



Assessing phonological correlates of syntactic change the case of Late Latin weak BE

Lieven Danckaert

► To cite this version:

Lieven Danckaert. Assessing phonological correlates of syntactic change the case of Late Latin weak BE. Syntactic Features and the Limits of Syntactic Change, Oxford University Press, pp.178-209, 2021, 10.1093/oso/9780198832584.003.0009 . hal-03327148

HAL Id: hal-03327148

<https://hal.science/hal-03327148>

Submitted on 27 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Assessing phonological correlates of syntactic change:
the case of Late Latin weak BE

Lieven Danckaert (CNRS, Université de Lille)

1. Introduction

The starting point of this paper is the observation that throughout the history of Latin, functional verbs of various kinds ('auxiliaries') and their non-finite complements can be linearised in either order. For instance, in (1) we see that a BE-auxiliary in a periphrastic passive can either precede or follow a past participle like *factus* 'become'. Similarly, in (2) we see the word orders 'modal - infinitive' and 'infinitive - modal' appear side by side in the same sentence (the relevant modal verb being *possum* 'CAN, be able'):

- (1) Ann-is XIII ante quam praetor **fact-us** **est**
years-ABL fourteen before than praetor.NOM become-NOM.M.SG be.PRS.3SG
legat-us esse <qui> pot-uisse-t? nisi admodum sero
legate-NOM be.PRS.INF how be.able-PLPRF.SBJV-3SG unless greatly late.ADV
praetor **est** **fact-us**, qu-od non arbitr-or.
praetor.NOM be.PRS.3SG become-NOM.M.SG which-ACC.N.SG not think-PRS.1SG
'How could he have been a legate fourteen years before he became praetor? Unless he became a praetor very late, which I don't think is the case.' (Cic. Att. 13.32.3)

- (2) Si L. Catilin-a [...] hac de re **pos-se-t**
if Lucius.NOM Catilina-NOM this.ABL about matter.ABL be.able-IPFV.SBJV-3SG
iudica-re, condemna-re-t L. Muren-am, si **interfic-ere**
judge-PRS.INF condemn-IPFV.SBJV-3SG Lucius.ACC Murena-ACC if murder-PRS.INF
pos-se-t, occid-ere-t.
be.able-IPFV.SBJV-3SG kill-IPFV.SBJV-3SG
'If Lucius Catilina could be the judge in this case, he would condemn Lucius Murena; if he were given the opportunity to murder him, he would indeed kill him.' (Cic. Mur. 83)

As is well known, in present day Romance varieties only the head-initial orders 'functional verb - lexical verb' are acceptable. Importantly, as was pointed out earlier in Danckaert (2016, 2017a,b), the transition from a mixed head-initial/head-final system to a strictly head-initial one is not always purely 'linear': in particular, in Late Latin (informally defined as the language used in Latin texts dating from *ca.* 200 - 600 AD) BE-periphrases of the type illustrated in (1) unexpectedly exhibit a very strong preference for the head-final order 'past participle - BE'.

The aim of this chapter is to revisit the relevant data, and to analyse them by referring to both the syntactic structure of the Late Latin clause and to the phonological properties of BE-auxiliaries. At a more general level I hope to show that at the very least in some cases, prosodic factors must be taken into account to correctly understand word order change. Crucially, in the wake of among others Taylor (2005), Speyer (2010) and Hinterhölzl and van Kemenade (2012), I do not want to say that word order change is to be explained in purely phonological terms: rather, it is important to pay attention to both syntax (phrase structure, constituency) and (prosodic) phonology.

The chapter is organised as follows. I will start by summarizing the main empirical facts as they emerge from my own corpus work (section 2). In section 3, I will present two pieces

of evidence which suggest that word order in Late Latin BE-periphrases is not purely a syntactic matter, but that it is also sensitive to phonological factors. In section 4 I will spell out my assumptions about Late Latin clause structure, and I will provide some background about how syntactic structures can be mapped onto prosodic representations. The actual analysis will be developed in sections 5 and 6. Section 7 is a brief conclusion.

2. Word order in the Latin clause: some basic generalizations

All corpus data to be presented in this paper are based on an analysis of 39 Latin text samples from a period of 800 years, starting with Plautus at around 200 BC, and ending with Gregory of Tours, who was active in the late sixth century AD. Apart from the comedies by Plautus and Terence, the corpus only consists of prose texts. A full description of the corpus can be found in Danckaert (2017a: 84-5).¹

2.1 The very basic picture

Let us then have a look at the VPAux/AuxVP-alternation in the history of Latin. In a nutshell, what emerges from the corpus is that combinations of a modal verb and an infinitive show the expected rise of head-initiality. In contrast, BE-periphrases behave very differently, in that they display a remarkably strong predilection for the order ‘PaPa - BE’ in Late Latin. However, this last observation does not hold in every single syntactic environment.

The data for the modal *possum* ‘CAN, be able’ are shown in Figure 1, where the values on the Y-axis correspond to the average frequencies of the head-initial order ‘modal - infinitive’ (in percentages).² As can be seen, despite there being a fair amount of synchronic variation throughout the entire period covered, both the straight and the smoothed regression lines reveal an unmistakable diachronic trend towards higher frequencies of the order ‘modal - infinitive’, with a rise from *ca.* 25% to well over 60%.

¹ For evidence that this corpus can reliably inform us about language variation and change, see Danckaert (2017a: 102-6, 282-5).

² As can be seen, only 37 samples were taken up in the data set plotted in Figure 1, namely those authors/texts for which I counted more than 10 combinations of *possum* and an infinitive. In this case, no data are represented for the texts of Anthimus (*ca.* 535 AD) and Antoninus of Piacenza (*ca.* 570 AD), which do not reach this threshold level. In the following graphs too, small subsamples are not represented.

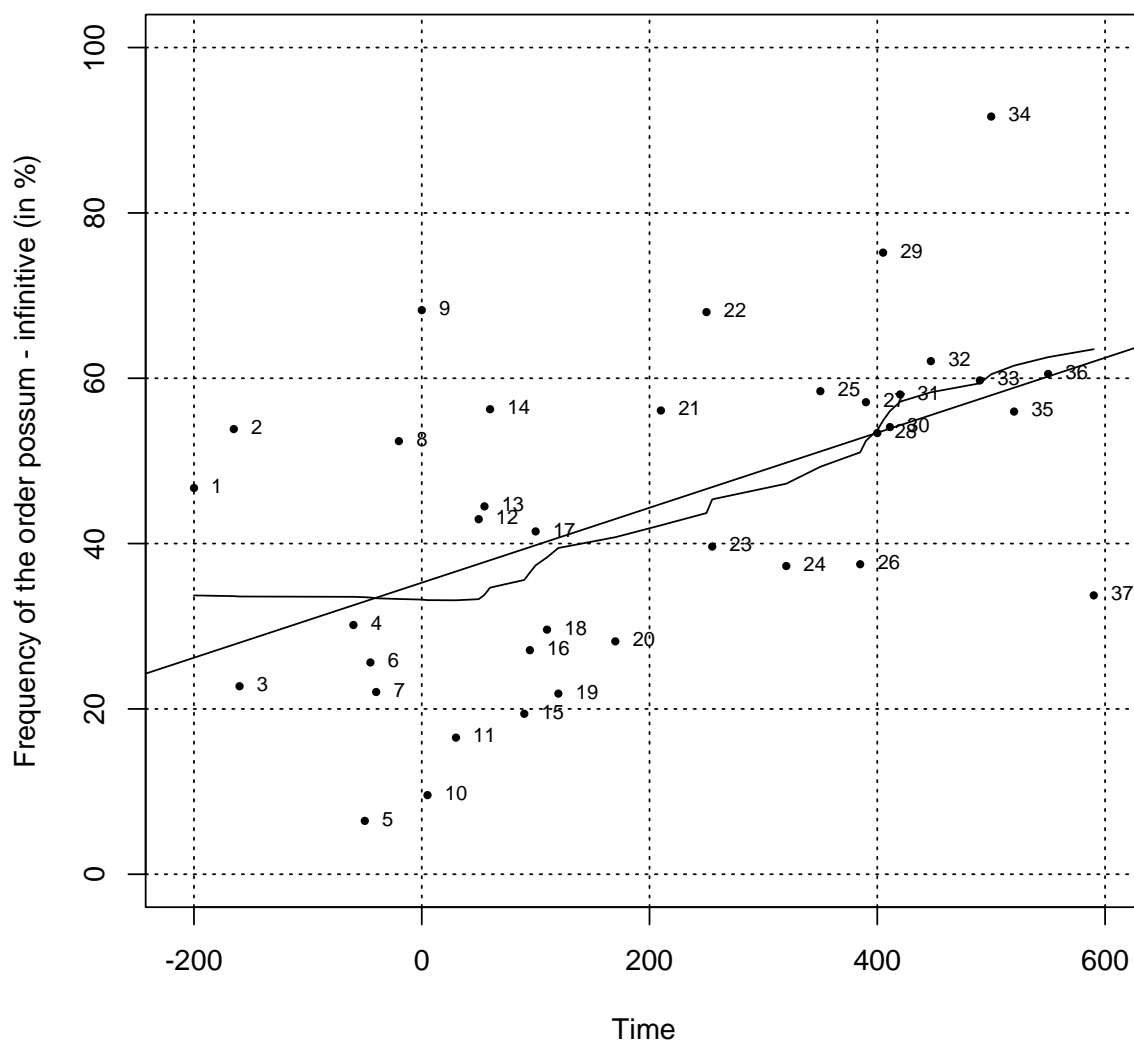


Figure 1: Incidence of the order '*possum* - infinitive', ca. 200 BC - 600 AD. Labels: 1= Plautus, 2= Terence, 3= Cato, 4= Cicero, 5= Caesar, 6= Varro, 7= Sallust, 8= Hyginus, 9= Vitruvius, 10= Livy, 11= Celsus, 12= Seneca, 13= Columella, 14= Petronius, 15= Frontinus, 16= Quintilian, 17= Pliny, 18= Tacitus, 19= Suetonius, 20= Gaius, 21= Tertullian, 22= Gargilius, 23= Cyprian, 24= SHA, 25= Palladius, 26= Egeria, 27= Jerome, 28= Augustine, 29= Vulgate, 30= GCC, 31= Vegetius, 32= Cassius Felix, 33= Victor of Vita, 34= Pompeius, 35= Caesarius, 36= Iordanes, 37= Gregory.

A similar picture is obtained with another modal verb, namely *debeo* 'MUST, have to', although it has to be said that here the diachronic trend towards more head-initiality is a bit weaker, witness the fact that the slope of the straight regression line is slightly less steep than that in Figure 1:

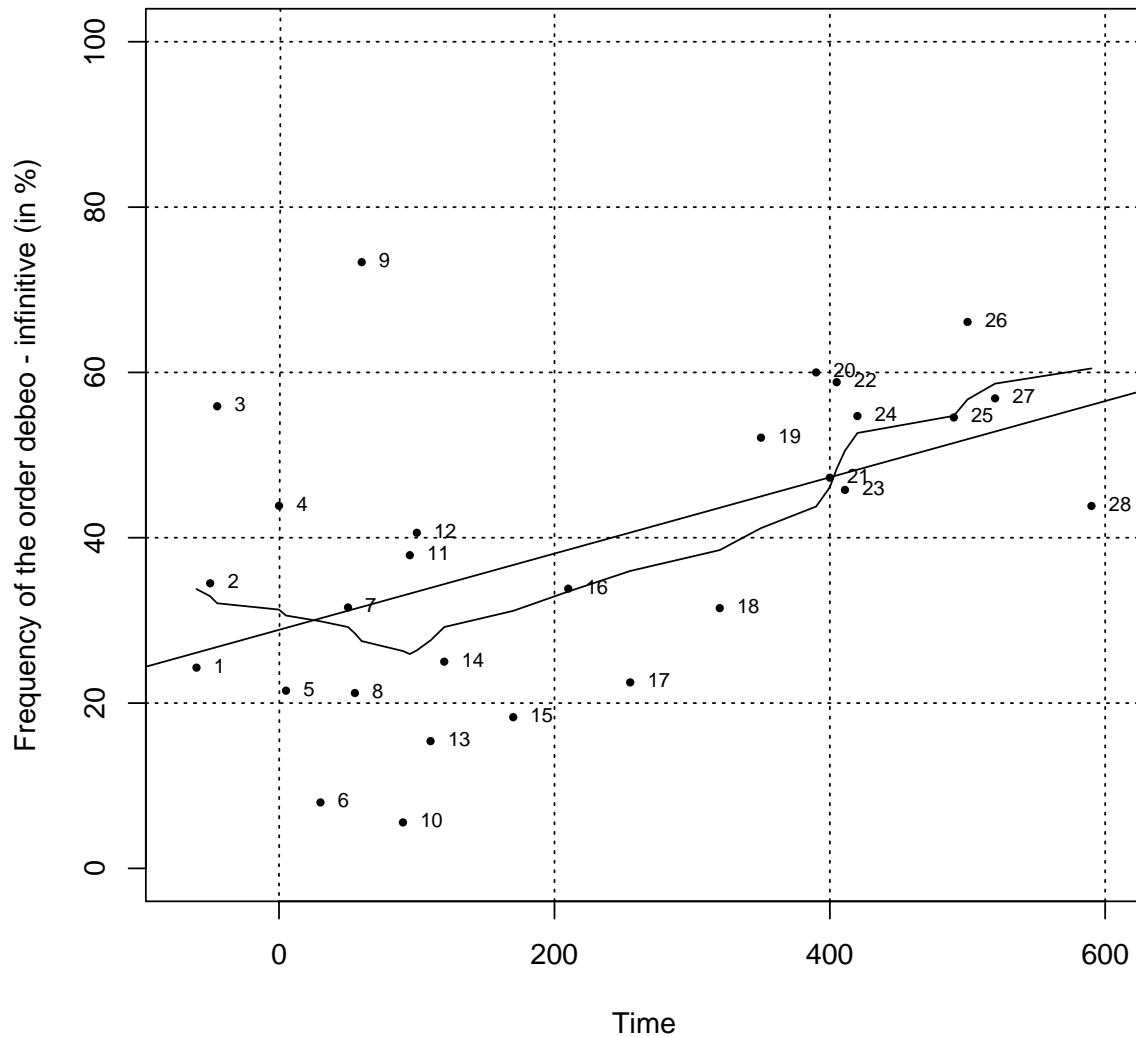


Figure 2: Incidence of the order ‘*debeo* - infinitive’, ca. 60 BC - 600 AD. Labels: 1= Cicero, 2= Caesar, 3= Varro, 4= Vitruvius, 5= Livy, 6= Celsus, 7= Seneca, 8= Columella, 9= Petronius, 10= Frontinus, 11= Quintilian, 12= Pliny, 13= Tacitus, 14= Suetonius, 15= Gaius, 16= Tertullian, 17= Cyprian, 18= SHA, 19= Palladius, 20= Jerome, 21= Augustine, 22= Vulgate, 23= GCC, 24= Vegetius, 25= Victor of Vita, 26= Pompeius, 27= Caesarius, 28= Gregory.

