see that the data points are scattered over a much wider space. Crucially however, the dispersion is not random: we see three distinct zones, which become even more clearly separated in Late Latin. Low down we see most of the data points for the perfect tenses. Especially perfect indicatives are very rare throughout most of the period under investigation. For those (Late Latin) data points where we observe frequencies higher than 10%, it is perfect infinitives of the type *amatus fuisse* which are more or less

productive.<sup>6</sup> In contrast, F-periphrases in the pluperfect and the future perfect seem fully productive from the first century AD onwards. Late Latin future perfects of the F-type are the norm.

The diachronic trajectory of the future perfects is of particular interest. Although we are missing a bit of the lower tail of the curve (i.e. the stage where the type *amatus fuerit* is completely absent is not documented), it seems clear that this pattern follows the characteristic ‘S-shaped’ distribution which is often found when a new linguistic form gradually replaces an older one (see for instance Kroch 1989). To the best of my knowledge, this is the first time ever that such an instance of (morpho)syntactic change in Latin has been documented in its entirety by means of corpus data. Importantly, the spread of the F-future perfects provides a nice case study of a perfectly regular pattern of language change, which in turn strongly suggests that the corpus described in section 2.2 can be considered to be representative of how the ‘real’ language naturally evolved.

To conclude this section, Graph 6 visualizes the results of a regression with ‘Time’, ‘Tense’ and an interaction term between the two as predictors. Observe that the modelled data hardly deviate from the observed data shown in Graph 5:



Graph 6: The effect of Tense on the likelihood of a BE-periphrasis to be of the F-type (probabilities over time).

Before we proceed, let me just point out that we can now make sense of the remarkable behaviour of Cassius Felix in Graph 1 in section 3.1. As it happens, this text exhibits a very

<sup>6</sup> This distinction is not shown in Graph 3. Plautus seems to be an exception here: of the 34 perfect F-periphrases in his works, 26 are indicatives (compared to 1 infinitive and 7 subjunctives). As Jim Adams (p.c.) points out to me, it is not entirely clear whether a Plautine *amatus fui* has the same interpretive characteristics as a Late Latin one. I leave this issue for future research.

high frequency of analytic future perfects, namely 77 out of a total of 102 BE-periphrases. All but one of these are of the F-type, whence the very high proportional frequency of F-periphrases in this author. In all likelihood, this is in the first place to be understood as a genre effect rather than anything particular about the grammar of Cassius Felix, future perfects being much more common in technical treatises than in for instance historical narratives.

### 3.2.3 A full model

For the sake of completeness, I will conclude this first case study by showing the results of a (fairly simple) multivariate regression model with the factors ‘Clause type’ (main vs. embedded) and ‘Mood’ (indicative, infinitive or subjunctive) as additional predictors (on top of ‘Tense’ and ‘Date’, as well as a number of meaningful interaction terms). Unsurprisingly, the factor ‘Voice’ did not turn out to be significant in a sufficiently rich model, neither as a main effect nor in any interaction term. The results are summarized in Table 5:

	<b>Log odds</b>	<b>p-value</b>	<b>Significant?</b>
(Intercept)	-0.4495481	1.08e-06	***
Date	0.0073501	< 2e-16	***
Tense (baseline: FUTPF)      PF	-3.2176144	< 2e-16	***
PLUPF	-1.7578817	< 2e-16	***
Mood (baseline: indicative)    infinitive	-0.2959081	0.1826	
subjunctive	-0.2923973	0.0187	*
Clause type (baseline: embedded)	-1.1326310	< 2e-16	***
Date*Tense_PF	-0.0062751	< 2e-16	***
Date*Tense_PLUPF	-0.0028250	2.34e-08	***
Date*Mood_infinitive	0.0058249	< 2e-16	***
Date*Mood_subjunctive	0.0027603	1.66e-13	***
Date*Clause type	-0.0025125	3.29e-09	***

Table 5: Summary of a logistic model predicting the probability for a BE-periphrasis to be of the F-type.  
Significance levels:  $p < 0.001 = \text{***}$ ;  $p < 0.01 = \text{**}$ ;  $p < 0.05 = \text{*}$ ;  $p < 0.1 = \text{'}$ .

The figures in this table are to be interpreted as follows. The first column lists the independent variables taken up in the model. For each categorical variable, it is indicated which level is taken as the reference category (baseline). The strength of a given predictor can be read off from the estimates, so-called log odds, given in the second column. The sign of this value indicates whether the presence of a particular variable favours (positive value) or disfavors (negative value) a BE-periphrasis to be of the F-type. The further away from zero, the stronger the effect. Finally, a p-value is provided which indicates to what extent we can be confident that the reported log odds are an accurate estimate.

The model summarized in Table 5 exemplifies the multivariate type of set-up which allows us to evaluate the role of a predictor X in explaining the distribution of two (and potentially more) competing linguistic variants, and at the same time to compare its predictive power with that of other covariates. If properly used, this routine constitutes a very powerful tool to describe and analyse complex and intricate data sets. Note that many other factors could be added to the mix we already have, including sociolinguistic ones. However, given the (for historical data exceptionally) good model fit we have now (ROC score (Area Under the Curve): 0.9124 (a value which corresponds to a so-called classification accuracy of over 91%); Tjur’s Coefficient of Discrimination: 0.4260041), we can assume that at least

descriptively, the rise of F-periphrases is basically a matter of two predictors, namely Time and Tense.

