Mikołaj Nkollo*, Beata Malczewska**

* Uniwersytet im. Adama Mickiewicza w Poznaniu, Wydział Neofilologii
** Uniwersytet Pedagogiczny im. Komisji Edukacji Narodowej w Krakowie,
Wydział Filologiczny

Towards phrasal attachment
of European Portuguese proclitics

A corpus-based inquiry
into diachronic phonology-syntax interface[*]

The aim of the present paper is to show how corpus analysis can be used to pinpoint the rules that govern the mapping of morphophonological and syntactic structures. The case in point is the European Portuguese (EP) clitic system and the changes it has gone through over the course of history. The topic central to this paper is interpolation, i.e. the insertion of an element between a clitic pronoun and the verb (schematically cl-X-verb). In Romance linguistics, clitic-verb non-adjacency is an unusual linear order whose significance

stems from the fact that it is assumed to lend support to the post-lexical (syntactic) rather than affixal (or word-internal) nature of clitics and clitic attachment (Miller, Monachesi 2003: 63–65, 80–81; Heap, Oliviéri, Palasis 2017: 213–215 specifically with reference to subject clitics in Italo-Romance and Gallo-Romance varieties). In the history of EP, interpolation has been, albeit at varying degrees, a viable structural model. For ca. 180 years (mid 17th to early 19th centuries), no item but the *não* sentential negator was eligible to disrupt the continuity of clitic-verb sequences (cf. Martins, 2016b: 422–423). Thus, two concurrent and semantically interchangeable linear models were active: (a) (proclisis trigger)-X-cl-verb and (b) (proclisis trigger)-cl-X-verb.

Present-day EP is known to show a strong asymmetry between its preverbal and postverbal clitic pronouns. One sign of this asymmetry is that while the preverbal clitic pronouns fail to exhibit clitic-specific allomorphic variation, enclitics interact with the verb form in an intricate, idiosyncratic way. Namely, enclitics both induce allomorphy in the stem they attach to and also undergo stem-induced allomorphy themselves (Spencer, Luís 2012: 205–206; Luís, Kaiser 2016: 221–222). In so doing, they display lexical phonology and an affix-like morphological behavior. By contrast, proclitics' connection to their host is much looser, so they are presumed to be inserted post-lexically. The second sign of clitic asymmetry in EP has to do with the fact that preverbal or postverbal placement of its pronouns hinges on specific syntactic constraints and is not dependent on the finiteness of the verb form, as in the remaining Romance varieties (Gerlach 2002: 63–64). While enclisis is the default position in root clauses, proclisis arises only when specific expressions, called 'proclisis triggers', appear preverbally in a sentence. These triggers include negative words, complementizers (subordinate conjunctions, prepositions), indefinite pronouns, some adverbs and quantifiers, wh-words and/or interrogative pronouns.

However, in the history of EP, morphophonological structures have not always been mapped onto the sentence level as described above. This paper is an attempt to trace the mechanisms that were instrumental in the emergence of contemporary clitic grammar and the abandonment of its earlier versions. What we are proposing here, on the basis of relevant corpus analysis, is that interpolation was a prominent factor prompting the change.

The paper is divided into four parts, moving from theory to history, and then delving into corpus analysis. Section 1 describes the significance of clit-

ic attachment from the lexicalist perspective, thereby implying a modular view of speech production and a clear-cut morphology vs. syntax divide. This part draws, to a large extent, on the criteria formulated by Zwicky and Pullum (1983: 503–504) in their seminal paper on how clitics diverge from affixes. Section 2 outlines how interpolation has changed throughout EP history. The focus is on the time between the mid-17th and the early 19th century, but some space is allotted to the other periods, as well. In Section 3, the light is on the corpus, and empirical results are shown alongside their preliminary analysis. Our corpus has been compiled on the basis of 39 texts released between 1614 and 1858. At present, it includes 659 manually retrieved examples of both (proclisis trigger)-cl-*não*-verb and (proclisis trigger)-*não*-cl-verb sequences. The two sequences are compared against each other in turn. At this point, it becomes apparent how the corpus-based data can be used to refine the very concept of interpolation and verify some of the ideas found in the previous literature devoted to this topic. Section 4 focuses on how the two models under discussion were patterned over appropriate contexts and in what environments the cl-*não*-verb sequence was given preference over proclisis. This discussion gives us vital insight into how present-day clitic system became dominant in the 19th century. Finally, in the concluding remarks the study results are summarized and confronted with modern corpus-based data. This ultimate analysis ties in with our findings.

1. The significance of clitic attachment

The issue of how clitics are generated and at which point of derivation they are inserted into the structure has always been a tricky one in Romance linguistics. Many proposals have been formulated in answer to this question (e.g. Duarte, Matos 2000; Bouzouita, Kempson 2006). For Anderson (2005: 22–24), the only difference between clitics and affixes lies in the type of the domain they attach to. Affixes are parts of morphological units (stems or inflected forms), but clitics tend to belong to larger syntactic constituents. Therefore, a rigorous distinction between morphological and phrasal attachment of clitic pronouns must be drawn (Gerlach 2002: 61–63). This is consistent with the idea argued for in diachronic linguistics (Duarte, Matos 2000: 122; Delfitto 2002: 47–48) that historically, clitics undergo a change of identity, of sorts, that takes them from free morphemes and turns them into

affixes. Obviously, not all clitics in contemporary Romance varieties are at the same point of this evolutionary process.

The remaining discussion has been inspired by the findings of Zwicky and Pullum (1983). The arguments repeatedly advanced in favour of either wordhood or affixhood of EP clitics are overwhelmingly reminiscent of the criteria that these two scholars used to solve the controversy over the difference between the English *not* and *n't*. Three of these criteria deserve special attention:

(i)   Unlike clitics, which are moderately selective about the syntactic category of their host, affixes show a high degree of selectivity with respect to their stems;

(ii) Morphophonological idiosyncrasy is more easily found in affixed words than in clitic-host combinations;

(iii) Likewise, arbitrary gaps in the set of conceivable combinations are more characteristic of affix-base combinations than of clitic-host ones.

Moreover, it has been observed that clitics, but not affixes, can be added recursively to material already containing clitics and that semantic idiosyncrasies seem to arise in affixed words at a larger scale than in host-clitic groups (Miller, Monachesi 2003: 92–93). In what follows, the above criteria are briefly related to the distributional facts and rules of clitic attachment in EP.

The first criterion says, in effect, that Romance clitic pronouns, unlike other phonologically deficient items (see examples below, with the definite singular masculine article), can only attach to a verb form. If the context forces the verb into omission, a form called 'strong pronoun' is mandatory (2).

(1)   ***o treinador*** (noun) *disse que* … 'The trainer said that …'
        o futuro (adjective) *treinador disse que* ... 'The future trainer said that …'
        o abaixo (adverb) *assinado treinador disse que* … 'The undersigned trainer said that …'

(2)   *O que é que tu viste*? 'What did you see?' – *O (intended meaning 'This').

Another constraint derived from the first criterion is that nothing but another clitic can be inserted between a clitic pronoun and the verb. One of the rare exceptions to this principle is interpolation (Sánchez-Rei 1999: 87–91), which is, incidentally, of central interest to the following discussion. A by-product of this criterion is that clitics cannot be coordinated (Luís 2014: 206).