As mentioned, the corresponding data for combinations of *esse* ‘be’ and a past participle are very different.^{3,4} Before we inspect the actual data, it is important to point out that one has to

³ The relevant facts have only received scant attention in the philological literature, and many aspects of word order variability in Late Latin BE-periphrases, most notably the effect of negation, were only first discussed in Danckaert (2016). Väänänen (1987: 107) notes that *factum est* is much more frequent in Egeria’s *Itinerarium* than *est factum*, and he ascribes this to the influence of the language of the Latin Bible, where the same skewed distribution can be observed. However, it should be clear that this explanation does little more than ‘kick the can down the road’, as it begs the question as to why the word order facts from the Latin Bible are the way they are. The same author’s claim about word order in BE-periphrases in the works of Lucifer of Cagliari is not backed up with any evidence, and my own corpus data (as summarised in Figure 3) contain robust counterevidence against Väänänen’s assertion that *l’ordre interverti devient usuel* (“the inverted order (i.e. ‘BE - PaPa’, ld) becomes productive”) during the first centuries AD: rather, exactly the opposite is true, as in certain environments

distinguish two types of periphrastic BE-constructions, which can both involve a deponent or a passive past participle. The two types are illustrated in (3) (cf. the parts in boldface):

- (3) Et tumultu-s quidem Gallic-us et Ligustin-us, qu-i
 and unrest-NOM indeed Gallic-NOM and Ligurian-NOM which-NOM.M.SG
 principi-o eius ann-i **exort-us** **fu-era-t**,
 beginning-ABL that.GEN year-GEN arisen-NOM.M.SG be-PLPRF-3SG
 haud magn-o conatu breu-i **oppress-us** **era-t**.
 not great-ABL attempt.ABL short-ABL suppressed-NOM.M.SG be-IPFV-3SG
 ‘And indeed the Gallic and Ligurian revolt, which had started in the beginning of that
 year, was quickly suppressed with minimal effort.’ (Liv. 41.19.3)

The BE-periphrasis in the main clause, *oppressus erat* ‘was (lit. had been) suppressed’ is the older pattern, which I will call an ‘E-periphrasis’. This type of formation is characterised by what we could call a ‘tense mismatch’: although the entire complex expression is a pluperfect, the BE-auxiliary in isolation is imperfective (as indicated by the gloss). In contrast, in the (deponent) expression *exortus fuerat* ‘had arisen’ (an ‘F-periphrasis’) both the complex unit and the bare auxiliary are pluperfects. As mentioned in Danckaert (2017a: 273-6), it is not clear whether apart from being morphologically different, the E and the F-pattern also differ semantically.

If we then look at the development of word order in E-periphrases, we observe that in strong contrast with modal verbs and their infinitival complements, the head-final pattern is strongly preferred by every single Late Latin author:

this order becomes rarer over time. Salvi (2004: 93, fn. 28) speculates that the preponderance of the order ‘PaPa - BE’ is restricted to the ‘most vulgar’ Late Latin texts (which is incorrect), and that in those texts, the lexical verb and the BE-auxiliary were reanalysed as a single (‘synthetic’) morphosyntactic unit. As pointed out in Danckaert (2017b: 231-3), this last hypothesis cannot be correct in the general case, as it makes incorrect predictions for word order in Late Latin negated BE-periphrases; there is however some reason to assume that it may be the correct analysis for some BE-periphrases in the Vulgate (Danckaert 2017c: 145-8).

⁴ Throughout this paper, I will not make a distinction between passive and deponent BE-periphrases: as argued at length in Danckaert (2016: 142-5; 2017b: 226-9), there are no (empirical) reasons to treat the two separately, neither when it comes to analysing word order variability, nor when modeling the alternation between ‘E-periphrases’ and ‘F-periphrases’ (on which, see below).

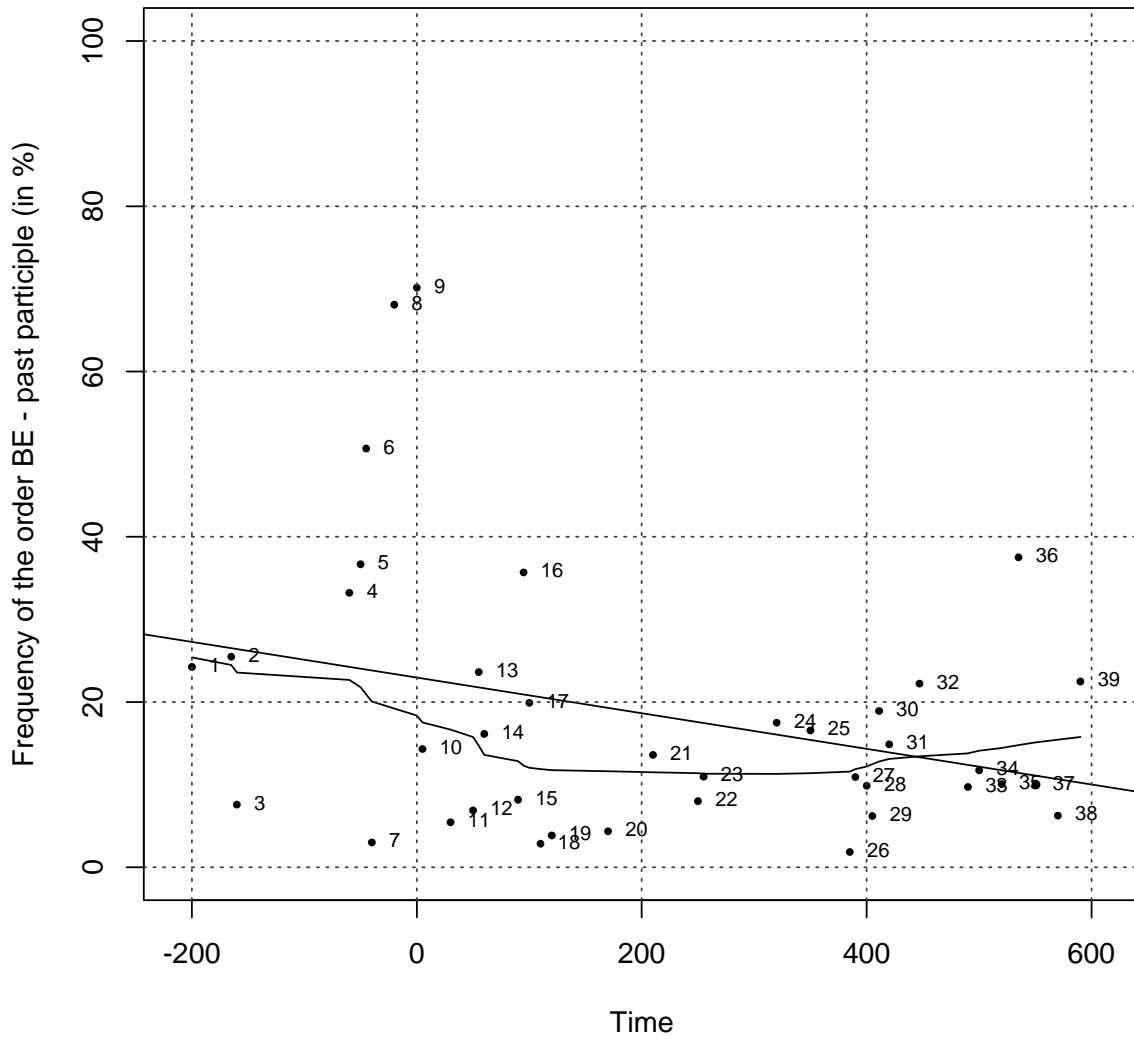


Figure 3: Incidence of the order ‘BE - PaPa’ in E-periphrases, *ca.* 200 BC - 600 AD. Labels: 1= Plautus, 2= Terence, 3= Cato, 4= Cicero, 5= Caesar, 6= Varro, 7= Sallust, 8= Hyginus, 9= Vitruvius, 10= Livy, 11= Celsus, 12= Seneca, 13= Columella, 14= Petronius, 15= Frontinus, 16= Quintilian, 17= Pliny, 18= Tacitus, 19= Suetonius, 20= Gaius, 21= Tertullian, 22= Gargilius, 23= Cyprian, 24= SHA, 25= Palladius, 26= Egeria, 27= Jerome, 28= Augustine, 29= Vulgate, 30= GCC, 31= Vegetius, 32= Cassius Felix, 33= Victor, 34= Pompeius, 35= Caesarius, 36= Anthimus, 37= Iordanes, 38= Antoninus, 39= Gregory.

Interestingly, there is a strong contrast between E and F-periphrases: as shown in Figure 4, in Late Latin the latter do not uniformly prefer the head-final order. Instead, we see that there is a lot of variation between individual authors, with average frequencies after 200 AD ranging from 0% (Anthimus at 535 AD) to 83.13% (Cassius Felix at 447 AD). Regardless of the correct interpretation of the variation observed in Figure 4, it should be clear that on the whole, Late Latin E and F-periphrases have very different word order preferences.

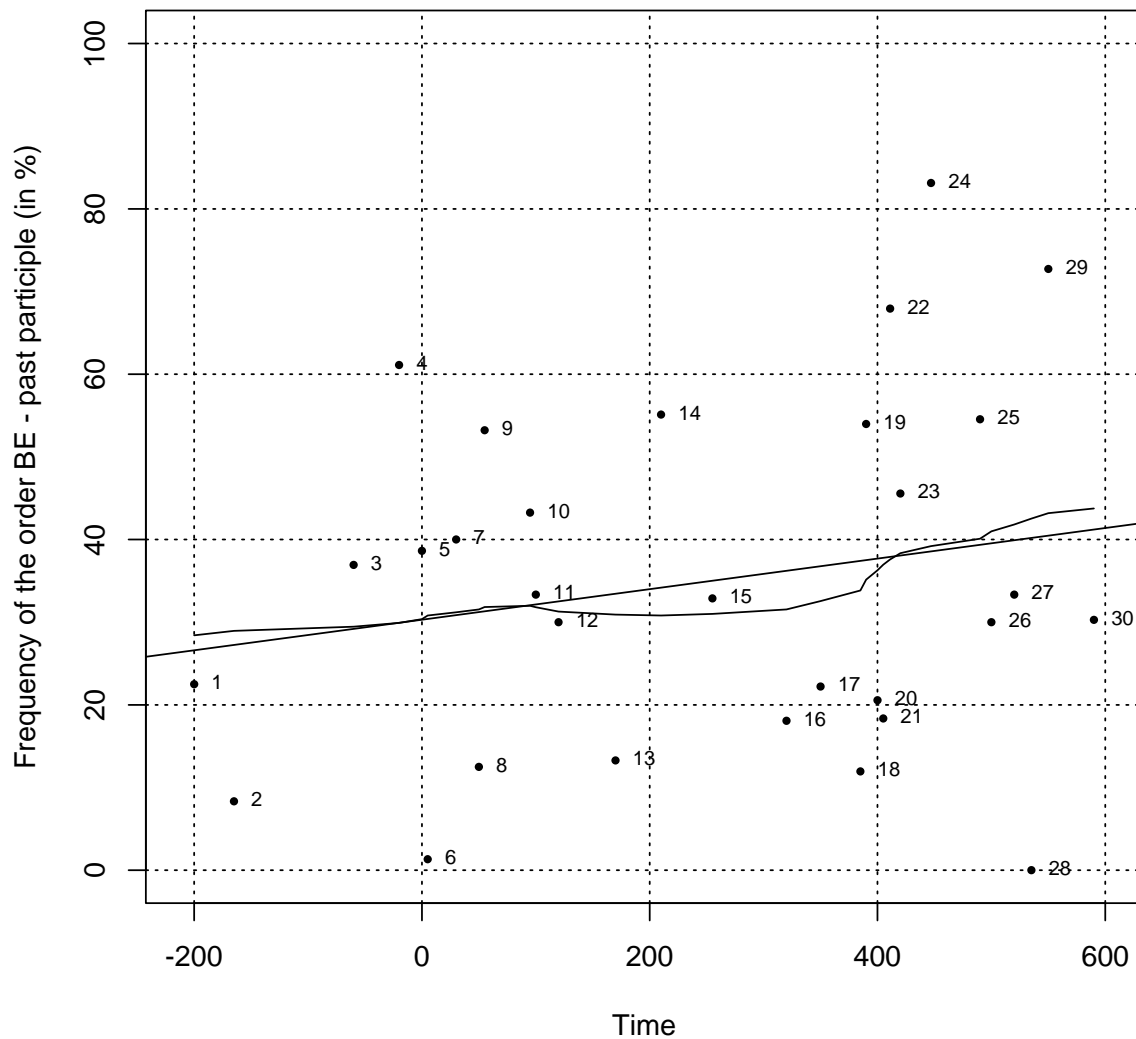


Figure 4: Incidence of the order 'BE - PaPa' in F-periphrases, *ca.* 200 BC - 600 AD. Labels: 1= Plautus, 2= Terence, 3= Cicero, 4= Hyginus, 5= Vitruvius, 6= Livy, 7= Celsus, 8= Seneca, 9= Columella, 10= Quintilian, 11= Pliny, 12= Suetonius, 13= Gaius, 14= Tertullian, 15= Cyprian, 16= SHA, 17= Palladius, 18= Egeria, 19= Jerome, 20= Augustine, 21= Vulgate, 22= GCC, 23= Vegetius, 24= Cassius Felix, 25= Victor, 26= Pompeius, 27= Caesarius, 28= Anthimus, 29= Iordanes, 30= Gregory.