### 3.3 *Intermediate conclusion*

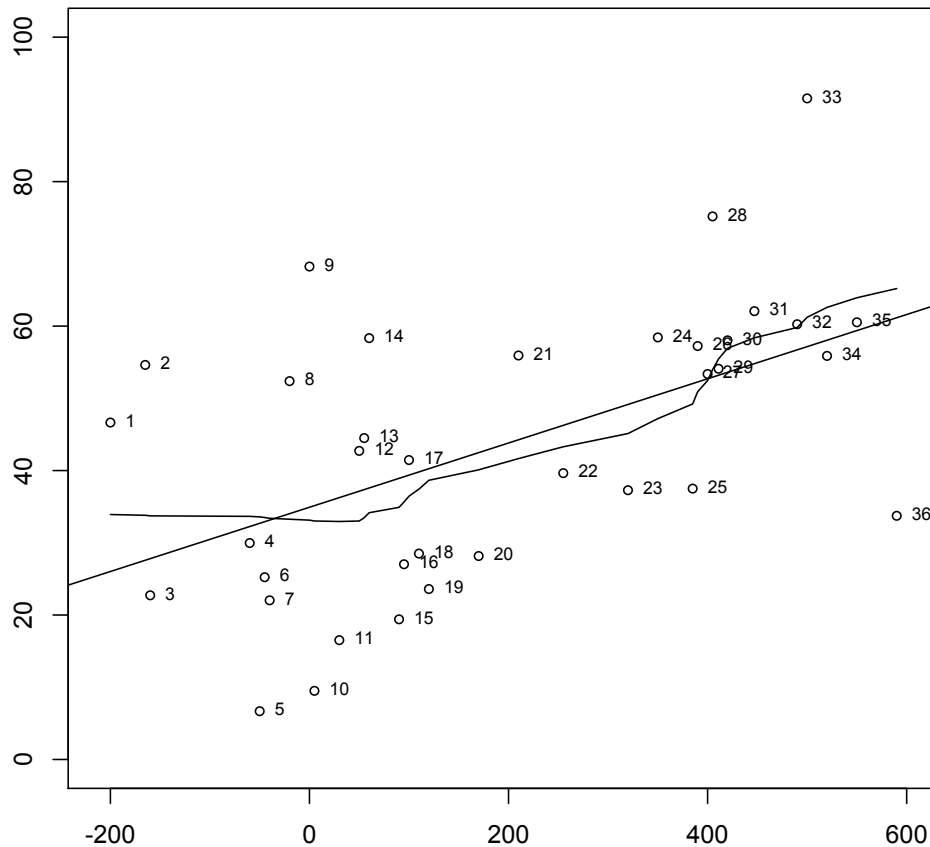
The main result arrived at in the past section is that the origins of the Latin F-periphrases are to be sought first and foremost in the tense system, and more particularly in the future perfects. It was also suggested that one factor occasionally mentioned in the literature, namely the distinction between deponents and passives, cannot be shown to be of diachronic interest. However, it is clear that these generalizations are only descriptive, and in need of further explanation. One particular question is whether, and if yes, to what extent, early E and F-periphrases were functionally equivalent. Did they set out as free variants, or did they originally have a different function? Addressing these questions unfortunately lies beyond the scope of this paper.

In the next section, I turn to a second topic related to the alternation between E and F in the history of Latin, namely the loss of synthetic passives in the *infectum*.

## **4 The genesis of Romance analytic present tense passives: insights from word order**

### 4.1 *A remarkable discrepancy*

As is well known, in the evolution from Latin towards Romance, a major change concerning the relative order of auxiliaries and dependent non-finite verbs takes place. In Latin, these two elements can be linearized in either order, whereas in present day Romance languages, the head-initial order ‘auxiliary - non-finite verb’ is the only available option. There is good evidence to assume that the shift towards this new system was already going on in Latin. For instance, in clauses with the modal auxiliary *possum* and an infinitival VP-complement, we can see a clear rise in head-initiality (i.e. the order ‘*possum* - infinitive’) in our corpus (for the sake of simplicity, I am lumping cases where the auxiliary and the non-finite verb are contiguous together with cases where they are not):