(3) a. *O João convidou-os, a ele e a ela*
        ART John invite.PFV.IND.3.SG-ACC.3.M.PL, to he and to she
        John invited them, him (strong pronoun) and her (strong pronoun)
    b. **O João convidou-o e a*
        DEF John invite.PFV.IND.3.SG-ACC.3.M.SG and ACC.3.F.SG

The second criterion deals with the special phonology of clitic-verb sequences. Since scholars generally agree that lexicon is the locus of phonological idiosyncrasy, special phonology has been repeatedly evoked to support the claim that clitic pronouns are word-internal rather than syntactically attached items. Two examples can be used to illustrate this point: the realizations of 3rd person direct object pronouns and the *nos* enclitic. These are dealt with in order below.

The 3rd person direct object pronouns, whose default realizations are *o*, *a*, *os*, *as* (M.SG, F.SG, M.PL and F.PL, respectively), take on the forms of *lo*, *la*, *los*, *las* if placed after the verb forms ending in consonants: *s*, *z*, and *r*. Moreover, the *l*-variant deletes the final consonant of the verb form. Even if linguistic signs belonging to other categories display the same sound sequence, they do not undergo analogous shifts[1].

(4) a. *Deves pô-la no forno* 'You should put it in the oven' / **Deves pôr*
        a no forno
        must.PRS.IND.2.SG put.INF ACC.3.F.SG in-ART.DEF.M.SG oven
    b. *Deves pôr a caçarola no forno* 'You should put the casserole in the
        oven'
        must.PRS.IND.2.SG put.INF ART casserole in-ART.DEF.M.SG oven

---

[1] One notable exception to this morphophonological process is also clitic-specific: *l*-insertion and final consonant deletion take place within clitic groups juxtaposing indirect and direct objects: *no-lo* 'it to us' and *vo-lo* 'it to you.PL', which block, respectively, **nos-o* and **vos-o*.

    c. *Deves pô-**la** caçarola no forno*

The *l*-insertion and final consonant deletion can also occur verb-internally in future and conditional forms. The phenomenon is known as mesoclisis (Piel 1989: 225; Duarte, Matos 2000: 132–134) and extends, alongside other pronouns, to stems that proceed from the Latin infinitive. Stem-allomorphy happens when 3rd person direct object clitics are inserted between the final [r] of the stem and inflectional endings: *dá-lo-ás* (**dar-o-ás*) '(you) will give it', *explicá-lo-íamos* '(we) would explain it'. Additionally, in individual verbs, the *l*-insertion and final consonant deletion fail to occur in some inflectional forms. For example, *quer* '(he) wants' does not turn into **quê-lo*, as expected under this rule, but surfaces instead as *quere-o*.

The second illustration of the criterion (ii) above has to do with the *nos* enclitic. The enclitic deletes the final [s] consonant in 1.PL verb forms it attaches to.

(5) a. *alegr**amo**-nos* 'we are pleased', **alegr**amos-nos***
     b. *alegr**ámo**-nos* 'we were pleased', **alegr**ámos-nos***

No comparable phenomena are observed elsewhere, even if non-clitic items happen to contain the same sound material: *Esta**mos nos** correios* '(We) are at the post office' / **Esta**mo nos** correios*.

To summarize, EP clitic pronouns both trigger and undergo allomorphy in contexts where the remaining types of linguistic signs fail to do so.

The arbitrary gaps, referred to in the third criterion above, relate to the fact that, for no apparent reasons, some conceivable person-number combinations are not attested (Miller, Monachesi 2003: 87–88). Instead, language users take advantage of other available sequences; for instance, a single clitic pronoun and a stressed one. To be sure, some of these gaps are due to restrictions on feature co-occurrence, but others lack any explanation whatsoever. Example (6a), below, posits a theoretical juxtaposition of the 1st person indirect object and 2nd person direct object ('you to me'). However, such sequences are disallowed in EP. Therefore, speakers opt for substituting *me* with a sequence involving a preposition and a stressed pronoun (i.e. *a mim*) that follows *trasladou-te*.

(6) a. **O Paulo trasladou-me-te* 'Paul transferred you to me'
          ART Paul transfer-PRF,IND.3.SG DAT.1.SG ACC.2.SG
      b. *O Paulo trasladou-te a mim*

Another gap involves an incomplete specification of the 3rd person indirect object clitic for the features [singular] and [plural]. If used as the only clitic in pronoun-verb or verb-pronoun combinations, its number is consistently marked in present-day EP. Yet, if used in clitic clusters, it remains unspecified for the category of number. Indeed, the clusters *lho*, *lha*, *lhos*, *lhas* fail to express how many individuals are actually concerned. For example, *Transmitiram-lho* could mean, 'They transmitted it to him/her' or 'They transmitted it to them'.

The phenomena discussed above are usually adduced to argue for word-internal status of clitics in EP. The arguments can be neatly subsumed under the principle known in linguistics as Strong Lexicalist Hypothesis (SLH, also known as Lexical Integrity Hypothesis): No syntactic rule can refer to elements of morphological structure (Dalrymple 2001: 83–84, 97–98).

However, if we assume that speech production is hierarchically ordered (i.e. that various modules are built upon one another; see Levelt 1989; Labelle 2001: 156, 161–162) and that it operates sequentially, SLH produces some interesting consequences for EP clitics. If clitic pronouns (and affixes) are parts of another expression, their processing must start very early. After all, the syntactic module can only operate once lexical items are fully specified morphologically. Thus, syntax cannot influence clitic pronouns in speech production. Even more radically, SLH requires that lexical competence (or mental lexicon) of EP users contain two separate entries: one for a bare verb (with no clitic pronouns) and another one, which combines a verb form and a clitic pronoun and whose use becomes mandatory if arguments have not been planned in an earlier module as full-fledged nouns and NPs. Syntax must operate on an entirely specified input ('operating' includes, in this context, movement, deletion, and addition of word-internal material; the application of any of these operations violates the word's integrity), because the behavior of clitic pronouns is so idiosyncratic that no productive rule is likely to explain all its intricacies (Bresnan 2001: 92). Put differently, the verb-clitic (or clitic-verb) sequence must be memorized individually, just as stem-affix combinations are claimed to be.

Yet, under the assumption that clitics combine with stems in much the same way as affixes do with their bases, numerous features remain unaccounted for. Indeed, substantial counter-evidence to what SLH asserts is found in literature. This counter-evidence highlights the properties that clitics share with words, thereby posing a serious challenge to the treatment of clitics as if they were affixes (Vigário 2003: 131–155).

The essential criticism lies in the ability of clitic pronouns to occupy variable positions with respect to the verb, something that affixes normally do not do. Likewise, the lack of phonological interaction in clitic-verb sequences further undermines the explanations that are based only on what happens in the postverbal domain. In fact, these two characteristics suggest that rather than being affixes, pronominal clitics are postlexical (just like ordinary, i.e. non-clitic, elements are). Another important factor is stress-related. It seems clitics behave more like full-fledged words when we consider stress patterns of present-day EP. In the case of certain inflectional forms, object clitics are inserted in defiance of the maximally three-syllable frame (Luís, Kaiser 2016: 214). In EP, word stress travels as far as the third syllable from the right. This rule holds for the inflected forms in verbal morphology, as well (*entregávamos* 'we were handing over' 1.PL imperfect indicative). Interestingly, this locus is maintained irrespective of whether enclitics attach to the forms with lexical stress on the antepenultimate. If enclitics were really the result of a lexical (word-internal) attachment, the result would be an onward shift of the word stress. Yet, such a shift does not take place (*entregávamo-vo-los* '(we) were handing it over to you.PL'.