2.2 The effect of negation

One very important additional observation concerns the influence of negation on word order in E-periphrases. As shown in Figure 5, only those E-periphrases which occur in an affirmative clause prefer to be head-final. In negated contexts the opposite tendency is observed, with all authors after 400 AD strongly preferring the head-initial order (only authors/texts with more than 8 negated E-periphrases taken into account):

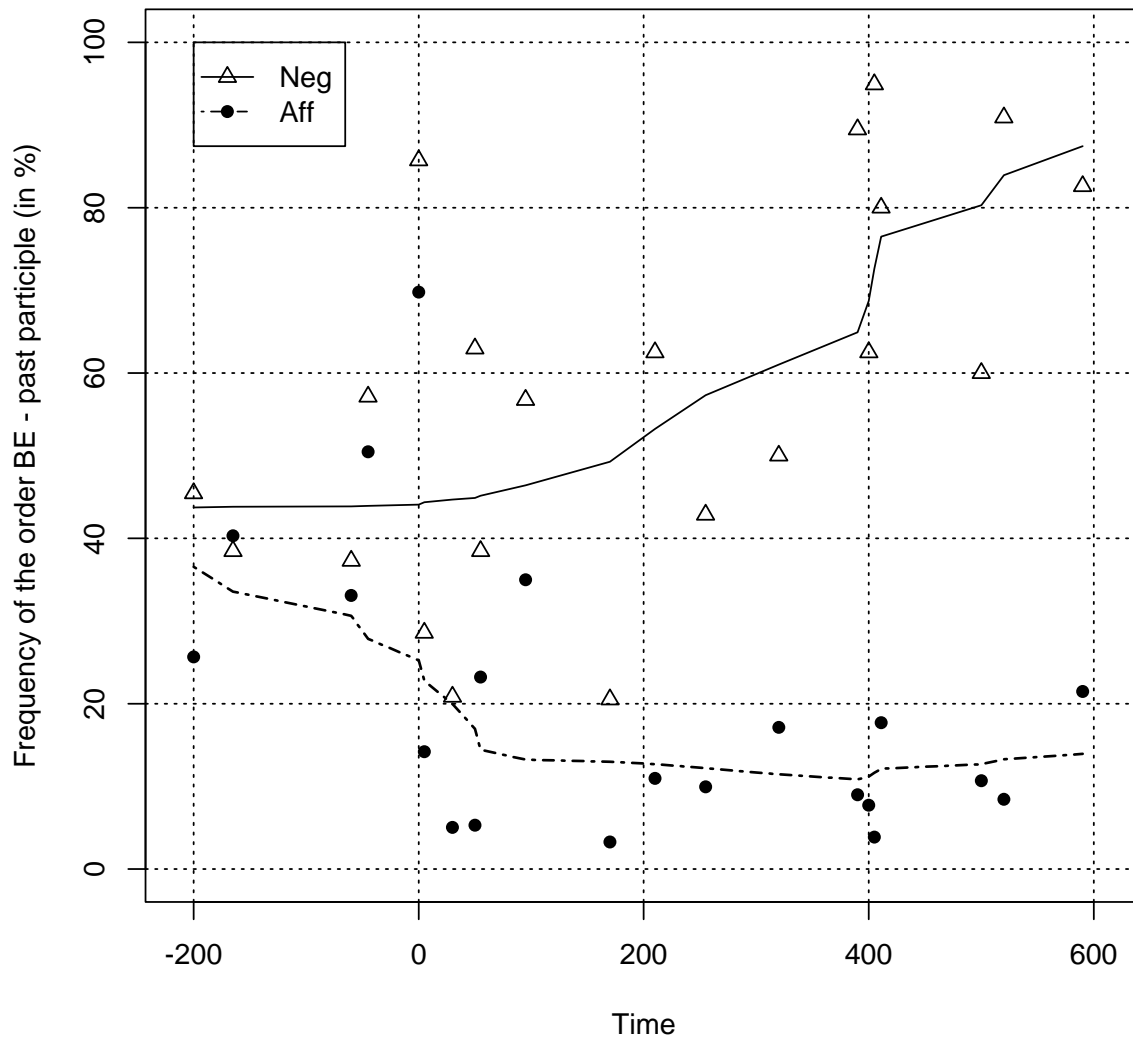


Figure 5: Incidence of the order ‘BE - PaPa’ in E-periphrases, *ca.* 200 BC - 600 AD: affirmative and negated clauses compared.

It is worth pointing out that there is no comparable effect of negation on word order in clauses with a modal verb and an infinitive: the relevant data are shown in Figure 6 (where I am averaging over clauses with *possum* ‘CAN, be able’ and *debeo* ‘MUST, have to’). As can be seen, this variable seems to have the opposite effect it has in E-periphrases, the head-final order ‘infinitive - modal’ being slightly more frequent in negated clauses than in affirmative ones in Late Latin (but note that it remains to be tested whether this effect is indeed statistically significant).⁵

⁵ The effect of negation on word order in F-periphrases is not easy to evaluate, due to the fact that negated F-periphrases are overall quite rare. For the sake of completeness, here are some data. In the period from 200 to 600 AD, I counted 1382 F-periphrases unaccompanied by *non*; 487 of those are head-initial (i.e. 35.24%). Out of 69 F-periphrases negated by *non*, 31 are head-initial (44.93%). This difference in absolute terms is not statistically significant (Fisher’s exact test, $p = .1218$).

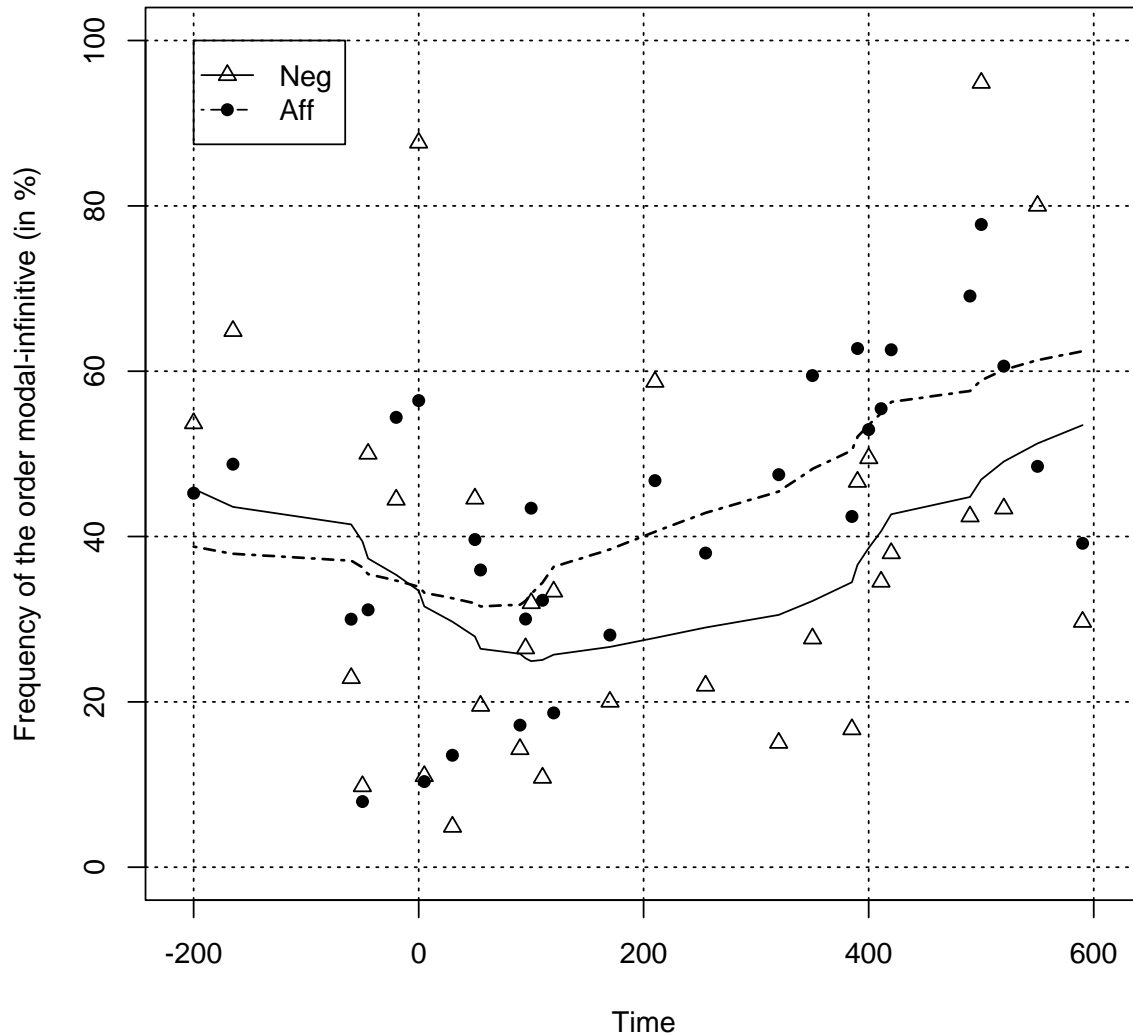


Figure 6: Incidence of the order ‘modal - infinitive’, *ca.* 200 BC - 600 AD: affirmative and negated clauses compared.

The two main *explananda* of this paper can then be summarised as follows. First of all, why do Late Latin E-periphrases prefer the order ‘PaPa - BE’ (Figure 3), a state of affairs which is all the more puzzling given that no trace of this word order pattern is left in modern Romance? Second, why is this preference reversed in negated clauses (Figure 5)?

The analysis that I will develop makes crucial reference to both syntactic and prosodic factors. In the next section, I will first provide evidence that phonology plays a role in bringing about word order in BE-periphrases, and that it is indeed possible to assess the role of prosodic factors on the basis of data from a written corpus.

3. The role of prosodic factors

3.1 Syllable count

My first argument in support of the claim that word order in Late Latin E-periphrases is sensitive to prosodic factors is the observation that the variable ‘number of syllables of the auxiliary’ is a good (and statistically significant, cf. Danckaert 2016: 160) predictor for word order. Specifically, it is only E-periphrases with a monosyllabic auxiliary which in Late Latin

display the observed strong tendency to follow the participle they are associated with.⁶ The data for the monosyllabic auxiliaries are shown in Figure 7. Note in passing that the overall resemblance between this graph and the one in Figure 3 is to be ascribed to the fact that the 8 monosyllabic forms of Latin *esse* (namely the present indicatives *sum*, *es*, *est* and *sunt*, and the present subjunctives *sim*, *sis*, *sit* and *sint*) happen to have a very high (combined) token frequency: with 20092 occurrences in the entire corpus, monosyllabic BE-auxiliaries clearly outnumber their polysyllabic counterparts (5153 tokens).

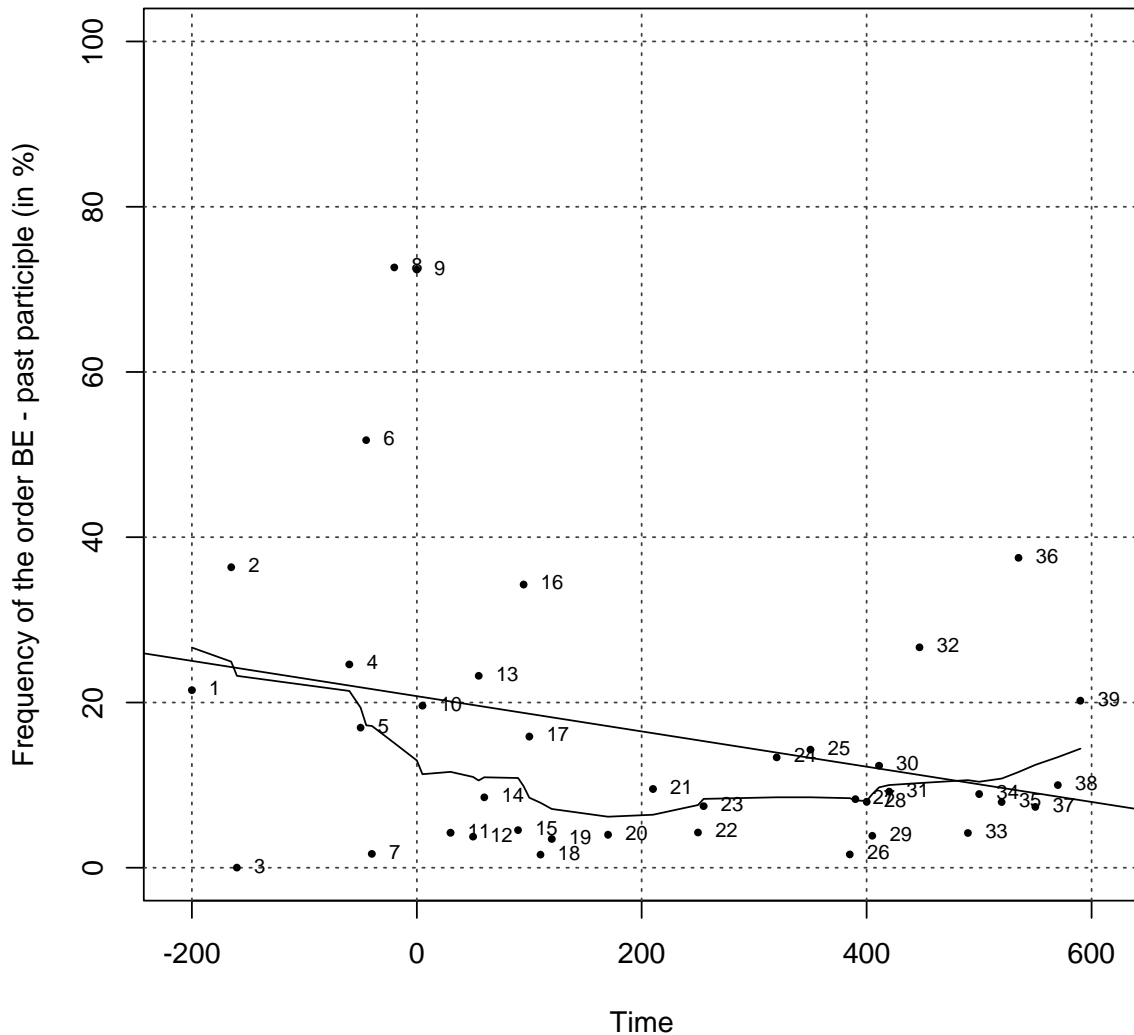


Figure 7: Incidence of the order ‘monosyllabic BE-auxiliary - PaPa’, *ca.* 200 BC - 600 AD. Labels: see Figure 3.