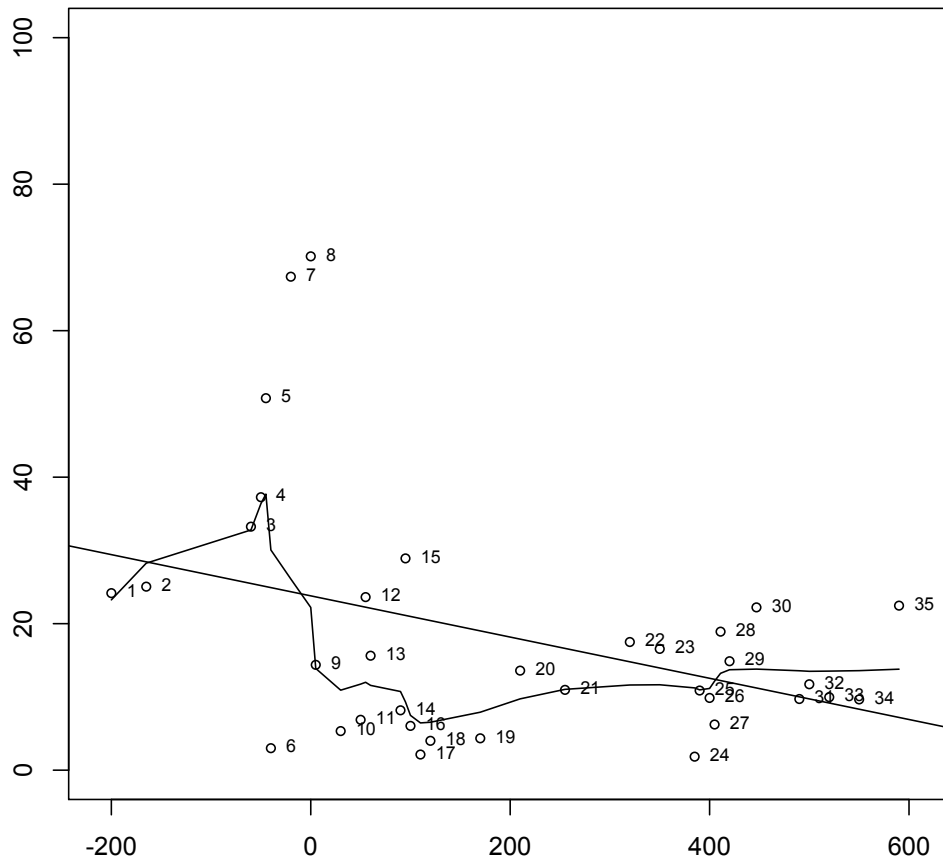


Graph 5: Relative frequency of the order 'possum - infinitive' (in percentages) over time. Case labels: 1= Plautus, 2= Terence, 3= Cato, 4= Cicero, 5= Caesar, 6= Varro, 7= Sallust, 8= Hyginus, 9= Vitruvius, 10= Livy, 11= Celsus, 12= Seneca, 13= Columella, 14= Petronius, 15= Frontinus, 16= Quintilian, 17= Pliny, 18= Tacitus, 19= Suetonius, 20= Gaius, 21= Tertullian, 22= Cyprian, 23= *Historia Augusta*, 24= Palladius, 25= *Itinerarium Egeriae*, 26= Jerome, 27= Augustine, 28= Vulgate, 29= *Gesta Conlactionis Carthaginensis*, 30= Vegetius, 31= Cassius Felix, 32= Victor Vitensis, 33= Pompeius Maurus, 34= Caesarius of Arles, 35= Iordanes, 36= Gregory of Tours.

Despite the fact that during the entire extended period of some eight centuries, there is a good deal of synchronic variation (the exact nature of which remains to be better understood), it is clear that the overall trend is one towards more head-initiality (witness the quasi-identical trajectories of the two regression lines), which is perfectly in line with what we know about the old and present day Romance languages.

However, no such increase in the order 'BE - past participle' can be observed in the case of E-periphrases. Rather, Late Latin texts exhibit a remarkably uniform preference for the order *amatus est*, in contrast with earlier periods, where more variation seems to be the norm:





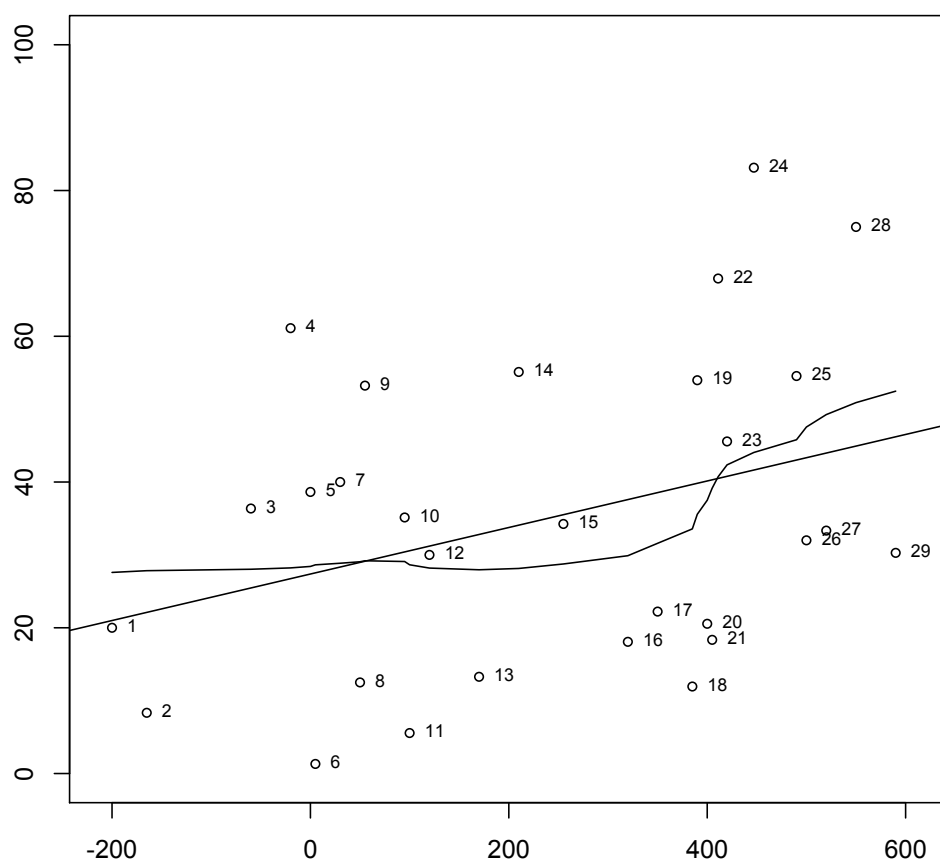
Graph 6: Relative frequency of the order 'BE - past participle' in E-periphrases (in percentages) over time. Case labels: 1= Plautus, 2= Terence, 3= Cicero, 4= Caesar, 5= Varro, 6= Sallust, 7= Hyginus, 8= Vitruvius, 9= Livy, 10= Celsus, 11= Seneca, 12= Columella, 13= Petronius, 14= Frontinus, 15= Quintilian, 16= Pliny, 17= Tacitus, 18= Suetonius, 19= Gaius, 20= Tertullian, 21= Cyprian, 22= *Historia Augusta*, 23= Palladius, 24= *Itinerarium Egeriae*, 25= Jerome, 26= Augustine, 27= Vulgate, 28= *Gesta Conlotionis Carthaginiensis*, 29= Vegetius, 30= Cassius Felix, 31= Victor Vitensis, 32= Pompeius Maurus, 33= Caesarius of Arles, 34= Iordanes, 35= Gregory of Tours.