2. EP interpolation: Historical overview

In defiance of what has been stated in the first criterion above (exclusive attachment of clitics to a verb form), the discontinuity of clitic-verb sequences is a long-standing pattern in EP. (This feature occurs with varying degrees of regularity all over the Ibero-Romance family, being attested in present-day Galician and in Old Spanish; see Bouzoutia, Kempson 2006: 253, Martins 2014: 42–44). Previous proposals of the chronological outline of interpolation are not entirely parallel to the common divisions of EP history (Castro 2006: 158–166, 188–194), where four main periods are identified: i) Old Portuguese (from the first text published in 1214 to approx. 1420); (ii) Middle

Portuguese (from 1420 to 1550); (iii) Classical Portuguese (from 1550 to the late 1700s); and (iv) Modern Portuguese (from approx. 1800 onwards).

In some of the previous research on interpolation, attention has been paid mainly to the medieval and Renaissance period (Fiéis 2001). These studies demonstrate that at the beginning of EP scriptural tradition, interpolation was a widespread phenomenon. What is more, at first, virtually any expression had the ability to follow preverbal pronouns, including the items that only rarely precede the verb (e.g. infinitival complements in multi-verb series or adjectival predicates; see 7a and 7e below; Martins 2016b: 421–426). This would later change and *não* would become the only eligible item to replace X in cl-X-verb sequences, but this initial, more permissive stage is going to be referred to as Phase I in the history of EP interpolation. Another characteristic of this period is that alongside single-constituent interpolations, multiple interpolations are documented, as well (Martins 2014: 42–44). Some of them would feature as many as three distinct syntactic constituents. This indiscriminate interpolation model would hold up until the mid-17th century, being represented in both literary and official texts consistently, though not evenly.

As described above, in this initial period there were hardly any constraints on the morphological status, syntactic function, and length of linguistic signs able to fill the X slot in the cl-X-verb sequence. However, it was not entirely a free-for-all. For example, left-periphery items are known to have been disallowed from appearing in medieval and Renaissance cl-X-verb sequences (Raposo 2000: 278–279). Likewise, the order of constituents in multiple interpolations was not entirely random. If one of the elements following a clitic pronoun was a subject, it would always come first within the sequence (7a, 7b and 7c). Correspondingly, if *não* appeared in multiple interpolations, it always occurred last, thereby directly preceding the verb (Fiéis 2001: 205–206; see 7a, 7c and 7d).

(7) a. *E sse **as nos comprar nõ quisermos** entõ uẽderdelas uos* (1329; example by Martins 2016b: 421)
     And if ACC.3.F.PL we buy.INF NEG want.FUT.SBJV.1.PL, …
     'And if we do not want to buy them, then it is you who must sell them'

b. … *deſpidio logo Cogeçofar vm capitaõ com quinhentos Turcos,*
*que **lhe Elrey de Zebit tinha** mandados de Meca, com regimen-*
to que ſe foſſem meter na cidade de Diu … (1614. Diogo do Co-
uto. *Decada sexta da Asia: dos feitos que os portugueses fizeraõ*
no descobrimento dos mares ... Cap. VI, col. 2; p. 43) http://purl.
pt/29503

with five hundred Turks, REL DAT.3.SG the-King de Zebit AUX send.
PST.PTCP.M.PL from Mecca

'afterwards Cogeçofar dispatched a captain with five hundred
Turks that had been sent to him from Mecca by King Zebit, with
the order to go and seize the city of Diu'

c. Cogeçofar lhe tornou a mandar dizer (…) & que quando ***lhe elle***
***naõ quiſeſſe** diffirir a ellas,* … (1614. Diogo do Couto. *Decada*
sexta da Asia: dos feitos que os portugueses fizeraõ no descobri-
mento dos mares ... Cap. VII, col. 2; p. 50)

… COMP when DAT.3.PL he.NOM NEG want.IPFV.SBJV.3.SG grant.INF
to they.F.PL

'Cogeçofar insisted that he should be told once again (…) that
when he did not want to concede it to them …'

d. ..., & *os juizes q̃ **o aſſi naõ cõprirẽ** pagaraõ por cada ves mil reis,*
... (1639. *Regimento dos Juizes das Aldeas, e Julgados do Termo*;
p.3) http://purl.pt/30213

ART judge.PL REL ACC.3.M.SG that-way NEG fulfill.FUT.SBJV.3.PL
pay-FUT.IND.3PL ….

'and the judges who do not observe it in this way will pay a tho-
usand reis each time'

e. ... *todollos adubyos que lhes **conpridoiros e neçesareos** forem*
(1476; example by Martins 2001: 459)

all.M.PL fertilizers REL DAT.3.PL accomplish.FUT.PTCP.M.PL. and
necessary be.FUT.SBJV.3.PL

'all the fertilizers that might be useful and necessary to them'

In Phase I, the *não* negator is known to have outnumbered by far any other
type of items, exceeding 90 per cent of possible contexts. As for diphasic dif-
ferences, there is evidence for this *não* preference to be better represented in
official documents as compared to literary texts. Moreover, estimations show

a slight yet systematic decrease in the frequency of the cl-X-verb model (with both *não* and the remaining expressions) from the 14th until the end of the 16th century (Martins 2016b: 422). Eventually, the decrease resulted, by the middle of the 17th century, in the *não* negator being left as the only element eligible to break the continuity of clitic-verb sequences. Given these shifts, multiple interpolations were slowly discontinued in Phase II as well.

The beginning of Phase III of the history of EP interpolation is dated to some of the literary texts (theatre plays and narratives) written in the first half of the 19th century. They are innovative in that the interpolation is no longer carried out exclusively with the aid of *não*, even though *não* continues to be privileged (Magro 2010: 116–117). Likewise, multiple interpolations become attested again. Yet, rather than being practised, as in Phase I, with any type of constituents, modern interpolation is constrained by the semantic profile of inserted elements. Aside from *não*, only some deixis carriers are allowed. 'Deixis carrier' refers here to expressions that are person-related (see 8c), space-related (8a), time-related (8b), as well as mood- and aspect-related (Martins 2016b: 423). Moreover, after a period of popularity in literary writings, modern-style interpolation seems to have changed its sociolinguistic status. At present, it is found either in dialectal spoken production (Magro 2010: 117) or, less frequently, in newspaper releases (Martins 2016b: 425). Although there are varying judgements as to its acceptability, its frequency has been seriously diminished. Also, unlike in Phases I and II, there is no clear-cut preference for 3rd person direct object pronouns (see below).