In contrast, as can be seen in Figure 8, with the exception of one data point after 200 AD, E-periphrases with a polysyllabic (and typically bisyllabic) auxiliary quite frequently appear in the head-initial order (in on average around 35% of the cases; only authors/texts with more than 8 polysyllabic E-auxiliaries taken into account). The almost entirely flat trajectory of the straight regression line suggests that word order variation in this data set is diachronically

⁶ Note that this observation does not entail that the abovementioned contrast between word order in (Late Latin) E and F-periphrases (cf. Figures 3 and 4) can be reduced to the fact that there are no monosyllabic auxiliaries of the F-type: see Danckaert (2016: 158-61) for statistical evidence corroborating this point.

stable, but given the considerable amount of synchronic variation throughout the entire period covered, it should be clear that it is very difficult to interpret what exactly is going on here. In any event, the contrast with the monosyllabic auxiliaries is unmistakable.

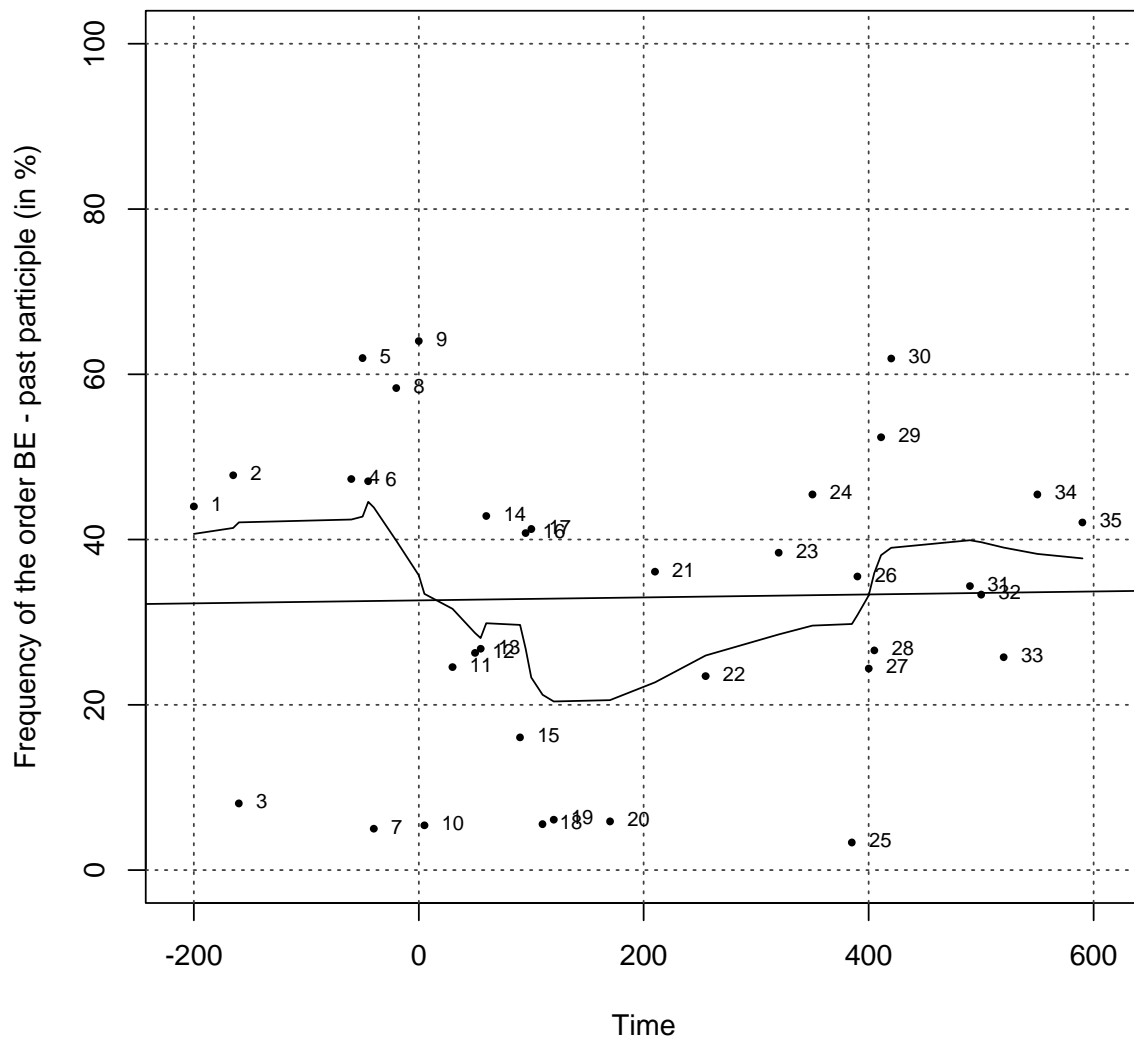


Figure 8: Incidence of the order ‘polysyllabic BE-auxiliary - PaPa’ (E-periphrases only), *ca.* 200 BC - 600 AD. Labels: 1= Plautus, 2= Terence, 3= Cato, 4= Cicero, 5= Caesar, 6= Varro, 7= Sallust, 8= Hyginus, 9= Vitruvius, 10= Livy, 11= Celsus, 12= Seneca, 13= Columella, 14= Petronius, 15= Frontinus, 16= Quintilian, 17= Pliny, 18= Tacitus, 19= Suetonius, 20= Gaius, 21= Tertullian, 22= Cyprian, 23= SHA, 24= Palladius, 25= Egeria, 26= Jerome, 27= Augustine, 28= Vulgate, 29= GCC, 30= Vegetius, 31= Victor, 32= Pompeius, 33= Caesarius, 34= Iordanes, 35= Gregory.

Assuming that the syntactic component of the grammar cannot count syllables, but that in contrast phonological rules or constraints can in fact be sensitive to the number of syllables of a given lexical item, we can conclude that we have a first piece of evidence in support of the claim that word order in Late Latin E-periphrases is at least partly a matter of phonology.

3.2 Evidence from prose rhythm

A second argument that goes in the same direction comes from the philological literature on prose rhythm, and in particular a phenomenon that was recently revisited in Holmes (2017).

As is well known, many Latin prose authors adhered to a number of conventions of ‘clausula formation’, which we can define as the preference to map the end of a sentence, and – less systematically – of longer sentence-internal phrases, onto certain rhythmical patterns (i.e. sequences of heavy and light syllables), and to avoid certain such sequences in the same positions.⁷

Table 1: (Some) preferred clausulae in Late Latin rhythmical prose.

Rhythmical pattern	Frequency
— ∪ . — — x	extremely common
— ∪ . — ∪ ∪ x	very common
´ x x . — ∪ — x	very common

Two examples instantiating the first type of clausula mentioned in Table 1 are given in (4), both from Boethius' *Philosophiae Consolatio* (ca. 524 AD):

- (4) a. $\text{ad super-ius element-um es-se-t ascensu-s}$
to high-COMP.ACC character-ACC be-IPFV.SBJV-3SG way.up-NOM
‘there would be a way up to the higher character’ (Boeth. Cons. 1.1)
- b. $\text{interi-us autem recept-a dulcesc-a-nt}$
interior-ADV PRT received-NOM.N.PL sweeten-PRS.SBJV-3PL
‘but once ingested they become sweet’ (Boeth. Cons. 3.1)

One very important observation concerns the behavior of forms of non-copular *esse* ‘be’ in Late Latin rhythmic prose. As was noted at least as early as Havet (1892), clause-final monosyllabic forms of *esse* can appear after the end of a clausula, apparently not taken into

⁷ For a general overview of the literature on Latin prose rhythm, see Oberhelman (2003). As pointed out by this author, other aspects of the rhythmical organisation of literary prose can be subsumed under the remit of ‘prose rhythm’ too, most notably the segmentation of long stretches of text into smaller prosodic units, the so-called ‘prose cola’ studied by among others Fraenkel (1932-’33) and Habinek (1985). Note however that the data discussed in this strand of the literature have never been revisited in the light of the modern literature on prosodic phonology to be discussed in section 4.2 below.

account for metrical (rhythmical) purposes (Havet 1892: 66-74; Di Capua 1914: 295, 1937: 21 n. 5; Hagendahl 1937: 96-9; Holmes 2017).⁹ Thus Havet (1892: 66-7):

‘Ordinairement, le monosyllabe *est*, *sunt*, etc., semble être censé nul, c’est à dire que d’une part la nature de la finale placée en contact avec lui paraît être indifférente (non seulement il est indifférent qu’elle soit longue ou brève [footnote omitted ld], mais encore, quand elle se termine par une voyelle ou une *m*, il est indifférent que le monosyllabe commence par une voyelle ou une consonne), et que d’autre part le mot pénultième agit, sur le précédent, comme il agirait s’il était final.’¹⁰

A similar remark is made in Di Capua (1914: 295):

‘Le voci monosillabe del verbo *sum*, poste in fine di clausola, non contano; e ciò sia che incomincino per vocale, sia per consonante.’¹¹

Di Capua (1914: 295) gives the following examples from Boethius’ *Consolatio* (with again the same type of clausula as in (4)):

- (5) a. — ∪ — — x ∥
poll-u-isse mentit-i sunt
soil-PRF-INF lied-NOM.M.PL be.PRS.3PL
‘they lied that [I] violated [my conscience]’ (Boeth. Cons. 1.4)
- b. — ∪ — — x ∥
felicitat-is oblit-us es
happiness-GEN forgotten-NOM.M.SG be.PRS.2SG
‘you have forgotten [your] happiness’ (Boeth. Cons. 2.3)

Importantly, practices in earlier authors seem to have been different, although it has to be added that the matter is not entirely straightforward to evaluate, mainly because many sequences ending with a monosyllabic form of *esse* constitute an acceptable clausula regardless of whether the final monosyllable is taken into account or not. Nevertheless, on the basis of small corpus study of Cicero, Bornecque (1907: 226-8, emphasis in original omitted) concludes that it is at the very least not the rule to disregard final forms of *esse*:

‘Dans 67 cas, la forme monosyllabique peut être comptée ou non. [...] Dans 37 cas, la forme monosyllabique doit être comptée pour une syllabe. [...] En résumé, il y a tendance à compter ces formes monosyllabiques.’¹²

⁹ But as pointed out in Hagendahl (1937: 96) (in his discussion of prose rhythm in Arnobius the Elder (*ca.* 310 AD)), instances of *est* and *es* (i.e. the second and third person singular forms of the present indicative) inside a clausula have to be retained in full in order to obtain a good clausula. Anticipating the upcoming discussion, we can assume that in such cases we are dealing with what I will refer to below as ‘strong BE’ (i.e. an instance of copular BE, or a remnant of the auxiliary used in earlier E-periphrases).

¹⁰ ‘Normally the monosyllables *est*, *sunt*, and the like seem to be considered null. On the one hand, the nature of the final syllable in contact with such a form seems not to matter: it is not only immaterial whether this final syllable is long or short [footnote omitted ld], but also, when it ends with a vowel or an *m*, it does not matter whether the <following> monosyllable starts with a vowel or with a consonant. On the other hand, the penultimate words behaves with respect to the preceding one as if it were in final position.’

¹¹ ‘The monosyllabic forms of the verb *sum* do not count when placed at the end of a clausula, regardless of whether they start with a vowel or with a consonant.’

This second piece of evidence confirms that especially monosyllabic forms of Late Latin *esse* display some idiosyncratic behavior which is difficult to account for if we were to assume that word order in E-periphrases is only a matter of syntax (say, variable headedness of the T-node).

4. Some background assumptions

In this section I will lay the foundations for the upcoming analysis: I will start by spelling out my assumptions about Late Latin clause structure, and then I will elaborate on how a prosodic representation can be derived from a syntactic tree. In section 5 I will then put this system to use to explain the patterns described in section 2.

The analysis of Late Latin clause structure that I will assume here is entirely based on the discussion offered in Danckaert (2017a: chapter 5). In this piece of work, it is argued that at some point during the first century AD, a new grammar (called ‘Grammar B’ for simplicity) becomes productive. This grammar gradually takes over from the older ‘Grammar A’, which was mainly productive in what is traditionally referred to as the Archaic and Classical periods (say 200 BC - 100 AD). Although Grammar A obviously did not disappear overnight, I will here make the simplifying assumption that the grammar we see at work in Latin texts from *ca.* 200 - 600 AD is indeed Grammar B.

4.1 Late Latin clause structure

I will take it that a clause generated by Grammar B minimally consists of the series of functional heads given in (6) (where I am abstracting away from left peripheral heads dominating SubjP):

$$(6) \quad \dots \text{Subj}^\circ > \text{F}^\circ_{[\text{EPP}]} > \text{Neg}^\circ > \text{T}^\circ > \text{Voice}^\circ > \nu^\circ > \sqrt{}^\circ$$

Let me briefly comment upon this fragment of the functional sequence. The lowest three heads constitute a three-tiered verb phrase, with an a-caterogial root at the bottom, a verbalizing head ν right above it, and a Voice-head on top of that. In this articulated thematic domain, an internal argument (if present) is merged in the complement position of $\sqrt{}^\circ$; external arguments are first merged in SpecVoiceP. Moving then to the articulated inflectional layer, we first encounter T, which I take to be the base position of BE-auxiliaries, both in passive and in deponent contexts. Right above T sits Neg: importantly for present purposes, whenever this head is lexicalised by *non*, in Grammar B it always forms a complex head with T (Danckaert 2017a: 249-54). In particular, T head moves to Neg and right adjoins to it, and as a result, negation surfaces as a proclitic to the hierarchically highest verb in the clause. This state of affairs is preserved in all present day Romance languages in which some descendant of Latin *non* is still used as the canonical marker of sentential negation.