Even upon very superficial inspection, it is obvious that the actual facts are in strong contrast with claims commonly found in the literature. Compare for instance Bauer (2006: 294): 'A statistical overview of the occurrence of *factus est* in a number of Latin texts from different periods shows a steady decline of the left-branching sequence'. This clearly cannot be correct. However, what we can also not say is that there is a real decline of the head-initial (right-branching) order, despite the straight regression line going down. As the smoothed line suggests, there simply is no single linear trend in these data. To the extent that any generalization can be made, it seems to be the case that the language evolves from a system with more to a system with less variation.

Note in passing that in the earlier period the highest frequencies of the order *est amatus* (including quite a few cases of *est ... amatus*) are found in Vitruvius and Hyginus, with Varro in third place. It is of course very tempting to hypothesize that the observed tendencies are related to the fact that the authors involved wrote technical treatises, a genre which can be assumed to require a less literary style than certain other texts from the same period. Although it would remain to be seen how such a correlation would have to be interpreted, these data would constitute a case where a synchronic factor like 'Genre' is correlated with the distribution of a given linguistic variant. Note however that the strength of this effect (if real) is certainly not to be overestimated: other technical writers such as Columella, Celsus and Frontinus (who were all active one or two generations after the previous set of authors)

display a different behaviour. In any event, considerations like the above illustrate the type of issues that would deserve to be looked into in more detail in future research.

Interestingly, we obtain a very different picture when we look at word order preferences in F-periphrases. Observe that this data set contains fewer authors/texts than the previous one: once again in order to control for inaccurate estimates stemming from small sample sizes, I only took into account data from authors/texts containing at least 10 F-periphrases, which excludes Caesar, Varro, Sallust, Petronius, Frontinus and Tacitus.



Graph 7: Relative frequency of the order ‘BE - past participle’ in F-periphrases (in percentages) over time. Case labels: 1= Plautus, 2= Terence, 3= Cicero, 4= Hyginus, 5= Vitruvius, 6= Livy, 7= Celsus, 8= Seneca, 9= Columella, 10= Quintilian, 11= Pliny, 12= Suetonius, 13= Gaius, 14= Tertullian, 15= Cyprian, 16= *Historia Augusta*, 17= Palladius, 18= *Itinerarium Egeriae*, 19= Jerome, 20= Augustine, 21= Vulgate, 22= *Gesta Conlotionis Carthaginensis*, 23= Vegetius, 24= Cassius Felix, 25= Victor Vitensis, 26= Pompeius Maurus, 27= Caesarius of Arles, 28= Iordanes, 29= Gregory of Tours.

Although we can presumably not say that the F-periphrases behave exactly like the ‘*possum* + infinitive’ combinations reviewed earlier, it is quite clear that there is a strong contrast between E-periphrases on the one hand, and *possum* and the F-periphrases on the other. More particularly, only the latter display the kind of behaviour that we expect from structures that are evolving from Latin towards the Romance languages, as witnessed by the slow but unmistakable increase of head-initial (‘auxiliary - non-finite verb’) word orders.

Note that it is very unlikely that the pattern observed in Graph 6, viz. the ‘fossilization’ of the order *amatus est* in Late Latin, is to be ascribed to language external factors, like for instance normative pressure, resulting in ‘mechanical’ imitation of earlier word order habits. Under such a scenario, one would expect the extent to which a given author imitates older models to co-vary with sociolinguistic variables, such as his/her level of education, or the intended audience. However, no such variation seems to emerge, at least not in any obvious

fashion. Rather, the Late Latin texts in the corpus all display a very similar behaviour (witness the dense cloud of dots in the lower right corner of Graph 6), which is perhaps all the more surprising given the very heterogeneous character of the text corpus, not least in terms of the geographical origins (place of composition) of the individual texts. Moreover, under this view the contrast between E-periphrases and modals/F-periphrases remains entirely mysterious (why wouldn't the latter be subject to normative pressure?). Finally, and most importantly, as we will see in section 5.2, the conditions under which the order 'auxiliary - past participle' in E-periphrases is favoured over the reverse order change over time, suggesting that the Late Latin system is also qualitatively different from the earlier grammar. As a result, I will stick to the null hypothesis and assume that the facts observed in the corpus accurately reflect the real evolution of the Latin language.

#### 4.2 A diachronic hypothesis

In order to account for the discrepancy between word order preferences in E-periphrases on the one hand, and all other (monoclausal) environments involving an auxiliary and a non-finite VP-complement on the other, I would like to propose that Late Latin E-periphrases exhibit the behaviour they do because they were on their way out from the language. For this reason, they did not take part in a particular type of syntactic change which otherwise affected the entire language system, namely the shift towards a strictly head-initial Tense node (i.e. a grammar that only generates the order 'auxiliary - non-finite verb'). In Danckaert (2015), it is argued that the syntactic and phonological corollaries of this evolution are to be understood in terms of cliticization of the BE-auxiliary to its (leftward) complement. I shall not further elaborate on the details of this analysis here.

The main consequence of this line of reasoning is that *amatus est* type structures cannot be the source of Romance formations like French *je suis aimé* and Italian *sono amato* (i.e. present tense passives which one could consider F-periphrases in the *infectum*), contrary to what seems to be the *communis opinio* on the topic (see for instance Winters 1984: 450-51, Hewson 1997: 315, Clackson and Horrocks 2007: 280 and de Melo 2012b: 84 for explicit claims along these lines). The standard account relates the formation of the Romance pattern to the well known fact that the original Latin *amatus est* periphrases were from the earliest stages ambiguous between a genuine verbal reading yielding an analytic perfect tense and an adjectival copular construction where the tense of the auxiliary exhaustively determines the tense of the entire expression (cf. section 2.3). Once the F-periphrases entered into competition with, and gradually took over from, the perfect tense *amatus est*, the latter disappeared, without this evolution affecting the adjectival *amatus est* structure. Rather, the latter reading survived, and was then 'recycled' to form a set of *infectum* passives, which in turn gradually ousted the pre-existing synthetic forms. The mechanism by which the extension from *perfectum* to *infectum* is driven is argued to be 'analogical leveling' of the verbal paradigm (on the role of analogy in bringing about this evolution, see Kravar (1966), Clackson and Horrocks (2007: 280) and de Melo (2012b)).