(8) a. *Agora porem que até a minha pobre bibliotheca já **se ahi vai rareando** e desfazendo vendida* (1837. António Feliciano de Castilho, *A Primavera*; p. 11)

   … library already REFL.3.SG there AUX become-rarer.PRS.PTCP and undo.PRS.PTCP sell.PST.PTCP.F.SG

   'Still, at present when even my poor library is vanishing, being taken apart and sold off'

   b. *Este desvio, por onde **me agora deixava** ir …* (1837. António Feliciano de Castilho, *A Primavera*; p. 24) http://purl.pt/29

   This diversion, by where.REL REFL.1.SG now let.IND.IPFV.1.SG go-.INF

   'This detour that I would allow myself to take now'

c. Alfageme: ... *oh*! *Se **o eu soubera**, se **o eu adivinhasse** ...* (1842.
   Garret Almeida. *O alfageme de Santarém ou a espada do Conde-
   stável* ... Acto IV, Scena VIII; p. 109) http://purl.pt/53
   If ACC.3.M.SG I.NOM know.1.SG.PPRF, if ACC.3.M.SG I.NOM guess-1.
   SG.PPRF
   'If only I knew that, if only I could guess that'

The evolution outlined in this section sets the scene for the question why,
starting from approx. mid-19ᵗʰ century, cl-*não*-verb sequences began to re-
cede. On the one hand, at that time *não* still evinced substantial frequency
in this setting (see Table 1 below for calculations). On the other hand, new
expressions became allowed that were able to fill the X slot in the cl-X-verb
sequence. It would seem that the emergence of this new quality should not
translate into more instances of clitic-verb adjacency. And yet, from that pe-
riod on, proclisis has started to take over, up to the point where it relegated
interpolation to an idiolectal and stylistically marked choice. No convincing
explanation of the relative abruptness of this shift has been proposed thus far.

Let us now briefly turn to the question why, over the course of history,
it was *não* that tended to disrupt clitic-verb sequences at a scale larger than
any other constituent. One reason may be its syllabic structure; that is go-
ing to be discussed in more detail in Section 4. Moreover, semantic proper-
ties of the negative marker are presumed to come into play. It has been re-
peatedly stressed in the literature (Dryer 1988: 99, 102) that negation carries
a heavy communicative load. It constitutes an essential part of the message:
if a hearer fails to recognize it in the flow of an utterance, the intended mean-
ing of a sentence is fundamentally misunderstood. Unlike in traditional logic,
where negation affects the entire clause, in natural language subjects are gen-
erally left out of its scope (Zanuttini 2001: 511). Instead, the proper domain
of negation is the verb-object combination. Moreover, typological research
demonstrates that semantically interrelated units tend to surface next to each
other within sentences. That is how the final position of *não* in multiple inter-
polations can be accounted for.

As for the syntactic principles underlying the adjacency of the *não* nega-
tor and the verb, they are perhaps best captured with the concept of c-com-
mand. The latter is a configurational schema accounting for the relationship
between nodes A and B in the following terms: A c-commands B if and only

if (i) neither A dominates B nor B dominates A, and (ii) the first node dominating A also dominates B (Matos, 2003: 774). If applied to *não*-verb sequences, c-command states that: (i) the verb is not syntactically dependent on *não*, (ii) *não* is not syntactically dominated by the verb, and (iii) both of them have their superordinate category in common, namely the VP.

3. The analysis of corpus data

Our corpus, compiled of 39 texts released between 1614 and 1858, contains at present 659 examples of both cl-*não*-verb and *não*-cl-verb sequences (see below on the equivalent cases covered by the latter model). All of these examples have been retrieved manually. This is because the texts available in the *Biblioteca Nacional Digital* are editable only to a limited degree and no automatic search can be conducted. The grand total of recorded clitic-verb combinations in all the texts, spanning 243 years, is 11.072. This means that the choice between cl-*não*-verb and *não*-cl-verb covers almost 6 per cent of all instances of clisis.

Table 1. Phase II: Chronological frequency
of interpolation compared to clitic-verb adjacency

	Sequences with clitics	cl-*não*-verb vs. *não*-cl-verb or *não*-V_{fin}-V_{inf}-cl (proclisis trigger assumed on the left of the sequence)	Interpolation in possible contexts (per cent)
17th century	3.378	156:24 (180)	86.66 %
18th century	4.701	224:97 (321)	69.78 %
19th century	2.993	100:58 (158)	63.29 %
Total	11.072	480:179 (659)	72.83 %

Source: Authors.

Now, the analysis guidelines and the achieved results are going to be discussed in more detail. First, the selection of texts was narrowed down so as to take into account only those where interpolation relied exclusively on the *não* negator. In this way, the chronological frames of Phase II could be traced with more accuracy. The last identified text, Bento Teixeira's chronicle, which does not comply with our selection criterion (with two items, *melhor* 'better' and

bem 'well' breaking the continuity of clitic-verb; see 9a-b) dates from 1650. Consequently, the chronicle had to be left out of the analysis even though in many earlier 17th-century texts, the *não* interpolation is exclusive.

(9) a. ..., *por ſeu grande talento, & valor, com que ſempre militou na In-*
         dia, achando-ſe em ocaſiões de guerra, em que **o bem moſtrou**,
         antes de entrar na Religiaõ. (1650. Bento Teyxeyra Feyo. *Relaçam*
         do naufragio que fizeram as naos ...; p. 41)
         … in which ACC.3.M.SG well show.PRF.IND.3.SG, before enter.INF
         in-ART.DEF.F.SG religion
         'by his extraordinary skill and valor, with which he always waged
         wars in India, finding himself in struggles where he could display
         his qualities, before entering the convent'
     b. ..., *tomando o q̃* **lhe melhor pareceo** *com grande feſta* (ibid; p. 48)
         http://purl.pt/28100
         take.PRS.PTCP REL DAT.3.SG better seem.PFV.3.SG with great feast
         'taking what seemed better to him with great joy'

The results of our corpus analysis also indicate the closing boundary of Phase II, thus pinpointing when the usage of *não* for interpolation started to lose its exclusivity. The first attestations of the present-day 'deixis-oriented' template have been found in a text from 1837. This turning point is *A Prima-vera*, a narrative by António Feliciano de Castilho (see 10a-b). Consequently, the narrative had to be left out, even though in many texts published in the following years (up to year 1858 in our corpus), *não* is exclusive in cl-X-verb sequences.

(10) a. ..., *uma sombra dos interiores commodos que* **me lá ficarão** *com*
         a poesia natural e singela? (António Feliciano de Castilho, *A Pri-*
         mavera; p. 18) http://purl.pt/29
         REL DAT.1.SG there rest.FUT.3.PL with DEF.F.SG poetry natural and
         unpretentious
         'the comfortable interiors which will stay with me along with
         a pure and unpretentious poetry'

b. *Melhor he hoje do que então era; não porque **o eu tornasse** á forja e á bigorna*, ... (ibid., *A Primavera*; p. 27)

not because ACC.3.M.SG I render.IPFV.SBJV.1.SG to-ART.F.DEF.SG smithy and to-ART.F.DEF.SG anvil

'not because I would give it back to the smithy and anvil'

Second, aside from the competition between interpolation and proclisis, the statistics in Table 1 take into account the *não*-V$_{fin}$-V$_{inf}$-cl sequences, i.e. the ones where clitic climbing fails to materialize. Clitic climbing takes places when pronouns attach to a verb other than the one that subcategorizes for the argument they specify (Martins 2000: 169–171). Apart from being semantically interchangeable (see 11a-c), the two models are alike in that they appear in EP when the finite verb and its non-finite complement are aligned. In EP, clitic climbing is an archaic feature (Martins 2016a: 19-20). Nonetheless, in our corpus, it is still an extant option, though less frequent than enclisis on non-finite verbs.