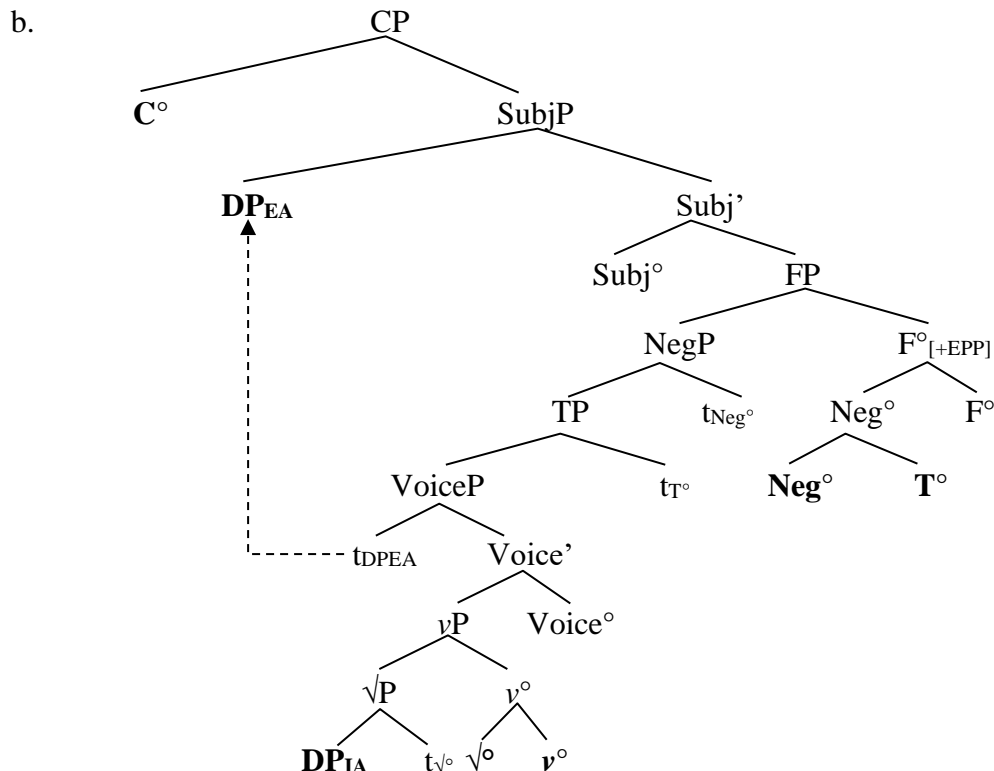
Finally, there are two functional heads which in a broad sense can be said to be associated with ‘subjecthood’. The first is F; as indicated, I take this head to be endowed with an EPP-feature, which entails that F(P) needs to be lexicalised overtly by some ϕ -feature bearing category. In the wake of among others Anagnostopoulou and Alexiadou (1998) and Biberauer (2003), I adopt the idea that the languages of the world differ as to the way in which this lexicalization process comes about: in a nutshell, some languages employ head

¹² ‘In 67 cases the monosyllabic form can be counted or not. [...] In 37 cases, the monosyllabic form has to be counted as a syllable. [...] In sum, there is a tendency to take these monosyllabic forms into account.’

movement to F° , whereas other languages merge a phrasal category in SpecFP. In Grammar B the former option is selected, yielding movement of the finite verb of a given clause to F, with or without a proclitic negator. Finally, subject arguments can optionally evacuate the thematic layer and move to the specifier of a dedicated subject position SubjP. Under this analysis, EPP satisfaction and presence or absence of A-movement are logically distinct (on this last point, see in particular Danckaert 2017c). All this is brought together in the tree in (7b) (overt terminals in bold), which details the structure of a Late Latin clause like (7a), featuring a negated (and in this case head-final) BE-periphrasis (and the overall word order ‘C-SOV-Neg-Aux’).¹³

- (7) a. cum Christ-us [Thoma-m apostol-um dubita-nt-em
 although Christ-NOM Thomas-ACC apostle-ACC doubt-PRS.PTCP-ACC
 de se] aspernat-us non si-t
 about REFL.ABL scorned-NOM.M.SG not be.PRS.SBJV-3SG
 ‘although Christ did not reject the apostle Thomas, who had doubts about him’
 (Aug. c. Faust. 16.8, *ca.* 400 AD)

¹³ The abbreviations ‘EA’ and ‘IA’ used in the tree in (7b) stand for ‘external argument’ and ‘internal argument’ respectively.



Let me conclude this section with a note on word order, or more precisely the linearization of syntactic structures. As indicated in the tree above, I take it that complement-head sequences do not have to be derived through movement (along the lines of Kayne 1994), but rather that they can be base-generated as such. Although this way of representing things may well be a simplification¹⁴, one major advantage of assuming variable headedness in the base is that it allows one to derive prosodic representations from syntactic trees in a much more straightforward fashion than if one were to assume antisymmetry.

In any event, it should be noted that head-final OVAux orders are not the only ones that are compatible with Grammar B. Rather, all nodes lower than and including F can be head-initial, as long as the ‘Final-Over-Final Constraint’ (FOFC) from Biberauer, Holmberg and Roberts (2014) is respected. Simply put, FOFC says that within one and the same extended projection, a head-final projection can only dominate other head-final projections. As documented in Danckaert (2017a: chapter 4), evidence for the claim that FOFC is at work in the hypothesised Grammar B comes from the observation that the word order ‘VOAux’ (i.e. a head-initial verb phrase dominated by a head-final TP (or FP)) is only very rarely attested in Late Latin texts.

4.2 From syntax to prosody

Let us consider how a syntactic structure like (7b) can be mapped onto a prosodic representation.¹⁵ In the wake of pioneering work by Selkirk (1980, 1981, 1986) and Nespor

¹⁴ For an antisymmetric approach to Latin word order, see again Danckaert (2017a). As mentioned there, one advantage of deriving complement-head sequences through syntactic movement is that it allows one to account for the decline of the VOAux order along the lines of Biberauer, Holmberg and Roberts (2014).

¹⁵ For additional discussion of prosodic constituency in Latin (albeit from a different perspective from the one adopted here), see Agbayani and Golston (2016).

and Vogel (2007 [1986]))¹⁶, I will assume that (phrasal) syntactic constituents of various sizes are mapped onto prosodic constituents of various sizes. Conversely, as a rule strings of words which do not form a syntactic constituent do not correspond to a prosodic constituent either (that is, unless some sort of ‘restructuring’ process applies, cf. Nespor and Vogel 2007 [1986]: *passim*).

The inventory of possible prosodic constituents is standardly summarised in what is known as the ‘prosodic hierarchy’. The largest constituent in this hierarchy is the utterance (U), which consists of one or more intonational phrases (I). These in turn are made up of phonological phrases (ϕ), which contain prosodic words (ω).¹⁷ Lower down we find metrical feet (F), and finally syllables (σ).¹⁸ By default, (non-deficient) morphological words are mapped onto prosodic words, syntactic phrases (simplex one such as DPs, PPs, or more complex ones like VPs (or higher nodes in the extended projection of a lexical verb)) surface as phonological phrases (see especially Truckenbrodt 1999), and clauses are mapped onto intonational phrases (see e.g. Selkirk 2011: 439 for a statement along these lines). Importantly, these rules are certainly not inviolable: rather, the exact prosodic shape of a given syntactic unit is governed by a number of factors, such as the length and/or internal complexity of a given XP, but also for example the speech rate in actual production.

As we will see in section 5, the representations that can be built with these prosodic categories resemble syntactic trees in that they are (rigidly) hierarchically organised, but they differ from them (i) in that they allow for non-binary branching nodes (which in syntax is not standardly allowed under current assumptions), and (ii) in that a given prosodic category can only dominate categories that occur lower in the prosodic hierarchy (i.e. unlike in syntax, in phonology there is no such thing as recursive embedding). As originally proposed in Selkirk (1981), this last point can be said to follow from of a stronger constraint known as ‘strict layering’. In the words of Selkirk (2011: 437):

(8) The strict layer hypothesis

A constituent of category-level n in the prosodic hierarchy immediately dominates only a (sequence of) constituents at category-level $n-1$ in the hierarchy.

In other words, intonational phrases immediately dominate only phonological phrases, which dominate only immediately prosodic words, and so on. As we will see immediately, the consequences of the strict layer hypothesis will play an important role in the upcoming analysis.

5. A lexical split: strong and weak BE

We can now return to the empirical data at issue, namely variable word order in Late Latin BE-periphrases. The core proposal is that at a certain point in the history of the language (presumably at the beginning of the common era, although the actual point of actuation is of

¹⁶ For more recent overviews of the state-of-the-art in the field of prosodic phonology, see Truckenbrodt (2007), Elordieta (2008) and Selkirk (2011).

¹⁷ Note that I am omitting the so-called ‘clitic group’ from this list of prosodic categories, which in Nespor and Vogel (2007 [1986]) is situated between the phonological phrase and the prosodic word. The status of this intermediate level is debated.

¹⁸ Below the level of the syllable one can furthermore distinguish morae (μ) and segments (S), which I will not further take into account here.

course difficult to pin down), a lexical split took place giving rise to the creation of two lexical items *esse*, which I will refer to as ‘weak BE’ and ‘strong BE’.

Weak BE is found in E-periphrases, and probably also in combination with a gerundive, as well as in expressions like *opus est* ‘need, have want’ and *necesse est* ‘be necessary’.¹⁹ Despite being the innovative variant, weak BE is used in environments which do not survive in Romance, and it does also not itself live on in the Romance languages.²⁰ In contrast, strong BE (which is a continuation of ‘old BE’) is found in various types of copular clauses, and in F-periphrases; it is also the ancestor of all Romance BE-verbs that derive from Latin *esse* (auxiliaries and copulae alike).²¹

As mentioned, I take it that the difference between strong and weak BE is specified in the lexicon. The core difference between the two variants concerns their respective phonological shape: weak but not strong BE is phonologically deficient. More precisely, as a first approximation to be refined below, we can say that weak BE constitutes a foot (F) that cannot project a prosodic word (ω) of its own, unless it is sufficiently heavy, that is, at least bisyllabic. For the time being, I will concentrate on monosyllabic forms of *esse*, and I will return to polysyllabic forms in section 6.4. Importantly, if we want to adhere to the ‘strict layer hypothesis’ spelled out in (8), the question now arises as to how a ‘stray foot’ can be incorporated in a larger prosodic representation, containing among other things phonological and intonational phrases, which by assumption cannot directly dominate anything else than prosodic words and phonological phrases respectively. In order to avoid this potential problem, we can assume that weak BE is subject to the constraint in (9).^{22,23}

¹⁹ As pointed out in Danckaert (2017a: 287-8), there are reasons to assume that the Late Latin lexical entry ‘weak BE’ hypothesised here is unrelated to the phenomenon of prodelision (or aphaeresis) with BE (Boldrini 1999: 54-5; Pezzini 2015). Specifically, the latter phenomenon exclusively affects forms of *esse* with a vocalic onset, whereas the word order behavior of Late Latin weak BE is only to a small extent sensitive to the nature of the first segment (vocalic or consonantal) of a given form of *esse*. Note also that the Late Latin word order data discussed here have nothing to do with the putative cases of second position cliticisation of *esse* discussed in Adams (1994) (on which, see Danckaert 2017a: 288-9).

²⁰ For a series of arguments that Latin E-periphrases are not the historical source of Romance analytic present tense passives of the type *sono amato* ‘I am (being) loved’, see Danckaert (2016, 2017b).

²¹ This proposal predicts that copular *esse* should not display the type of (word order) behavior reviewed in sections 2 and 3: for example, we predict the word order pattern ‘subject - *esse* - predicate’ to be fully productive in Late Latin. In future work I hope to test whether this prediction is borne out.

²² On foot extrametricality, see also Buckley (1994) and Hayes (1995: 56-60, 105-8). Compare Hyde (2011: 1027) on extrametricality more generally (emphasis mine):

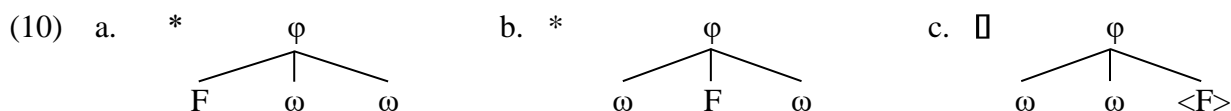
‘The result of extrametricality is essentially invisibility to the application of subsequent rules. When a constituent is designated as extrametrical, it is excluded from the domain of rules that might incorporate it into higher levels of prosodic structure. An extrametrical segment cannot be associated with a mora; for example, an extrametrical syllable cannot be footed, and an *extrametrical foot cannot be included in a prosodic word*.’

²³ One interesting question is why the constraint in (9) makes explicit reference to the right edge of phonological phrases, or put differently, why weak BE cannot be extrametrical when it sits at the left edge of a p-phrase. As pointed out in Hyde (2011), cross-linguistically there is a general – but no absolute – tendency for extrametricality to obtain at the right edge of

(9) PF-constraint on placement of weak BE

Weak BE can only survive as a single metrical foot at the level of the prosodic word if it appears at the right edge of a phonological phrase, where it counts as ‘extrametrical’.

To make this more concrete, we can say that a prosodic representation involving a ‘bare foot’ instance of weak BE can only be well-formed if it looks like (10c) (where extrametricality is indicated by the angle brackets): in (10c) no violation of strict layering obtains, because the extrametrical foot is invisible for the phonological component at the point where p-phrases are construed.



With all this in place, let us now look at a number of sample derivations to see how the proposed machinery can account for the patterns introduced in section 2.

6. Weak BE at the syntax-prosody interface

6.1 □ PaPa - weak BE (- XP)

I will start with the derivation of examples in which a monosyllabic BE-auxiliary in an E-periphrasis appears at the very end of a clause. Let us assume that the ‘VPAux’-example in (11) (from Iordanes, active at around 550 AD) was generated by Grammar B, and that the clause-initial subject does not sit in its thematic base position but rather in SpecSubjP.