My main problem with this account is that it leaves the word order facts reviewed in the previous section entirely unexplained (as it happens, in the literature on this topic, word order is simply never discussed). In addition, note that at no point is any independent evidence for the putative extension from an adjectival present tense passive towards a verbal passive ever offered. The semantics of these two structures are clearly very different, and although the proposed shift does not seem *a priori* unlikely, one would want some additional details about how exactly this evolution proceeded.

As an alternative, I take the Romance present tense passives to be new formations, created by analogical extension of the already existing F-periphrases in the *perfectum* half of

the paradigm. In other words, I do not dispute the claim that analogy plays an important role in bringing about the relevant evolution, but I reject the view that one and the same object was ‘transformed’ and given a new function (or more accurately, that the range of meanings/functions of one and the same object was broadened).

An advantage of this slightly different approach is that we now have a handle on the word order facts. The earliest F-periphrases in the *infectum* seem to date from the fourth century (cf. de Melo 2012b). We can assume these new formations to inherit the word order probabilities of the structure they were modeled on, namely the older *perfectum* F-periphrases. As a result, they regularly take part in the general changes that affect Late Latin, including the shift towards generalized head-initiality characteristic of the present day Romance languages. Moreover, we do not run into the semantic problem just mentioned: one verbal passive was modelled on another verbal passive, and adjectival formations do not entire the picture.

## 5 Towards a better understanding of word order in BE-periphrases

In this final section, I would like to take the discussion in section 4 one step further, by applying the same type of analysis that I used in section 3 to study the origins of the F-paradigm, namely multivariate logistic regression, to the issue of the word order alternation ‘BE- past participle’ - ‘past participle - BE’. Although the facts in this empirical domain are clearly much more complex, some interesting patterns arise. One factor that can be shown to be a good predictor for this alternation is whether or not the auxiliary involved is monosyllabic.

### 5.1 An interesting testimonium

I will start the discussion by looking at a very interesting passage from the grammarian Sacerdos, who in all likelihood lived in the third century AD (cf. Kaster 1988: 352-3).<sup>7</sup> In the second book of his *Artes grammaticae*, the author discusses, among other things, prose rhythm (starting at GLK 6: 492.25). He begins by observing that in his day, the aesthetic preferences in this domain are different from what was customary earlier (the point of reference being Cicero). More particularly, earlier authors are said to have been reluctant to sacrifice the expressive power or rhetorical effect of a given word order pattern for the sake of ‘euphonic’ considerations of prose rhythm. According to Sacerdos, the rules governing clausula composition cannot be so overridden in his own days.

Interestingly, as a specific example of a word order pattern which is no longer tolerated in his day, Sacerdos mentions monosyllables occurring in clausulae. All the examples he uses involve monosyllabic BE-E auxiliaries (GLK 6: 493.11-20):

quod in primis est uidelicet nostro tempore uitiosum, Tullius non dubitauit uerbo monosyllabo finire structuram, ut ‘ab istius petulantia conseruare non licitum est’ et ‘quae cum his ciuitatibus C. Verri communicata sunt’ et ‘quod P. R. iam diu flagitat, extincta atque deleta sit’. hae compositiones demutatae facient nostri temporis structuram sic, ex tribrachy et ditrochaeo ‘ab istius petulantia non est licitum conseruare’, ex trochaeo et dactylo et ditrochaeo ‘quae sunt C. Verri cum his ciuitatibus copulata’, ex trochaeo et bacchio a longa ‘id quod P. R. iam diu flagitat, extincta sit atque deleta’.

Tullius did not hesitate to end a period with a monosyllabic word, which in our time is clearly a major offence. We thus find ‘ab istius petulantia conseruare non licitum est’, ‘quae cum his ciuitatibus C. Verri communicata sunt’ and ‘quod P. R. iam diu flagitat, extincta atque deleta sit’. When altered, in our days these compositions will produce structures such as the following: one consisting of a tribrach

<sup>7</sup> Many thanks to Jim Adams for pointing this passage out to me.

and a ditrochee, like *'ab istius petulantia non est licitum conseruare'*, a trochee followed by a dactyl and a ditrochee, yielding *'quae sunt C. Verri cum his ciuitatibus copulata'*, or a trochee and a bacchius starting with a long syllable, as in *'id quod P. R. iam diu flagitat, extincta sit atque deleta'*.

Importantly, the three examples from Cicero that Sacerdos considers *uitiosum* all involve the sequence 'past participle - BE-auxiliary' (*licitum est, communicata sunt* and *deleta sit*). So how do we interpret this passage, given what we have seen in the previous section? First of all, note that it is not entirely clear whether the author refers to all monosyllabic words (verbs, nouns, adjectives etc. alike), which would be the case if we interpreted the phrase *uerbo monosyllabo finire structuram* literally, or whether it narrowly refers to monosyllabic forms of *esse* only, which one could hypothesize on the basis of the examples that Sacerdos gives. In any event, observe that what the author refers to is clearly not a categorical ban on the order 'past participle - auxiliary': rather, he only states (or at least implies) that in his time, the relevant order is (strongly) dispreferred if (i) the auxiliary involved is monosyllabic and (ii) the entire sequence appears at the end of a period or colon, where it is subject to the prescriptive rules on clausula formation. Note in passing that in his own writings, Sacerdos does not seem to avoid the order 'past participle - BE'. In his *Artes*, we find a total of 71 BE-periphrases. In line with what we have seen earlier, word order preferences here are correlated with whether a given periphrastic expression is of the E or of the F-type. Out of a total of 60 E-periphrases, 52 (86.7%) appear in the order 'past participle - BE'. Conversely, less than half (5 out of 11, i.e. 45.5%) of the F-periphrases appear in this order. This difference is statistically very significant (Chi-square with Yates correction,  $p = .006$ ). Finally, note that the grammarian rewrites the Ciceronian passages not by simply reversing the order of the participle and the auxiliary (which would only remove the latter from the very last position of the structure, but not from the clausula altogether). The third rewritten example even contains the 'head-final' sequence *extincta sit*.