(11) a. *E tratando de fugir ao noſſo caſtigo, **o naõ puderaõ eſcuſar**, ...* (1625. Bertolameu Guerreiro. *Jornada dos vassalos da coroa de Portugal, pera se recuperar a cidade do Salvador, ...* CAPITVLO XXXXII; p. 134) http://purl.pt/17352

And trying to flee our punishment, ACC.3.M.SG CAN.3.PL.PFV forgive.INF

'And trying to avoid the punishment we would inflict on them, they could not forgive it'

b. *.., como vivem as crianças no ventre materno, ſem reſpirar; e que **naõ os poderáõ matar** ...* (1759. Pedro Norberto de Aucourt e Padilha. *Raridades da natureza, e da arte, divididas pelos quatro elementos*. Parte III.§.VI.; p. 376) http://purl.pt/13915

... and REL NEG ACC.3.M.PL can.FUT.IND.3.PL kill.INF

'… and that they will not be able to kill them'

c. *..., lhe foy conveniente depoſitála por dez dias no Convento do Prado de S. Vincente da Ordem dos Prégadores, do qual **naõ pode retirála** ...* (1701. *Epitome da portentosa vida, e milagres de Sta. Catharina de Ricciis, ...*; p.2) http://purl.pt/346

… out of-DEF.M.SG REL NEG can.3.SG.PFV withdraw.INF-ACC.3.F.SG
'…to leave her for ten days in the Dominican convent, out of which
he could not withdraw her'

If cases of enclisis on non-finite verb forms are disregarded, the absolute number of appropriate contexts suffers a noticeable reduction. As a result, the percentages tilt in favor of interpolation. The tables below depict calculations for two most frequent types of pronouns – *o, a, os, as* 3rd person direct objects and *se* – in appropriate settings.

Table 2. Occurrences and frequency of interpolation of 3rd person direct object
and reflexive clitics to cases of clitic-verb adjacency
(proclisis and enclisis on non-finite verb combined)

39 texts (1714–1858)	Occurrences total	cl-*não*-verb	*não*-cl-verb	*não*-verb-verb-cl
o, a, os, as	137	112 (81.75%)	18 (13.13%)	7 (5.1%)
se	333	227 (68.16%)	84 (25.22%)	22 (6.6%)

Source: Authors.

Table 3. Occurrences and frequency of interpolation of 3rd person direct object
and reflexive clitics to cases of clitic-verb adjacency (proclisis only)

39 texts (1714–1858)	Occurrences total	cl-*não*-verb	*não*-cl-verb
o, a, os, as	130	112 (86.15%)	18 (13.84%)
se	311	227 (72.99%)	84 (27.01%)

Source: Authors.

Third, our corpus data appears to disconfirm some of the beliefs prevalent in previous writings on Romance clitic grammar, where interpolation is presented to be mandatorily triggered. This representation has some observational basis and seems to rely on the following reasoning: if proclisis and interpolation can be applied interchangeably, and if proclisis must be triggered in EP, the same must happen to interpolation. In more technical terms, a special linguistic sign called *exordium* is needed to force pronouns into a preverbal position (Hinzelin 2010: 332–335), whether verb-adjacent or not. This as-

sertion, however, is not supported by our data. In fact, our corpus does have attestations of non-triggered interpolations. Below are some of the examples (the first one does not belong to the corpus).

(12) a. *E tratando de fugir ao noſſo caſtigo,* **o naõ puderaõ eſcuſar**, ... (1625. Bertolameu Guerreiro. *Jornada dos vassalos da coroa de Portugal, pera se recuperar a cidade do Salvador*, ... Cap. XXXXII; p. 134) http://purl.pt/17352

   And trying to flee our punishment, ACC.3.M.SG CAN.3.PL.PFV forgive.INF

   'And trying to avoid the punishment we would inflict on them, they could not forgive it'

   b. *O Conde da Ribeyra* **ſe naõ deſcuydava** *ſem deſcançar em prevenir quanto era neceſſario para taõ grande occaſiaõ.* (1714. *Relaçam da campanha de Alem-Tejo no Outono de 1712*; p. 41) http://purl.pt/26483

   DEF.M.SG count da Ribeira REFL.3 NEG neglect.3.SG.IPFV.IND without take-a-rest.INF

   'Count da Ribeira was attentive and tireless in preparing everything that was necessary for such a special occasion'

   c. *...; e como os Pruſſianos ſe achavaõ divertidos com a inumeravel multidaõ de outros inimigos, que por todas as partes os acometiaõ,* **lhes naõ era** *poſſivel fazerlhes reziſtencia* (1758. *Noticia da grande vitoria, alcançada pelos prussianos contra russianos*; p. 3). http://purl.pt/20902

   REL from all.F.PL DEF.F.PL side.PL ACC.3.M.PL assault.IPFV.IND.3.PL DAT.3.PL NEG be.IPFV.IND.3.SG possible make-INF DAT.3.PL resistance

   'since the Prussians' attention was diverted by large crowds of other enemies assaulting them from all sides, they were unable to put up resistance'

The genesis of these structures lies in the persistence of proclisis in some of the syntactic positions where enclisis is obligatory nowadays. Throughout Phase II, proclisis appeared (though not exclusively) in the following situations: (i) V1 (verb first) root clauses following a subordinate clause; (ii) V1

coordinate clauses (introduced by the conjunctions *e* 'and', *mas* 'but', *pois* 'then'); and (iii) root declaratives with an overtly expressed preverbal subject. In the latter case, proclisis was the default placement no matter what the subject type was (be it a subject pronoun, a proper name, or a common noun) and whether it was definite or not. Yet, despite the dominance of preverbal placement, enclisis was gradually gaining ground. Its increasing popularity is best witnessed by the texts where both preverbal and postverbal pronouns in the same V1 structural environments were applied alternately. Examples 13 and 14 below demonstrate fluctuations in clitic placement across contexts (i) and (ii) above.

(13) a. *Poſſuindo o meſmo Imperio Tiberio Ceſar, **ſe ſubmergiraõ** em huma noite na Aſia doze Cidades pela violencia desſte Meteóro.* (1748. Relação do formidavel, e lastimoso terremoto succedido no Reino de Valença ...; p.4)

possess-PRS.PTCP ART same empire Tiberius Caesar, REFL.3.PL sink.PRF.IND.3.PL in one night

'Under the reign of Emperor Tiberius Caesar, in one night twelve cities in Asia were sunk by this meteor'

   b. *Perſeguindo os Colonnas ao Papa Bonifacio VIII **levantou-ſe** por muitos dias em Italia hum taõ grande tremor de terra.* (1748. Relação do formidavel, e lastimoso terremoto succedido no Reino de Valença ...; p.4) http://purl.pt/347

follow.PRS.PTCP ART colum.PL to-ART.DEF.M.SG pope Boniface VIII rise.PRF.IND.3.SG REFL.3.SG for many.PL day.PL

'In the wake of the throngs trailing to Pope Boniface VIII, a huge earthquake struck Italy, lasting many days'