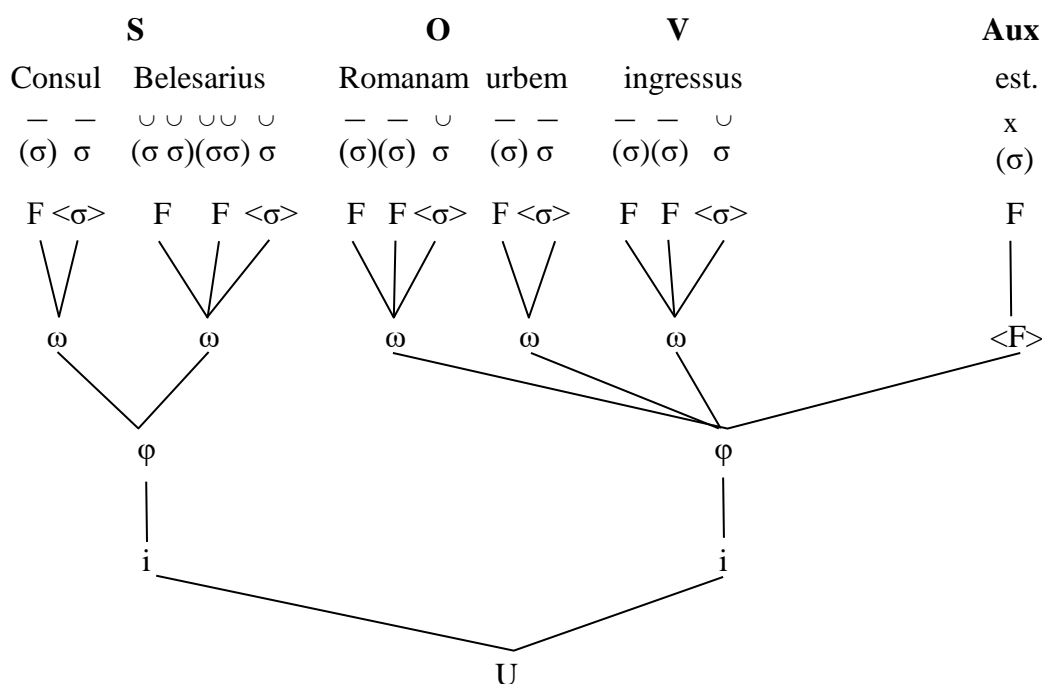
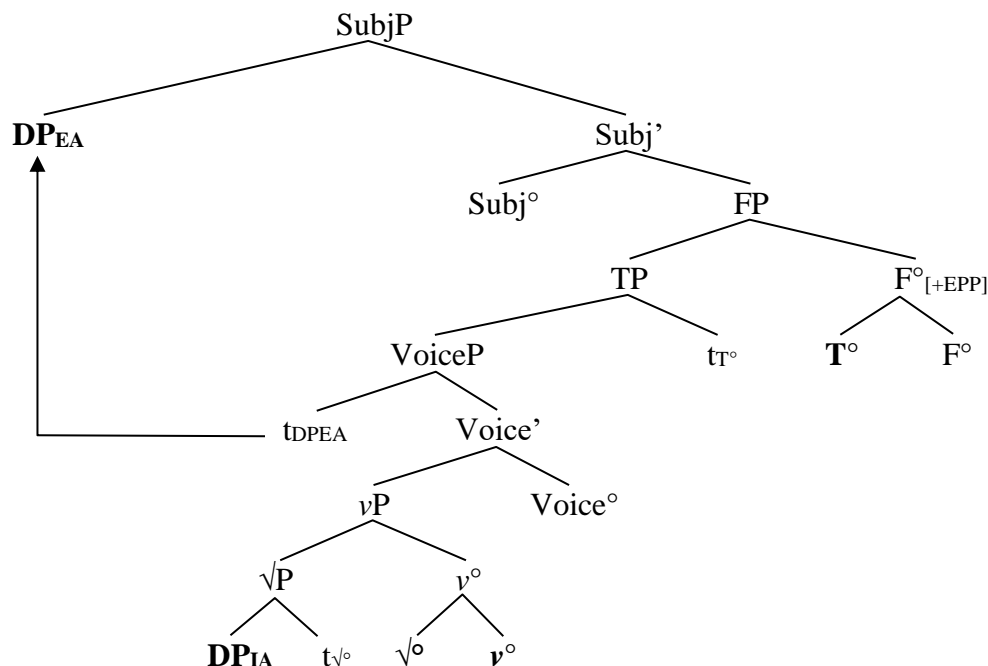
- (11) Consul Belesari-us Roman-am urb-em ingress-us
 consul.NOM Belesarius-NOM Roman-ACC city-ACC entered-NOM.M.SG
 est.
 be.PR.3SG
 ‘The consul Belesarius entered the city of Rome.’ (Iord. Rom. 373)

Full syntactic and prosodic structures of this example are provided in (12), where footed syllables are represented as ‘(σ)’, and extrametrical ones as ‘<σ>’.²⁴

phonological domains, for reasons that are at this point not well understood. I would like to leave this question open.

²⁴ Note that I am assuming that Latin trochaic feet can be uneven (cf. Jacobs 2000; Marotta 2000; *pace* Mester 1994): nothing crucially hinges on this. On foot formation (and stress assignment) in bisyllabic words with a light first syllable and an extrametrical final syllable, such as *sui* ‘her (GEN)’ and probably also *patris* ‘father (GEN)’ in (13)-(14), see Hayes (1995: 111 (‘incorporation of extrametrical material’)). Note also that the upcoming representations could be further enriched with information about vowel elisions in prevocalic extrametrical syllables, a complication which is orthogonal to the point at issue.

(12)



Let me briefly comment upon this structure. First of all, it should go without saying that the prosodic representation offered is to some extent hypothetical (but probably not too far off the mark). In any event, whenever weak BE appears at the end of a clause, there is obviously no way it can end up at the left edge (cf. (10a)) or in the middle of a phonological phrase (as in (10b)). Rather, it will automatically appear at the right edge of a phonological phrase, which in (12) is one that corresponds to the syntactic node FP.

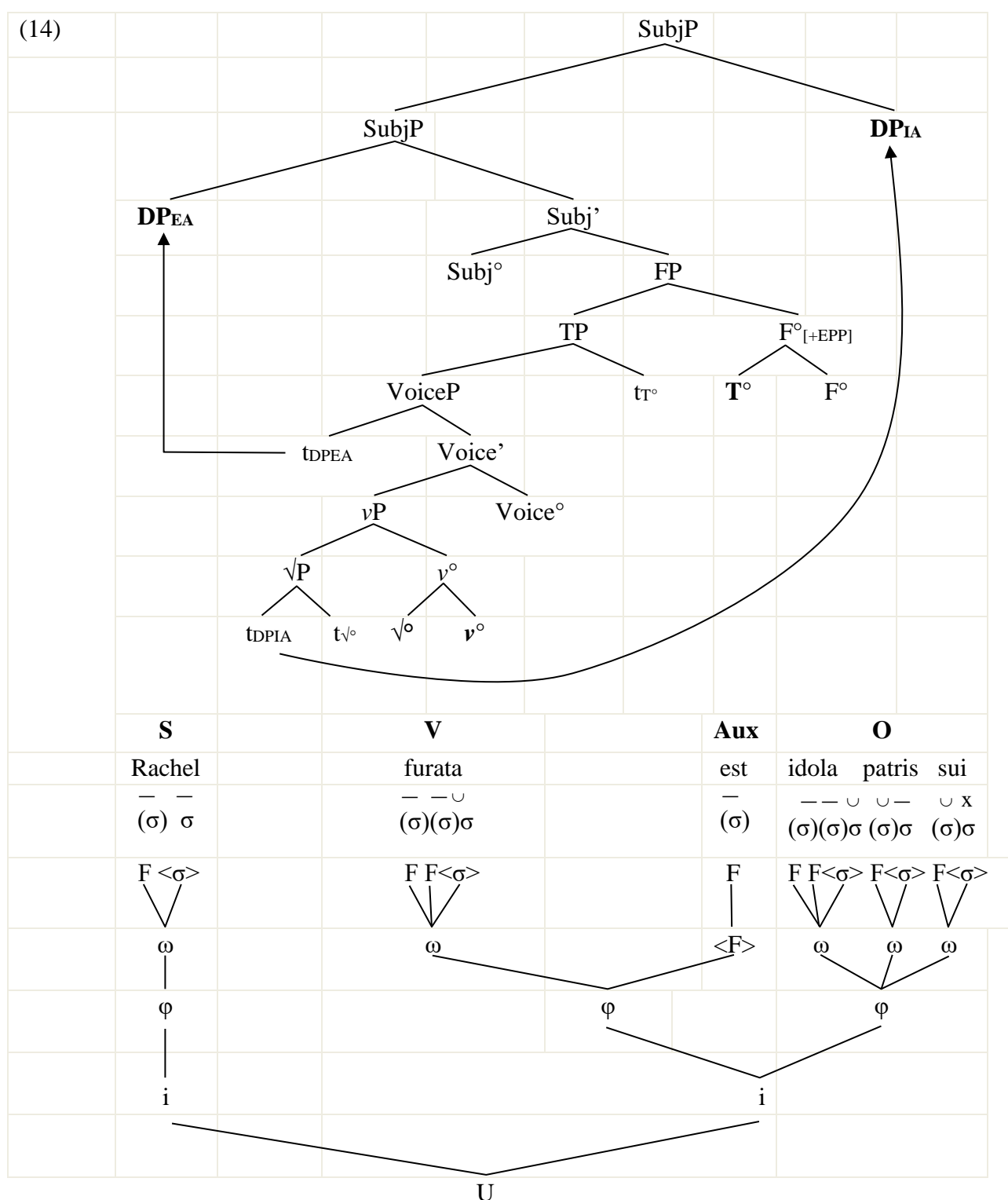
A side question that arises at this point is whether preverbal subjects are to be mapped onto an independent intonational phrase or not. As pointed out in Elordieta, Frota and Vigário (2005), modern Romance varieties differ in this respect. For example, in European Portuguese a preverbal subject by default constitutes a phonological phrase (and nothing more), whereas

in Castilian Spanish it typically appear as a larger unit, namely an intonational phrase. As can be seen in (12), I (tentatively) assume that preverbal subjects in Grammar B are mapped onto an independent intonational phrase, at least when occurring in SpecSubjP (in which case they could be said to be associated with an ‘aboutness’ reading). I will return to this particular point shortly.

Importantly, it certainly isn’t the case that the string ‘PaPa - weak BE’ has to occur clause-finally. Although I cannot at this point provide exact quantitative data, it is clear that examples such as (13), in which a head-final E-periphrasis is followed by additional phrasal material (in this case a direct object), are not hard to come by:

- (13) Rachel furat-a est idol-a patr-is su-i
 Rachel.NOM stolen-NOM.F.SG be.PRS.3SG idols-ACC father-GEN her-GEN
 ‘Rachel stole her father’s idols.’ (Vulg. (VT), *Isaias* 59.3)

It seems reasonable to assume that a post-INFL phrase such as the extraposed object in (13) corresponds to an independent phonological phrase. The full structure of this type of example would look like in (14) (where I assume XP extraposition to involve a process of rightward movement and adjunction to the top node of the phrase marker, with obvious simplifications):



This type of structure is correctly predicted to be compatible with a head-final BE-periphrasis featuring an instance of weak BE: in order for weak BE to end up at the end of the p-phrase derived from the syntactic node FP, it actually suffices for the extraposed constituent to be mapped onto a category which is at least as big a phonological phrase. Given the assumption that it is – at least as a default rule – syntactic constituents (rather than non-constituents) that are mapped onto prosodic constituents, we must assume that an alternative prosodic

representation in which weak BE leans rightward to be incorporated at the left edge of the constituent projected by the extraposed direct object would be ill-formed.

6.2 * Weak BE - PaPa

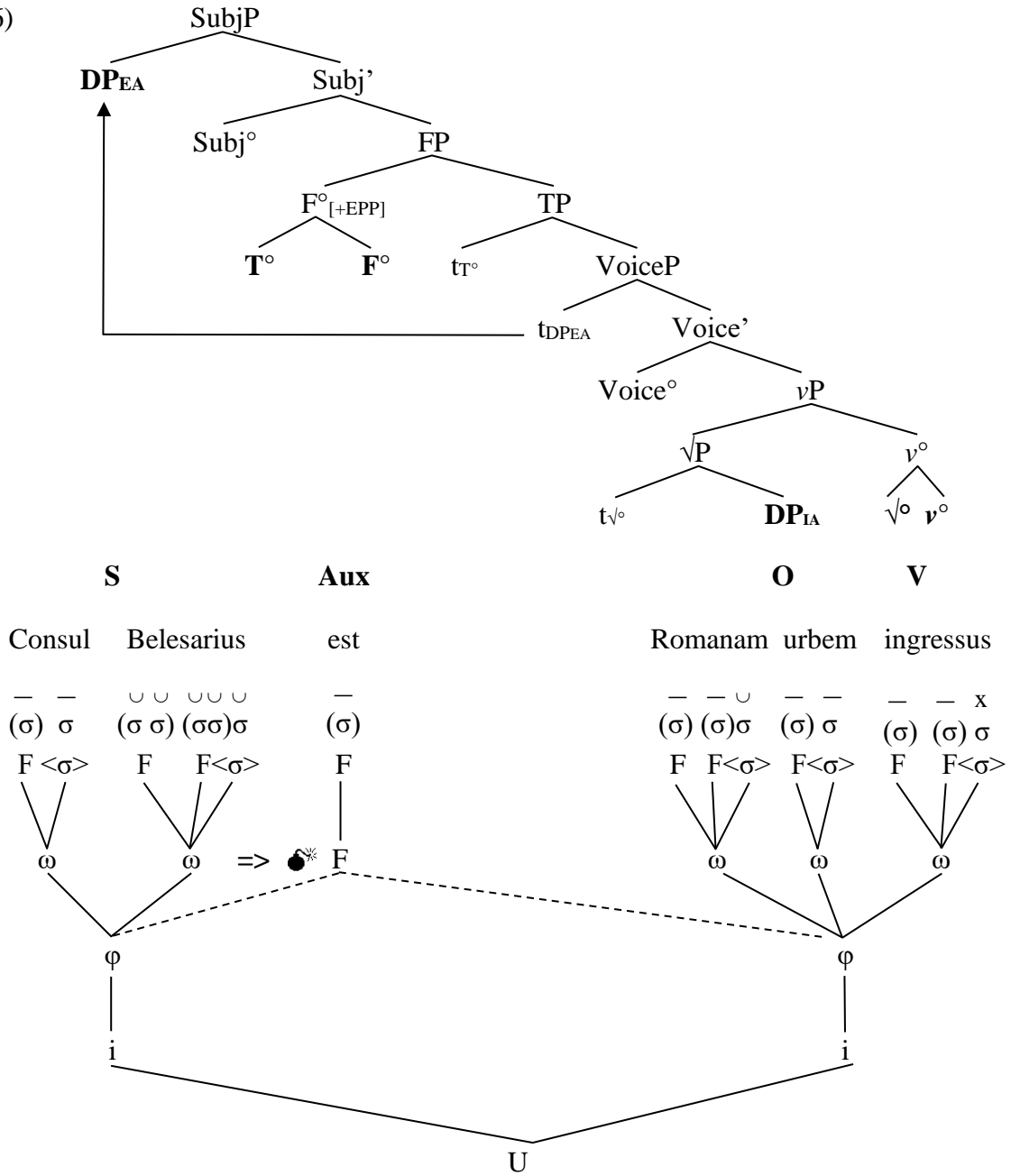
Next, let us consider why head-initial E-periphrases with a monosyllabic BE-auxiliary are so rare in Late Latin, or in other words, why affirmative ‘S - BE_{monosyllabic} - OV’ orders are only very infrequently attested in Late Latin texts (compare the constructed example (15), which only differs from its attested counterpart in (11) in that it features the order AuxVP rather than VPAux).