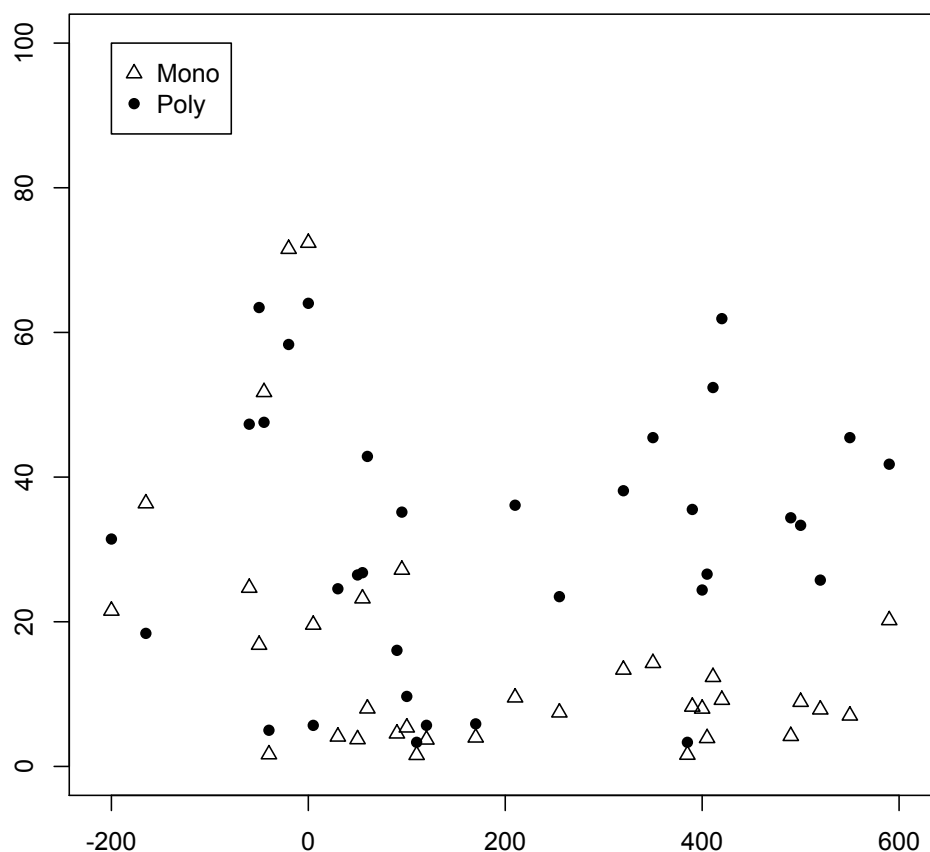
To conclude, the main lesson we can draw from this passage is that it might be very interesting to study the interaction between prose rhythm and word order in BE-periphrases. Although a comprehensive study of this topic is clearly a very complex enterprise which goes well beyond the scope of this study, what we can do is test whether on top of (or perhaps even instead of) the two factors governing word order in BE-periphrases that have been taken into account up to this point (namely (i) time and (ii) the distinction between E and F-periphrases), the phonological shape (monosyllabic or otherwise) of the auxiliary also plays a role. In the next section, I will show that this factor does indeed have explanatory power, albeit not quite in the way that one could expect on the basis of Sacerdos' prescriptions.

## 5.2 Predicting word order in BE-periphrases: a first attempt

Let us first have a look at the raw data. In Graph 8, the data set summarized in Graph 6, which reported on the overall diachronic development of word order preferences in E-periphrases, is split up into two subgroups, distinguishing monosyllabic and polysyllabic BE-auxiliaries.<sup>8</sup>

---

<sup>8</sup> Needless to say, in F-periphrases the auxiliary can never be monosyllabic, which is why only E-patterns are considered here.



Graph 8: Relative frequency of the order ‘auxiliary - past participle’ in E-periphrases over time: Monosyllabic and polysyllabic auxiliaries compared.

What we see is that without a single exception, Late Latin authors show higher frequencies of the order ‘auxiliary - past participle’ if the auxiliary involved is polysyllabic. No such thing is true in Early and Classical Latin, where there is once again more variation. At this point, we can provisionally conclude that the *testimonium* of Sacerdos discussed in the previous section does not pose any major threats for the conclusion arrived at the end of section 4. Although it remains to be seen whether or not Late Latin texts which pay attention to prose rhythm do indeed avoid the sequence *amatus est* at the end but not in the middle of a period (which as I said would indeed be very interesting to investigate systematically), there clearly is no such thing as a general Late Latin aversion to the sequence ‘past participle - monosyllabic auxiliary’. Quite the contrary: periphrases with a monosyllabic auxiliary heavily favour the head-final order (that is, when we average over *in pausa* and non *in pausa* environments).