(14) a. *Outras reformamſe, e **ſe lhes diminuem** as muitas liberdades* (1758. Luís António Verney. *Resposta as Reflexoens que o R.P.M. Fr. Arsenio da Piedade Capucho fez ao Livro* ... Reflexam I. *Da Qualidade do Autor*; p. 9)

Other.F.PL retire.PRS.IND.3.PL-REFL.3.PL and REFL.3.PL DAT.3.PL diminish ART many.F.PL liberty.PL

'The others retire, and many of their freedoms dwindle'

   b. *Enſinalhe os preceitos Retoricos dos eſtilos, &c. e **provalhe** tudo com os exemplos de S. Joam Crizoſtomo* (1758. Luís António Ver-

ney. *Resposta as Reflexoens que o R.P.M. Fr. Arsenio da Pieda-*
de Capucho fez ao... Reflexam VI. *Da Retorica*; p. 31) http://purl.
pt/24753
teach.PRS.IND.3.SG-DAT.3.SG ART rule.PL rhetorical of-ART.DEF.M.PL
style.PL and prove.PRS.IND.3.SG-DAT.3.SG with ART example.PL
'He teaches him the principles of rhetoric and style and shows him
everything on examples from St. John Chrysostom'

In this situation of concurrence, the speakers were left to their own devic-
es as to the rules deciding the correct clitic placement. Now it becomes appar-
ent how these complex sequences originated, some consisting of as many as
three elements – *não*, the pronoun, and the verb. It follows that if no special
exordium was required for an ordinary proclisis to take place, various linear
orderings were allowed in sequences with the *não* negator.

Additionally, in the older stages of EP, subordinate clauses were occasion-
ally introduced with no overt *que* complementizer[2]. The omissions reached
such a frequency that they carried over into negated clauses. In such con-
texts, interpolations may have appeared even when an overtly expressed pro-
clisis trigger was absent (15 and 16). However, this case cannot be equated
with 'out-of-the-blue' genuine interpolations and requires a separate treat-
ment elsewhere.

(15)   ..., *cõ proteſtos, lhe foi requerido **o não fizeſſe*** ... (1644. *Relaçam em*
       que se refere parte dos gloriosos successos, que na Provincia da Be-
       ira tiverão contra Castelhanos, as armas de S. Magestade, ...; p. 3)
       http://purl.pt/12510
       ... DAT.3.SG AUX-be.PRF.IND.3.SG require.M.SG.PST.PTCP ACC.3.M.SG
       NEG do.IPFV.SBJV.3.SG
       'With protests he was required not to do that'

(16)   *Já ponho ponto, & ſinto **me naõ occorreſſe** de antes*, ... (1736.
       Fr. João Manoel. *Vaticinio exposto, confirmado, e defendido: expo-*
       sto á Universidade de Coimbra ...; p. 22)

---

   [2] To the best of our knowledge, the phenomenon has not thus far attracted the atten-
tion of scholars; an apparently similar structure has been extensively dealt with by Artea-
ga (2009: 23–27), for Old French; it remains to be investigated whether French and Por-
tuguese omissions were indeed equivalent.

now put.PRS.IND.1.SG point be-sorry.PRS.IND.1.SG DAT.1.SG NEG hap-
pen.IPFV.SBJV.3.SG before
'I get to the point now, and I regret that it did not occur to me before'
http://purl.pt/6819

Fourth, in the corpus, the average ratio of interpolation to cases of clitic-verb adjacency is 2.68:1 (480 vs. 179 occurrences). It would follow that *não* is interpolated in 72.83% of all possible contexts. Yet, in spite of the semantic equivalence of these two models, the data shows they were not fully interchangeable. They differed with respect to the pronoun type they tended to employ: 3rd person direct objects appeared in the interpolation scenario at a greater rate than the remaining pronouns. Table 2 above shows that in the case of this type of pronoun, the interpolation of *não* reached 81.75% (112 out of 137 occurrences). By contrast, the remaining pronouns did not depart significantly from the average ratio, thus showing a more balanced distribution. After subtracting the pronouns *o*, *a*, *os*, *as* as a whole, the numbers amount to 368 occurrences of cl-*não*-verb versus 154 occurrences of *não*-cl-verb, i.e. 70.5% to 29.5%. At first glance, the difference between 81.75% and 70.5% is not a dramatic one. However, it translates into 4.48:1 in favour of interpolation for the *o*, *a*, *os*, *as* pronouns, and 2.39:1 for the remaining pronouns. Two questions that naturally emerge at this stage are: Why were particular clitics favoured in particular contexts? How do we account for the fact that direct object pronouns stand out as more frequent?

In what follows, interpolation is going to be discussed mainly in connection with the *o*, *a*, *os*, *as* series. The first clue to their overrepresentation may well lie in the syllabic structure of *não* and direct object pronouns. The former ends in a branching nasal [ãw̃] nucleus with no coda, whereas the pronouns consist only of a rhyme with a zero onset. Throughout the history of EP, encounters of verb-final nasal diphthongs with postverbal 3rd person pronouns are known to be the loci of allomorphic variation[3]. In the case under

---

[3] For some scholars, the conditions under which 3rd person direct object pronouns take the [n]-onset are open to debate. For example, Luís & Kaiser (2016: 221) deny the relevance of the 'final nasal diphthong' criterion. They trace this kind of allomorphy to the configuration of features '3rd person' and 'plural' of the verb form, claiming that empty onset enclitics are preferred if this requirement is not met: *Tem-os* (*tem-nos) *visitado* 'has been visiting them'.

discussion, the [n] onset is added to the pronoun. In present day standard EP, the change is restricted to verb-clitic sequences. Even if the same sound material occurs in expressions other than clitic pronouns (e.g. definite articles, prepositions), allomorphy fails to occur.

(17) a. *Viram-nas* (**Viram-as*)
 see.PRF.IND.3.PL ACC.3.F.PL
 '(They) saw them'
 b. *Viram as quebras* (**Viram-nas quebras*)
 see.PRF.IND.3.PL ART-DEF.F.PL cracks
 '(They) saw cracks (in it)'
 c. *Viu-as* (**Viu-nas*)
 see.PRF.IND.3.SG ACC.3.F.PL
 '(He) saw them'

4. The interpretation of the data: Moving along the constraints

The results of the corpus analysis are indicative of the close connection between direct object pronouns and interpolation. Now, it remains to be seen how these results advance our knowledge of the mechanisms behind the shift from Phase II to Phase III.

In modern EP, hardly any phonological change can be observed in preverbal position, and such as it is is not clitic-specific. Irrespective of the category of the segments being juxtaposed, the phonological change is a strategy that enables speakers to get rid of a hiatus. This change involves glide-insertion in the middle of sequences made of [ə] and the following vowel, whether word-internally or at word boundaries. As far as clitic-verb sequences are concerned, [j] is only optionally inserted (e.g. *me ofereceram* '(they) offered me' either assuming the form [məɔ] or [mjɔ]). Likewise, a phonological liaison of this kind may occur after function words followed by 3rd person direct object pronouns (*se a vires* 'if (you) see her', either surfacing as [səɐ] or as [s(ə)jɐ]). The optionality of glide insertion is assumed to speak to the phrasal (postlexical) phonology of preverbal clitics (Luís & Kaiser, 2016: 220).