(15) ?? [consul Belesarius] est [Romanam urbem] ingressus

Given the way Grammar B was characterised in section 4.1, (15) is predicted to be syntactically grammatical, and given what we know about variable word order in Late Latin (generally more AuxVP than VPAux, and generally more OV than VO, cf. Danckaert 2017a), we actually predict this word order pattern to be one of the more frequently attested ones.

If we assume that the structure of a hypothesised example like (15) only differs from the one in (11)-(12) in the headedness of the node Voice, T and F (i.e. all the projections above vP which in (12) are head-final), we obtain a tree like (16):

(16)



As can be seen, this structure is not in compliance with the constraint in (9), as weak BE does not appear at the right edge of a phonological phrase: as indicated by the dashed lines, the relevant foot cannot be incorporated in either of the neighboring phonological phrases.

First, if syntactic and prosodic constituency were to coincide, weak BE would end up at the left edge of the p-phrase projected by FP. The result is a violation of the strict layer hypothesis (viz. a ‘stray F’ at a level where there should only be ω ’s), given that foot extrametricality can only obtain at the right edge of a p-phrase. We can now also be a bit more precise as to how exactly unacceptability comes about in cases like (15). Concretely, (9) should be taken to work as a post-syntactic filter (along the lines of e.g. Bošković 2001). There is nothing wrong with the syntactic tree in (16), which is fully compatible with Grammar B, and we have no reason to assume that the syntactic component has access to the phonological features associated with the lexical items it manipulates. However, the corresponding phonological representation is ill-formed, with unacceptability as a result.

Alternatively, in (16) weak BE could lean to the left and end up at the right edge of the phonological phrase projected by the subject. The problem with this scenario is that the nodes SpecSubjP and F° (with or without some additional maximal projection(s) in the main spine of the clause) do not together form a syntactic constituent, and as a result they are not mapped onto an independent prosodic unit, at least not by default. Corpus data provide empirical evidence in favor of this last point. If we split up the set of E-periphrases in two subsets, namely those that occur in clauses featuring a preverbal (pre-INFL) subject and those that don't, and if we look at word order preferences in both subsets, we obtain the following distribution (only samples with more than 8 preverbal subjects (external or internal arguments) taken into account):

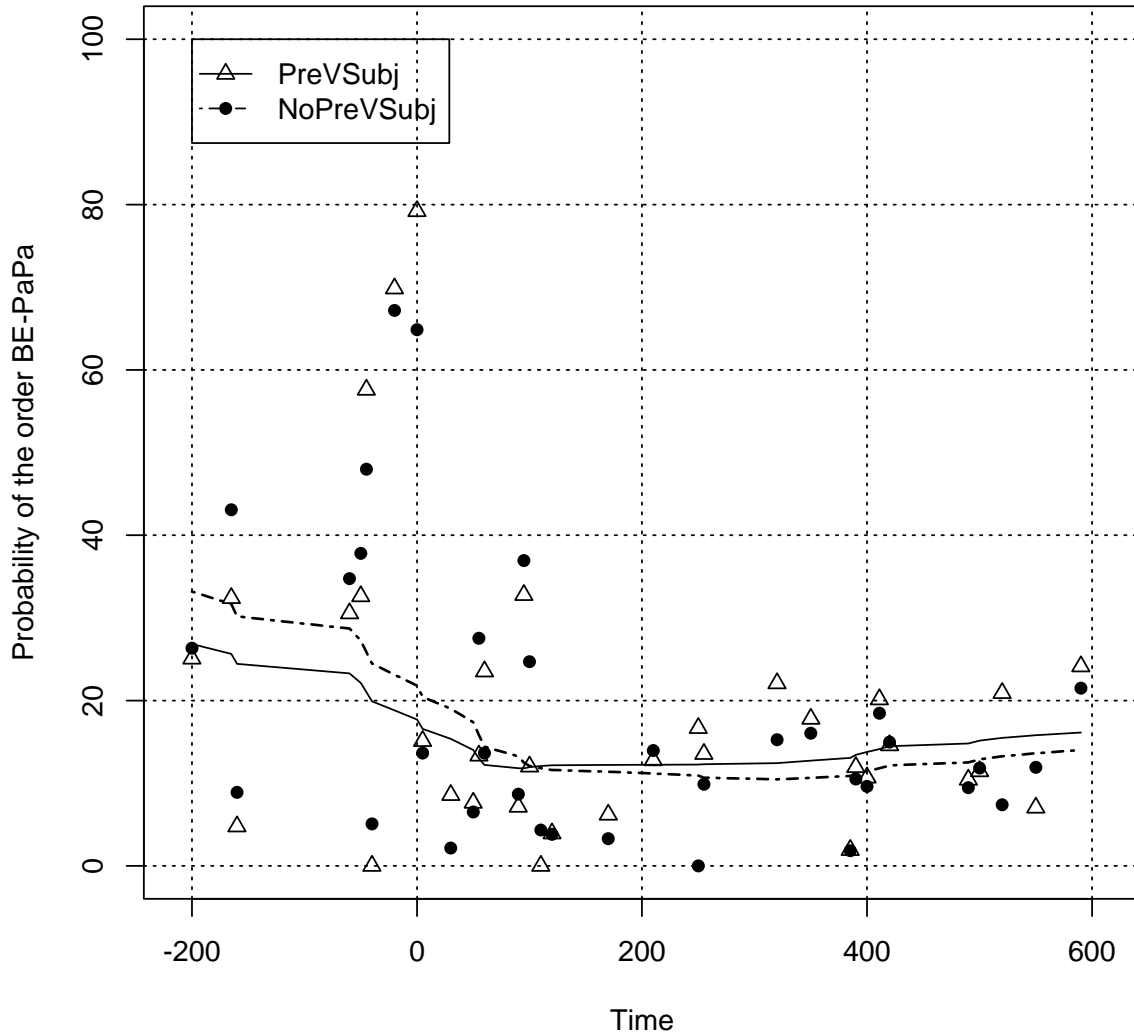


Figure 9: Incidence of the order ‘BE - PaPa’ (E-periphrases only), *ca.* 200 BC - 600 AD: clauses with and without an overt preverbal subject compared.

As we can see, in Late Latin the presence of a preverbal subject seems to favor the order ‘BE - PaPa’, but the effect is clearly only very small, and it is not entirely obvious to assess whether it is statistically significant. If we build a logistic regression to model the variable distribution of ‘PaPa - BE’ and ‘BE - PaPa’ orders in E-periphrases, with the factor ‘[± preverbal subject]’ as one of the independent covariates, alongside ‘Date’ (200 BC - 590 AD), ‘Negation’ (affirmative vs. negated), ‘SyllAux’ (polysyllabic vs. monosyllabic BE-auxiliary), ‘Clause type’ (embedded vs. main clauses), ‘Voice’ (deponent vs. passive BE-periphrases), as well as

a number of interaction terms, we obtain the results summarised in Table 2 (between brackets in the first column: level of categorical predictors for which the estimated log odds are given):

Table 2: Results of a logistic regression modeling the incidence of the order ‘BE - PaPa’ in Latin E-periphrases (*ca.* 200 BC - 600 AD).

Predictor	Estimate	Std. Error	p-value
(Intercept)	-0.7820513	0.0461583	<2e-16 (***)
Date	-0.0010144	0.0001592	1.88e-10 (***)
Negation (negated)	0.1558348	0.1656251	0.346762
SyllAux (monosyllabic)	-0.6755053	0.0421043	<2e-16 (***)
Clause type (main)	-0.1866203	0.0369689	4.46e-07 (***)
PreVSubj (present)	-0.0507338	0.0400040	0.204721
Voice (passive)	0.2568799	0.0394483	7.43e-11 (***)
Date:Negation	0.0027817	0.0008338	0.000849 (***)
Date:SyllAux	-0.0016884	0.0001783	<2e-16 (***)
Date:PreVSubj	0.0011452	0.0001686	1.10e-11 (***)
Negation:SyllAux	0.9301830	0.2106361	1.01e-05 (***)
Date:Negation:SyllAux	0.0027917	0.0009584	0.003580 (**)

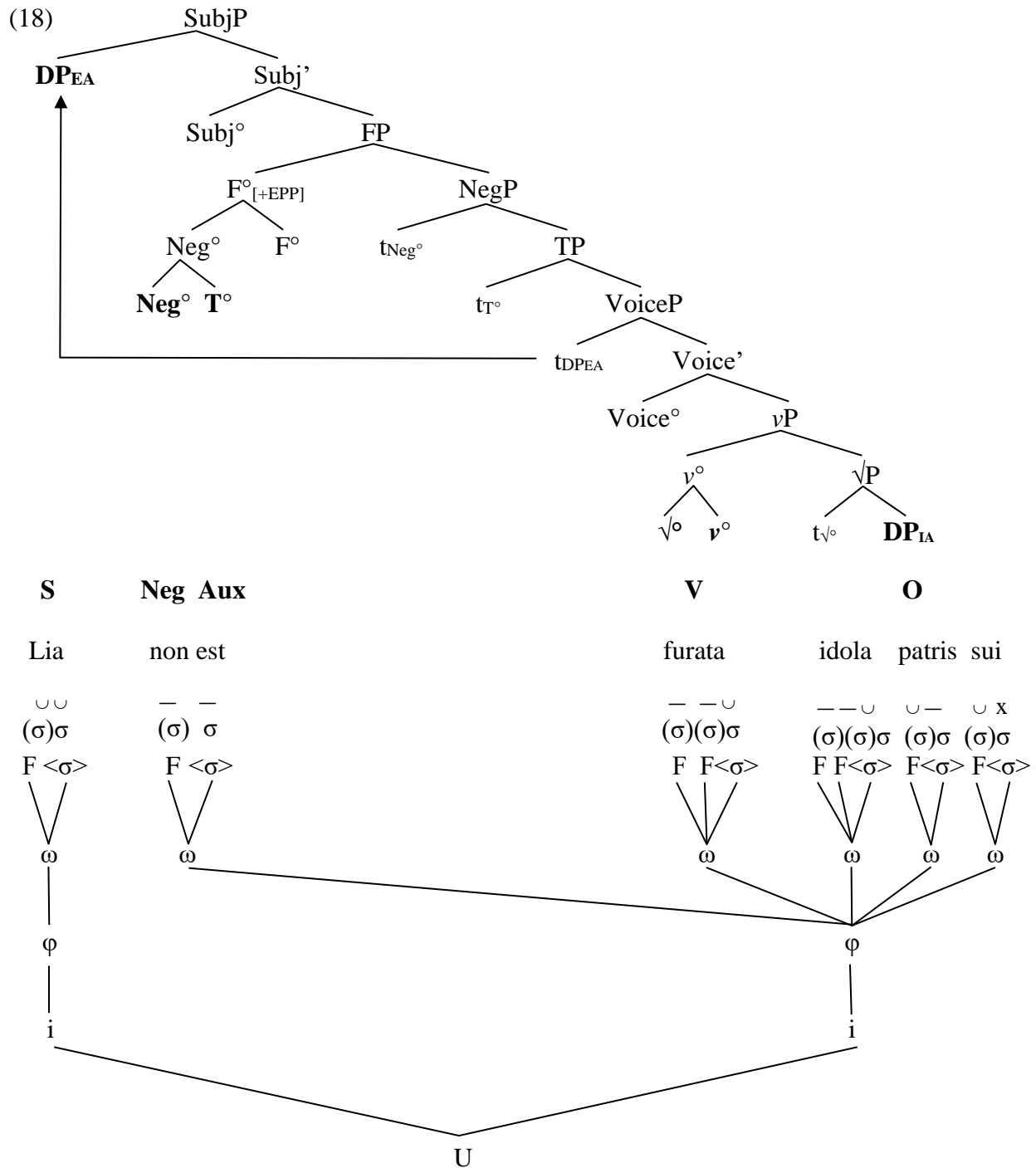
As indicated in the shaded rows, the main effect of the factor [\pm preverbal subject] is not significant, but an interaction term with ‘Date’ is, indicating that the effect of preverbal subject XPs on word order in BE-periphrases changes over time (cf. the crossing smoothed regression lines in Figure 9). In whatever way this effect is to be interpreted, as mentioned its size is not particularly big, and in any event it is much smaller than the effect of a much better predictor for word order, namely negation (cf. Figure 5).

6.3 □ Neg - Weak BE - PaPa

This brings us to our second major *explanandum*, namely the reason why in Late Latin E-periphrases the order ‘Neg - BE - PaPa’ is actually the preferred one. Consider for example a Late Latin ‘S - Neg - BE - VO’-clause like (17), which features once again a monosyllabic BE-auxiliary:

- (16) Et ideo Li-a non est furat-a [idol-a
 and therefore Lia-NOM not be.PRS.3SG stolen-NOM.F.SG idols-ACC
 patr-is su-i], sed Rachel.
 father-GEN her-GEN but Rachel.NOM
 ‘And therefore it was not Lia who stole the idols of her father, but Rachel.’ (Caes. Ar. *Sermo* 88.4)

The crucial element here is of course the status of the preverbal negator: as mentioned in section 4.1, in Grammar B this element is a proclitic, which standardly forms a complex head with the hierarchically highest verb in its host clause, which itself obligatorily moves to F. The structure we then obtain looks like (18):



As can be observed, weak BE heads a left-branching T-node: in the derived prosodic tree these two elements form a tight unit, and together they project a prosodic word. In this prosodic constituent the BE-auxiliary surfaces as an extrametrical final syllable of the prosodic word *non est* ‘not is’, in which it is properly incorporated, without even being footed. Crucially, univertation of weak BE and negation takes place in the syntax (viz. through syntactic head movement), and thus before spell-out, and before the construction of the prosodic tree. Put differently, weak BE can escape the ‘PF trap’ (9) by attaching to Neg° in the syntax. Consequently, there is no way that weak BE can surface as the type of metrical foot to which (9) can apply.

The analysis thus capitalises on the difference between syntactic and prosodic cliticization: although both Late Latin (non-emphatic) *non* and weak BE are phonologically deficient (in the sense that they constitute a single metrical foot that does not project a prosodic word), they have a different status. In the case of negation, phonological deficiency seems correlated with the syntactic operation of head movement (in a way that ultimately remains to be better understood), whereas the PF requirement on weak BE is evaluated post-syntactically.

Note in passing that in Late Latin the predominance of head-initial E-periphrases in negated clauses is presumably just an effect of time (cf. Figure 5): put differently, the rise of the order ‘*non* - BE - PaPa’ is probably best interpreted to reflect the ‘normal’ evolution towards the Romance languages, rather than an effect of negation *per se*. This interpretation receives support from the fact that in the regression analysis reported on in Table 2, the predictor ‘Negation’ only yields significant effects in interaction with ‘Date’ and ‘Number of syllables of the auxiliary’ (and both together), but not on its own. As a result, negation can be said to license (or save) the order ‘BE - PaPa’, but it does not trigger it.