However, at this point a more interesting question arises, namely whether the facts just reviewed make the story developed in section 4 concerning the interaction between word order and the distinction between E and F-periphrases, superfluous. Given that monosyllabic BE-auxiliaries only occur in E-periphrases (alongside bi- and trisyllabic ones), and never in F-periphrases (where the auxiliary always counts 2, 3 or 4 syllables), one can rightfully ask the question whether the putative different word order preferences between E and F-periphrases is not just an artefact of the number of syllables of the auxiliaries involved. In other words, if much of the observed variation, and more particularly the unexpected Late Latin preference for the head-final order can be accounted for by distinguishing monosyllabic and polysyllabic auxiliaries, the distinction between E and F-periphrases might well be dispensed with. Some frequency facts suggest that this might actually be true: out of 27,545 BE-periphrases in my corpus, in the vast majority (20,048 to be precise) of the cases, the auxiliary is monosyllabic.

Much of the Late Latin head-finality could therefore be purely an effect of this factor, in whatever terms it is to be further explained.

In order to verify whether the factor ‘E vs. F’ has any explanatory power in predicting word order preferences alongside the factor ‘monosyllabic vs. polysyllabic auxiliary’, one has to evaluate the role of both of these in one and the same — once again multivariate — model. In order to increase the accuracy of the results, I included a number of other independent variables in the model building procedure, such as ‘Mood’ (the distinction between indicatives, subjunctives and infinitives) and ‘Clause type’ (main versus embedded clauses) and ‘Negation’ (presence of absence of the sentential negator *non*). Only those authors/texts were taken up which contain at least 8 negated BE-periphrases. As a result, data from Caesar, Sallust, Hyginus, Petronius, Frontinus, Pliny, Tacitus, Suetonius, Palladius, the *Itinerarium Egeriae*, Vegetius, Cassius Felix, Victor Vitensis and Iordanes were not included (which still leaves us with a total of 22 authors/texts). In addition, all ‘contracted’ forms of the type *amatust* from Plautus and Terence were excluded. The results obtained are summarized in Table 6:

	Log odds	p-value	Significant?
(Intercept)	-0.6050179	< 2e-16	***
Date	-0.0006961	8.40e-06	***
Negation (baseline: affirmative)	0.2081590	0.195167	
Monosyllabic auxiliary (baseline: polysyllabic)	-0.7579604	< 2e-16	***
E vs. F (baseline: E)	-0.2820330	0.002917	**
Date*Negation	0.0017328	0.002180	**
Date*Monosyllabic	-0.0015821	< 2e-16	***
Date*E/F	0.0008890	0.001128	**
Negation*Monosyllabic	0.7344238	0.000532	***
Date*Negation*Monosyllabic	0.0039825	1.33e-07	***

Table 6: Summary of a logistic model predicting the probability of obtaining the order ‘BE - past participle’.  
Significance levels:  $p < 0.001 = \text{***}$ ;  $p < 0.01 = \text{**}$ ;  $p < 0.05 = *$ ;  $p < 0.1 = \text{'}$ .

First of all, the model summarized here cannot be considered optimal in terms of goodness of fit (ROC score (Area Under the Curve): 0.6918; Tjur’s Coefficient of Discrimination: 0.07838583; compare the much higher values obtained for the model discussed in section 3.2.3). In all likelihood, this is at least partly because I only took into account one factor related to how BE-periphrases interact with clausula rules, namely (one aspect of) the phonological make-up of the auxiliary involved. To arrive at a better understanding of the relevant phenomenon one would also have to look at (i) the phonological properties of the participle and (ii) the linear position (*in pausa* or otherwise) of the entire BE-periphrasis. Moreover, I did not make any distinction between contiguous BE-periphrases (cases where the auxiliary and the past participle are linearly adjacent) and non-contiguous ones. Many additional factors worth investigating come to mind (as always, sociolinguistic variables are expected to add to our understanding of the data).

Secondly, and most importantly, we see that the ‘E vs. F’ factor does indeed survive in this richer model (both as a main effect and in an interaction term with time), which suggests that there is no reason to reconsider the conclusion arrived at in section 4. Note however that it is clear that its weight is less strong than the predictor concerning the number of syllables of the auxiliary (witness the extremer log odds estimated for the latter, cf. the values in the shaded cells).

Third, as can be seen in Table 6, the model includes a main effect of negation, as well as a number of interaction terms including this predictor. The factors ‘Mood’ and ‘Clause type’

did not turn out to be significant and were therefore left out. As it happens, the presence or absence of sentential negation is actually a very strong predictor for word order, and it interacts in a very interesting way with the factor ‘Number of syllables’. In addition, the role of negation is subject to an interesting diachronic development (cf. the significant interaction term with ‘Time’), which for reasons of space cannot be discussed here. However, we now see very clearly that the conditions under which one is likely to obtain or not obtain the order ‘BE - past participle’ in earlier centuries are very different from the ones governing this alternation in Late Latin, which can be considered compelling evidence against the view that the predominance of Late Latin *amatus est* is a matter of ‘mechanical’ imitation of earlier, ‘classical’ models (see also section 4.1 above). On the contrary, there is every reason to assume that the syntax of Late Latin BE-periphrases is both qualitatively (influence of negation, effect of number of syllables, E vs. F) and quantitatively (more vs. less variation) different from that of Early and Classical Latin.

## 6 Conclusion

In this paper, I have discussed a number of aspects concerning the diachronic development of BE-periphrases in Latin. As shown in section 3, the data on the origin and spread of the F-pattern constitute a clear case where the course of syntactic change is characterized by ‘overt’ rather than ‘submerged’ continuity, as texts from the classical era canonically represent an intermediate stage between early and later sources. At a technical level, this was shown to be true by adding the continuous variable ‘Time’ to a regression model predicting the alternation between E and F-periphrases in a large corpus of texts, and by observing that this variable has significant predictive power. As discussed in section 4, data on word order in clauses with an auxiliary (BE in F-periphrases, or the modal *possum*) and a non-finite verb also suggest that language change in Latin follows an entirely continuous path.