Yet, before the advent of Phase III, preverbal direct object pronouns were phonologically tied to the preceding non-verbal sound material on a much wider scale. All of their non-verbal hosts shared the ability to make pronouns

preverbal (proclisis triggers). Examples follow in 18a–b and 19a–b, below. They are taken from Diogo Couto's (1614) chronicle (the text is representative of Phase I).

(18) a. ... *deſcuidandoſe por entaõ da ciſterna, **pello permitir** Deos noſſo Senhor aſsi, ...* (1614. Capítulo VI, coluna 2; p. 43)
disregard.PRS.PTCP for then of-ART.DEF.F.SG cistern, because--ACC.3.M.SG allow.INF God our Lord so
'forgetting the cistern for a while, since the Lord our God allowed it so …'

b. *… & ao Ruy Freire encomendou de palaura, que trabalhaſſe por lhe tornar com a repoſta, por que importaua muito, & que lhe faria merce, **pello segurar**.* (1614. Capítulo VI, coluna 2; p. 47)
…, for-REL much (and) DAT.3.SG do.COND.3.SG favour, because ACC.3.M.SG guarantee.INF
'He issued a word-of-mouth order to Rui Freire to make an effort and return with the answer he was so anxious about, and to do him a favour of keeping it safe'

(19) a. *..., mas ſairaõlhes os capitaẽs Portugueſes, & Eſpanhoes **aos recolher**, o que naõ poderaõ fazer ſem ſe trauarem cõ os imigos …* (1614. Capítulo VI, coluna 1; p. 42)
… but go out.PRF.IND.3.PL-DAT.3.PL ART captain.PL Portuguese and Spanish at-ACC.3.M.PL collect.INF
'but the Portuguese and Spanish commanders went in front of them, carrying out the sick, which they could not do without crossing paths with the enemy'

b. *…, & que elle o mouera **ao vir** auiſar de hũa grande traiçaõ que lhe eſtaua ordenada …* (1614. Capítulo VI, coluna 1; p. 46)
COMP he.NOM move.PPRF.3.SG.IND at-ACC.3.M.SG warn of INDF.F.SG big treason
'that he spurred him to come and warn him of an imminent betrayal'

What these examples demonstrate is preverbal pronoun allomorphy, which is believed to bear witness to the morphological attachment of the preverbal clitic to a preceding proclisis trigger. 3rd person direct object clitics are pho-

nologically integrated into the items on their left (as are enclitics). 'Phonologically integrated' stands here both for the shifts in the preposition (*por o* turns into *pelo*) and for the [V2] semi-vocalization undergone by the pronoun in the *ao* sequence (phonologically [ɐw])[4]. If taken separately, the preposition *a* and the pronoun *o* in EP surface as [ɐ] and [u] respectively. Therefore, morphological attachment extended over preverbal clitics in this particular context (after prepositions *por* and *a*) in Phases I and II.

As for the other elements that interacted with direct objects, the *não* negator is one of the best documented and important ones (see 20a–c). Likewise, the second member of correlative negative coordinate structures (Sánchez López 2017: 649–650; Matos 2003: 772) *não… nem* 'neither… nor' allowed for allomorphy in preverbal direct object clitics (see example 21c).

(20) a.   … ***naõ no fazendo*** *deſde o dia que lhe puzerem a dita pena* …
          NEG ACC.3.M.SG do.PTCP.PRS since DEF.M.SG day … (1639. *Regimento dos Juizes das Aldeas, e Julgados do Termo*; p. 6)
          '… not doing it since the day he will be given this punishment' (*o* evolves into *no* under the influence of the nasal nucleus in *não*) (http://purl.pt/30213)

      b.   …, *p(er)o deffendemos q̃ o marido d(e)poys que o soub(er) q̃ sa molh(er) fez adulterio **nõna tenha** a sa mesa nẽ a seu lecto e o q̃ o nõ fez(er) **nõna possa** depoys accusar* … (13th century. *Foro Real*. Livro 4, Capítulo 7, Fólio 136r) Source: CIPM http://cipm.fcsh.unl.pt/
          'COMP POSS.F.3.SG wife do.PFV.3.SG adultery NEG-ACC.F.3.SG have.SBJV.PRS.3.SG …. NEG-ACC.F.3.SG can.SBJV.PRS.3.SG afterwards accuse.INF'

---

[4]   This process, known as [V2] semi-vocalization, operates in modern EP in word final position (*pediu* [iw] '(he) asked', *pneu* [ew] 'tyre', *céu* [sɛw] 'sky'). It takes place when out of two contiguous vocalic segments – [V1] and [V2] – the second is a round vowel. Except for clitic items (*vi-o* [viw] '(I) saw it'), [V2] semi-vocalization does not occur if the two vowels belong to different words (*táxi usado* [*iw] 'worn out taxi'). [V2] semi-vocalization seems to be especially favoured if [V1] is an *i* or an *a*, and less so if it is an *e* or an *o*. All in all, this sound change has the properties of a post-lexical process: (i) it is general (not clitic-specific); (ii) it can apply optionally (*rio* [ʁiu] / [ʁiw] 'river'); (iii) and it affects enclitics (Vigário 2003: 99–102).

'But we maintain that the husband, after he learns that his wife committed adultery, should not allow her by his table or into his bed, and if he does, he cannot accuse her afterwards'

c. ... *se p(er)ventura se p(er)der ou ouerẽ subr'elha algua dolta ou duuida possa seer prouado pella nota ond(e) foy sacada [e] aquella **nõna mostre nẽna tena** en logar u a ueya nenhũa das partes* (13ᵗʰ century. *Foro Real.* Livro 1, Capítulo 8, Fólio 79v) Source: CIPM

... DEM.DIST.F.SG NEG-ACC.3.F.SG show.SBJV.PRS.3.SG nor-ACC.3.F.SG have.SBJV.PRS.3.SG in place where-REL ACC.3.F.SG see.SBJV.PRS.3.SG either of-DEF.F.PL side.PL

'If by chance it is lost or any doubt over it arises, it must be proven by a note where it was taken from, and this note can neither be shown publicly nor stored in a place where either side could see it'

Instances of allomorphic variation in preverbal pronouns, as described above, occur unsystematically, yet persistently, in the texts from Phases I and II. As a matter of fact, not every case of *não*, *nem*, and *a* would cause pronouns to undergo phonological shifts. Thus, it seems that the decision as to whether pronouns should be phonologically linked to the preceding non-verbal lexemes was left to speakers' discretion.

(21) a. ... *e eſtas **nem as hà**, **nem as pòde** haver ſem meyos* (1672. Antóno Vieira. *Carta CXX. Ao Marquès de Gouvea*; p. 437)

    and DEM.F.PL nor ACC.3.F.PL be.PRS.IND.3.SG nor ACC.3.F.PL can.PRS. IND.3.SG be.INF without mean.PL

    'And these are not there and neither can there be any without means'

b. ..., *ſoccorria às neceſsidades de **quem o auiſauão**.* (1614. Nicolao Agostinho. *Relaçam summaria da vida do ... Senhor Dom Theotonio de Bragãça*. Cap. XIII; pág. 113)

    guarantee.IPFV.IND.3.SG to-DEF.F.PL need.PL of who ACC.3.M.SG warn.IPFV.IND.3.PL

    '… he filled the needs of those who would warn him' http://purl. pt/11490

Now, coming back to the list of allomorphy-inducing expressions, we see pronouns combining, again, in a completely unpredictable manner. They select their left-hand side host independently of the host's category membership and of its syllabic structure. In fact, apart from *por* and *a*, EP has a series of one-syllable prepositions (proclisis-triggers) whose phonological properties are at first sight capable of triggering pronoun allomorphy, but which do not perform according to expectations. Even though *em* [ɐ̃j̃] 'in' and *sem* [sɐ̃j̃] 'without' are monosyllabic and though the second of them ends in a nasal diphthong, written records offer no attestation of their being hosts for preverbal pronouns. In other words, sequences like *sem na ver 'without seeing her', *se no bateres 'if you hit him', etc. are missing. This *ad hoc* selection lends further support to Zwicky and Pullum's observation that morphophonological idiosyncrasies are found more easily in affixed words than in clitic groups.