6.4 Polysyllabic weak BE?

Before wrapping up this paper, let me very briefly say a couple of words about E-periphrases featuring a bi- or trisyllabic auxiliary, and by extension, about which type of ‘phonologically deficient’ category weak BE belongs to. As pointed out in for instance Selkirk (1996), it is possible for monosyllabic function words in English not to constitute an independent prosodic word. In contrast, from the moment a function word is minimally bisyllabic, it always seems to project a p-word. More recently it has been argued that the broader typological picture is slightly more complex. Zec (2005) distinguishes two types of functional categories, which she calls ‘free’ and ‘bound’ function words. As the terminology suggest, the latter are characterised by a lesser degree of phonological independence. Which class a given functional category belongs to is specified in the lexicon.

Interestingly, items belonging to the class of bound function words can in fact fail to project a p-word even if they are bisyllabic, whereas bisyllabic free function words always project an independent p-word. The question then arises as to whether Latin weak BE can be classified as a free or rather as a bound function word. Monosyllabic forms of *esse* may not directly inform us about this question, but bisyllabic ones potentially could: if these could be shown to behave more like monosyllabic forms of weak BE than for instance bisyllabic forms of modal verbs, we could reasonably hypothesise that weak BE can be characterised as a bound function word, in the sense of Zec (2005). Recall that we had observed a clear difference between monosyllabic and polysyllabic forms of *esse* in E-periphrases, at least as far as word order is concerned (cf. Figures 7 and 8). However, it is certainly not the case that polysyllabic E-auxiliaries have the same word order preferences as for instance Late Latin modal verbs (cf. Figures 1 and 2).

Another interesting question is whether the presence of negation favors the head-initial order ‘BE - PaPa’ when the auxiliary involved is polysyllabic, as it does with monosyllabic auxiliaries. Unfortunately, the evidence is not entirely unequivocal, which is mainly so because some of the samples in my corpus are very small. Let us have a look at some figures: because of the ‘small sample size issue’ just mentioned, I shall only give figures for the first four centuries in the corpus considered together, and I shall compare these to the corresponding values for the last four centuries. The data concerning the monosyllabic auxiliaries that we have concentrated on so far are summarised in Table 3. As we know, the main thing to observe here is that ‘PaPa - BE’ orders are the norm in Late Latin affirmative clauses, whereas the reverse order is preferred in negated clauses. No doubt tellingly, this last effect can only be observed in Late Latin:

Table 3: Word order in E-periphrases with a monosyllabic auxiliary.

	200 BC - 200 AD		200 - 600 AD	
	PaPa - BE	BE - PaPa	PaPa - BE	BE - PaPa
affirmative	7992	2124	8738	781
negated	140	101	57	159

With bisyllabic auxiliaries, which have a considerably lower token frequency, a similar picture emerges, but the observed tendencies are less extreme. As shown in Table 4, in both the earlier and the later period ‘PaPa - BE’ orders outnumber ‘BE - PaPa’ patterns in affirmative clauses (with a ratio of about 2:1). In contrast, in negated environments the order ‘BE - PaPa’ is preferred in both periods (but observe that there are only 56 negated tokens for the last 400 years; however, the observed outcome (22 head-final tokens vs. 34 head-initial ones) can be shown to be statistically significant (Pearson’s chi-squared test, $p = .0297343$)). Especially the preponderance of negated ‘BE - PaPa’ orders in the earlier centuries very much remains to be accounted for. In any event, the observed findings are at least compatible with the hypothesis that bisyllabic forms of BE can fail to project a prosodic word, as otherwise we would expect ‘PaPa - BE’ orders to be more frequent in negated clauses too.

Table 4: Word order in E-periphrases with a bisyllabic auxiliary.

	200 BC - 200 AD		200 - 600 AD	
	PaPa - BE	BE - PaPa	PaPa - BE	BE - PaPa
affirmative	2446	1237	976	522
negated	57	99	22	34

If we then turn to BE-auxiliaries counting three syllables, the picture seems to change:

Table 5: Word order in E-periphrases with a trisyllabic auxiliary.

	200 BC - 200 AD		200 - 600 AD	
	PaPa - BE	BE - PaPa	PaPa - BE	BE - PaPa
affirmative	393	155	802	394
negated	9	9	31	25

Here too the values in the cells on the bottom row are particularly small (and therefore very difficult to interpret), but it does seem to be the case that the favoring effect of negation is now absent, witness the fact that in Late Latin, negated ‘PaPa - BE’ patterns are slightly more frequently attested than ‘BE - PaPa’ orders (viz. 31 vs. 25 tokens; we cannot exclude that this outcome is due to chance (Pearson’s chi-squared test, $p = .0773609$)). As it is in all likelihood never possible for a trisyllabic (lexical or functional) word not to constitute an independent p-word, this state of affairs is of course expected.

7. Conclusion

In this paper I have revisited a number of Late Latin word order data involving verbal periphrases with *esse* ‘be’. I started by pointing out that these tend to occur in the head-final order ‘PaPa - BE’, i.e. a variant of the pattern ‘VPAux’ which in Late Latin is generally on the decline. Importantly, in negated clauses it is the reverse order ‘BE - PaPa’ which is (strongly) preferred. I proceeded to develop an account of these observations which makes reference to both syntactic and prosodic properties of Late Latin clauses featuring a BE-periphrasis. Although probing the phonology of a dead language poses evident methodological challenges, I have shown that there is indeed evidence for the claim that prosodic factors are at play in bringing about the observed word order data.

The core proposal is that the BE-auxiliary involved in the relevant formations is a newly created lexical item which I called ‘weak BE’, and which differs from older forms of *esse* in that it is – or at least can be – phonologically deficient and thus unable to project an independent prosodic word. It was shown that given a number of assumptions about the syntax-prosody mapping, the observed word order facts can neatly be accounted for. I concluded the paper with some discussion of the differential behavior of monosyllabic, bisyllabic and trisyllabic BE-auxiliaries. I tentatively suggest that it may well be the case that certain bisyllabic forms of BE may also fail to be mapped onto a prosodic word. At a theoretical level, the present proposal supports the idea that much like syntactic variation, prosodic variation is encoded in the lexicon too (in line with Zec 2005).

References

- Adams, James N. (1994). *Wackernagel’s Law and the Placement of the Copula esse in Classical Latin*. Cambridge: Philological Society.
- Agbayani, Brian and Chris Golston (2016). ‘Phonological Constituents and their Movement in Latin’, *Phonology* 33: 1-42.
- Alexiadou, Artemis and Elena Anagnostopoulou (1998). ‘Parameetrizing AGR: Word Order, Movement and EPP-Checking’, *Natural Language and Linguistic Theory* 16: 491-539.
- Biberauer, Theresa (2003). *Verb second (V2) in Afrikaans: A Minimalist Investigation of Word Order Variation*. Doctoral dissertation, University of Cambridge.
- Biberauer, Theresa, Anders Holmberg, and Ian Roberts (2014). ‘A Syntactic Universal and its Consequences’, *Linguistic Inquiry* 45: 169-225.
- Boldrini, Sandro (1999). *Prosodie und Metrik der Römer (aus dem Italienischen übertragen von Bruno W. Häuptli)*. Leibzig: Teubner.
- Bornecque, Henri (1907). *Les clausules métriques latines*. Lille: Université de Lille.
- Buckley, Eugene (1994). ‘Persistent and Cumulative Extrametricality in Kashaya’, *Natural Language and Linguistic Theory* 12: 423-64.
- Bošković, Željko (2001). *On the Nature of the Syntax-Phonology Interface*. Amsterdam: Elsevier.
- Danckaert, Lieven (2016). ‘Variation and Change in Latin BE-periphrases: Empirical and Methodological Considerations’, in James Adams and Nigel Vincent (eds), *Early and Late Latin: Continuity or Change?*, Cambridge: Cambridge University Press, 132-62.
- Danckaert, Lieven (2017a). *The Development of Latin Clause Structure: A Study of the Extended Verb Phrase*. Oxford: Oxford University Press.
- Danckaert, Lieven (2017b). ‘The Origins of the Romance Analytic Passive: Evidence from Word Order’, in Eric Mathieu and Robert Truswell (eds), *Micro-Change and Macro-Change in Diachronic Syntax*. Oxford: Oxford University Press, 216-35.
- Danckaert, Lieven (2017c). ‘Subject Positions in the History of Latin’, *Catalan Journal of Linguistics* 16, 125-161.

- Di Capua, Francesco (1914). 'Il *Cursus* nel *de consolatione philosophiae* e nei trattati teologici di Severino Boezio', *Didaskaleion* 3: 269-303.
- Di Capua, Francesco (1937). *Il ritmo prosaico nelle lettere dei papi e nei documenti della cancelleria romana dal IV al XIV secolo*, vol. I. Roma: Facultas theologica pontificii athenaei seminarii romani.
- Elordieta, Gorka (2008). 'An Overview of Theories of the Syntax-Phonology Interface', *Journal of Basque Linguistics and Philology* 42: 209-86.
- Elordieta, Gorka, Sónia Frota and Marina Vigário (2005). 'Subjects, Objects and Intonational Phrasing in Spanish and Portuguese', *Studia Linguistica* 59: 110-43.
- Fraenkel, Eduard (1932-'33). Kolon und Satz: Beobachtungen zur Gliederung des antiken Satzes I-II, *Nachrichten der Göttinger Gesellschaft der Wissenschaften. Philosophisch-historische Klasse* 1932, 197-213; 1933, 319-54.
- Habinek, Thomas (1985). *The Colometry of Latin Prose*. Berkeley, CA: University of California Press.
- Hagendahl, Harald (1937). *La prose métrique d'Arnobé: Contributions à la connaissance de la prose littéraire de l'empire* (= Göteborgs Högskolas Årsskrift 42, 1936: 1). Göteborg: Elanders Boktryckeri Aktiebolag.
- Havet, Louis (1892). *La prose métrique de Symmaque et les origines métriques du cursus*. Paris: Bouillon.
- Hayes, Bruce (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago, IL: University of Chicago Press.
- Hinterhölzl, Roland and Ans van Kemenade (2012). 'The Interaction between Syntax, Information Structure, and Prosody in Word Order Change'. In Terttu Nevalainen and Elisabeth Closs Traugott (eds), *The Oxford Handbook of the History of English*. Oxford: Oxford University Press, 803-21.
- Holmes, Nigel (2015). *esse* and *-que* in Late Latin prose rhythm. Paper presented at ICLL 18, Université de Toulouse - Jean Jaurès, 10 June 2015.
- Holmes, Nigel (2017). 'Esse and -que in Late Latin Prose Rhythm', *Pallas* 103: 251-9.
- Hyde, Brett (2011). 'Extrametricity and Non-Finality'. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume and Keren Rice (eds), *The Blackwell Companion to Phonology*. Oxford: Wiley-Blackwell, 1027-51.
- Jacobs, Haïke (2000). 'The Revenge of the Uneven Trochee: Latin Main Stress, Metrical Constituency, Stress-Related Phenomena and OT', in Aditi Lahiri, (ed.), *Analogy, Levelling, Markedness: Principles of Change in Phonology and Morphology*. Berlin: Mouton de Gruyter, 333-52.
- Kayne, Richard (1994). *The Antisymmetry of Syntax*. Cambridge, MA: The MIT Press.
- Marotta, Giovanna (2000). 'Sulla massimalità dei piedi trocaici: Il caso del latino', *Lingua e Stile* 35: 387-416.
- Mester, Armin (1994). 'The Quantitative Trochee in Latin', *Natural Language and Linguistic Theory* 12: 1-61.
- Nespor, Marina and Irene Vogel (2007 [1986]). *Prosodic Phonology (with a New Foreword)*. Berlin: Mouton de Gruyter.
- Oberhelman, Steven (2003). *Prose Rhythm in Latin Literature of the Roman Empire: First Century B.C. to Fourth Century A.D.* Lewiston (NY): Edwin Mellen.
- Pezzini, Giuseppe (2015). *Terence and the Verb 'to be' in Latin*. Oxford: Oxford University Press.
- Salvi, Giampaolo (2004). *La formazione della struttura di frase romanza: Ordine delle parole e clitici dal latino alle lingue romanze antiche*. Tübingen: Niemeyer.

- Selkirk, Elisabeth (1980). 'Prosodic Domains in Phonology: Sanskrit Revisited'. In: Mark Aronoff and Mary-Louise Kean (eds.), *Juncture: A Collection of Original Papers*. Saratoga, CA: Anma Libri, 107-29.
- Selkirk, Elisabeth (1981). 'The Nature of Phonological Representation'. In Terry Myers, John Laver and John Anderson (eds.), *The Cognitive Representation of Speech*. Dordrecht: North Holland, 379-88.
- Selkirk, Elisabeth (1986). 'On Derived Domains in Sentence Phonology', *Phonology Yearbook* 3: 371-405.
- Selkirk, Elisabeth (1996). 'The Prosodic Structure of Function Words', in James Morgan and Katherine Demuth (eds), *Signal to Syntax: Bootstrapping from Syntax to Grammar in Early Acquisition*. Mahwah, NJ: Erlbaum, 187-213.
- Selkirk, Elisabeth (2011). 'The Syntax-Phonology Interface'. In Goldsmith, John, Jason Riggle and Alan Yu (eds), *The Handbook of Phonological Theory*. Oxford: Wiley-Blackwell, 435-84.
- Speyer, Augustin (2010). *Topicalization and Stress Clash Avoidance in the History of English*. Berlin: Mouton de Gruyter.
- Taylor, Ann (2005). 'Prosodic Evidence for Incipient VO Order in Old English', *English Language and Linguistics* 9: 139-56.
- Truckenbrodt, Hubert (1999). 'On the Relation between Syntactic Phrases and Phonological Phrases', *Linguistic Inquiry* 30: 219-55.
- Truckenbrodt, Hubert (2007). 'The Syntax-Phonology Interface'. In de Lacy, Paul (ed.), *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, 435-56.
- Väänänen, Veikko (1987). *Le journal-épître d'Égérie (Itinerarium Egeriae): Etude linguistique*. Helsinki: Academia Fennica.
- Zec, Draga (2005). 'Prosodic Differences among Function Words', *Phonology* 22: 77-112.