On the other hand, word order in E-periphrases was shown to be more problematic: especially in Early and Classical Latin, the alternation between the orders *amatus est* and *est amatus* is still poorly understood. Enriching the model discussed in section 5 with factors such as information structure (does the information conveyed by the participle convey salient information or not?), as well as the register and genre of a given text can be hoped to add to our understanding of the variation observed. As pointed out, the role of prose rhythm would also be interesting to look into.

Finally, although it is clear that quantitative corpus data should always be gathered and interpreted with great care and caution, I hope to have shown that the type of methodology presented and applied in this paper can be a very useful tool to formulate detailed and robust *explananda*. Assuming the extant body of Latin texts to be a reliable source of both synchronic and diachronic variation, we can hope that future corpus studies will further enhance our understanding of syntactic and sociolinguistic variation in Latin.

## References

- Bauer, B. (2006), ‘“Synthetic” vs. “analytic” in Romance: the importance of varieties’ in Gess, R. and Arteaga, D. eds., *Historical Romance Linguistics: Retrospective and Perspectives* (Amsterdam: Benjamins), 287-304.
- Brix, J. (1901), *Ausgewählte Komödien des T. Maccius Plautus 4: Miles Gloriosus*, 3rd edn revised by M. Niemeyer (Leipzig: Teubner).
- Brunet, E. and Mellet, S. (n.d.), *Hyperbase, version 5.5: textes latins* (companion to CD-ROM) (Nice: Université de Nice Sophia-Antipolis).



- Brookes, J. (2014), *Probabilistic and Multivariate Modelling in Latin Grammar: The Participle-Auxiliary Alternation as a Case Study*. Unpublished Ph.D. diss., The University of Manchester.
- Clackson, J. and Horrocks, G. (2007), *The Blackwell History of the Latin Language* (Oxford: Blackwell).
- Coussé, E. (2008), *Motivaties voor volgordevariatie: een diachrone studie van werkwoordwoordsvolgorde in het Nederlands*. Unpublished Ph.D. diss., Ghent University.
- Danckaert, L. (2015), 'The genesis of Romance analytic BE-passives: evidence from word order' submitted to Mathieu, E. and Truswell, R. eds., *From Micro-change to Macro-change*, (Oxford: Oxford University Press).
- de Melo, W. (2012b), 'Kuryłowicz's first 'law of analogy' and the development of passive periphrases in Latin' in Probert, P. and Willi, A. eds., *Laws and Rules in Indo-European*, (Oxford: Oxford University Press), 83-101.
- de Sutter, G. (2005), *Rood, groen, corpus! Een taalgebruiksgebaseerde analyse van woortvolgordevariatie in tweeledige werkwoordelijke eindgroepen*. Unpublished Ph.D. diss., Catholic University of Leuven.
- de Sutter, G. (2009), 'Towards a multivariate model of grammar' in Dufter, A., Fleischer, J. and Seiler, G. eds., *Describing and Modeling Variation in Grammar* (Amsterdam: Benjamins), 225-54.
- Gehrke, B. (2013), 'Still puzzled by adjectival passives?' in Folli, R., Sevdali, C. and Truswell, R. eds., *Syntax and its Limits* (Oxford: Oxford University Press), 175-91.
- Gehrke, B. (to appear), 'Adjectival participles, event kind modification and pseudo-incorporation', *Natural Language and Linguistic Theory*.
- Herman, J. (2002), 'La disparition du passif synthétique latin: nouvel essai sur l'écrit et le parlé en latin mérovingien', *Estudis romànics* 24, 31-46.
- Hewson, J. (1997), 'From Latin to Modern Romance' in Hewson, J. and Bubeník, V. eds., *Tense and Aspect in Indo-European Languages* (Amsterdam: Benjamins), 314-30.
- Kaster, R. (1997), *Guardians of Language: The Grammarian and Society in Late Antiquity* (Berkeley: University of California Press).
- Kravar, M. (1966), 'Sintaksa tvorbe -tus fui u latinskom [La syntaxe de la formation -tus fui en latin]', *Živa Antika* 16, 223-32.
- Kroch, A. (1989), 'Reflexes of grammar in patterns of language change', *Language Variation and Change* 1, 199-244.
- Labov, W. (1966), *The Social Stratification of English in New York City* (Washington, D.C.: Center for Applied Linguistics).
- Laughton, E. (1964), *The Participle in Cicero* (Oxford: Oxford University Press).
- Leumann, M. (1921), 'Part. perf. pass. mit fui im späteren Latein', *Glotta* 11, 192-4.
- Lodge, G. (1904), *Lexicon Plautinum*, 2 vols. (Leipzig: Teubner).
- McGlynn, P. (1963-67), *Lexicon Terentianum*, 2 vols. (London: Blackie).
- Muller, H. (1924), 'The passive voice in Vulgar Latin', *The Romanic Review* 15, 68-93.
- Pintzuk, S. (2003), 'Variationist approaches to syntactic change' in: Joseph, B. and Janda, R. eds., *The Handbook of Historical Linguistics* (Oxford: Blackwell), 509-28.
- Verbraken, P. (1976), *Études critiques sur les sermons authentiques de saint Augustin* (= Instrumenta Patristica 12) ('s-Gravenhage: Nijhoff).
- Winters, M. (1984), 'Steps toward the Romance passive inferrable from the *Itinerarium Egeriae*', *Romance Philology* 37, 445-54.
- Weinreich, U., Labov, W. and Herzog, M. (1968), 'Empirical foundations for a theory of language change' in: Lehmann, W. and Malkiel, Y. eds., *Directions for Historical Linguistics: A Symposium* (Austin: University of Texas Press), 95-188.