(22) a. *Naõ cuidey que havia tanta peçonha na noſſa terra, nem tanta induſtria em **a lançar** fóra* ... (1672. António Vieira. Carta CVIII. Ao Marquès de Gouvea; p. 405)
Nor so-much.F.SG cleverness in ACC.3.F.SG throw.INF outside
'I've never thought that the world was so full of venom and that people had so much skill in spitting it out'

b. ..., *a caſa do Arcebiſpo a negocio para que elle o tinha mandado hir, ſocedeo entrar pella caſa a tempo, que não vio criado algum nella para dar recado ao Arcebiſpo & ſem **o ſaber** entrou no apoſento*, ... (1614. Nicolao Agostinho. *Relaçam summaria da vida do ... Senhor Dom Theotonio de Bragãça*. Cap. XIV; p. 124)
… without ACC.3.M.SG know.INF in-DEF.M.SG chamber
'He happened to come in at a time when no one was there to see him to let the Bishop know, and he got into the room unawares'

All in all, rather than being dictated by syntax or phonology, the selection of hosts for preverbal pronouns appears to be an entirely random matter of lexicon. Speakers had to rely on rote learning rather than on a specific rule which would state how the clitics *o*, *a*, *os*, *as* should be inserted in preverbal environments and what form they should take. Obviously, this situation produced very unsteady results. For want of an internalized, fixed list of allomorphy-inducing lexemes at their disposal, various authors came up with dif-

ferent provisional solutions. This is a plausible explanation of why divergent realizations – *quem o* / *quem no*, *não o* / *nãono*, *ao* / *a o* – were so persistent in EP's history[5].

In such circumstances, allomorphy-avoidance[6] in preverbal positions might have been prompted by the systematic interpolation of *não* in the 17th and through the 18th century. At least that is how we interpret the prominence of *o*, *a*, *os*, *as* in cl-*não*-verb sequences as documented in our corpus. All of the remaining pronouns, including clitic clusters, are less troublesome and do not run the risk of undergoing allomorphic variation. It comes as no surprise, then, that they show up with less frequency in cl-*não*-verb sequences than 3rd person direct objects. Moreover, semantic properties of *não* are particularly conducive to a verb-adjacent position. The combined strength of all these factors contributed to a situation where interpolation was commonly found, even at the expense of the 'clitic by the verb' principle.

Therefore, each of the three equivalent sequences (e.g. *que o não viu*, *que não no viu,* and *que não o viu*) discussed thus far is associated with the dominance of at least one of the three general constraints organizing EP clitic grammar (see Table 4). These constraints were unlikely to be satisfied all at once. In pursuit of their communicative goals, speakers would privilege one

––––––––––––––––––

   [5] Moreover, the inconsistent architecture of proclisis trigger – pronoun – verb sequences required speakers to develop advanced processing skills. The recognition of suitable lexemes that could force clitics to undergo phonological shifts is tricky enough, and then there is the added difficulty of the homonymy of definite articles and 3rd person direct object pronouns. (The latter still poses a challenge to foreign students of EP). The evidence in our corpus would suggest that authors did not agree about how these two categories should combine and how the [n]-onset forms should be distributed in preverbal syntactic configurations.

   [6] In diachronic linguistics, the term has been more frequently evoked with reference to paradigm-internal organization. Recently, its relevance has been called into question by Da Tos (2013: 64), suggesting that the evolution of inflectional paradigms is not goal-oriented ('there is nothing teleological in it'). Yet, the very act of choosing between freely interchangeable variants; e.g. EP 1.SG.PRES.IND *oiço* and *ouço* '(I) can hear'; Fr. 1.SG.PRES. IND *m'assieds* and *m'assois* '(I) sit down') is exclusive, meaning the remaining variant(s) fail(s) to occur. Thus, 'allomorphy-avoidance' may simply be a mere by-product of the preference for a regular form. Usage-based theories of grammatical change, to which we subscribe, stress the importance of repetitive use of a certain variant in producing a new pattern. If some critical number of the variant occurrences is attained, a durable change may eventually come about. Momentary developments along these lines can hardly be called purpose-driven.

or two of the three constraints. The other they would play down or ignore altogether.

Table 4. Characteristics of interpolation, standard clitic-verb sequence, and [n]-onset preverbal pronouns with reference to the constraints on clitic placement in EP.

Realization of preverbal 3rd person direct object pronouns	Dominant principle		
	Homogeneous realization in both preverbal and postverbal position	Clitic-verb adjacency	Phrasal attachment in preverbal position
proclisis trigger – *não* – clitic – verb (with clitic allomorphy)	+	+	–
proclisis trigger – *não* – clitic – verb (with no allomorphy)	–	+	+
(proclisis trigger) – clitic – *não* – verb	–	–	+

Source: Authors.

Yet, even the interpolation of *não* did not automatically obviate preverbal allomorphy. In one type of context, preverbal allomorphic variation was maintained regardless. The case in point involves the preposition *por*. In present-day EP, it surfaces in the contracted form of *pelo* (*pela*, *pelos*, *pelas*) if preceding the definite article. In the previous stages of EP's history, the preposition appeared in its contracted form also when followed by 3rd person direct object pronouns. Curiously, our corpus offers no counter-examples to the *pelo não*-verb combination. Put otherwise, in no case did the sequence *por* (preposition) followed directly by 3rd person direct object pronoun and the negative marker surface as *por o*. See the examples of this particular arrangement recorded in our data below.

(23) a.　… *hũa petiçam affinada per duzentos cidadões, em que todos pediað o mofteiro, po**lo naõ auer** neftas partes, ...* (1640. Relaçam verdadeira do milagroso portento; p. 6) http://purl.pt/16740

ask.IPFV.IND.3.PL ART monastery for-ACC.3.M.SG be.INF in-DEM.F.PL side.PL

'a petition signed by two hundred citizens, in which they all asked for a monastery to be built, because there were none in the area'

b. ..., *havendo muitos delles que naõ ouvem miſſa, nem prégaçaõ em todo o anno pola naõ terem* (1653. António Vieira. Carta IX. *A El-Rey ſobre as neceſſidades eſpirituaes do Maranhaõ*)

… in entire ART year for-ACC.3.F.SG NEG have.INF.3.PL

'there being many who never hear Mass or sermon in a whole year, because there are none to be heard'

c. ... *para acabar de deſanimar os inimigos impacientes, os grana-deyros da cortadura, pelos não ter viſto como os mais ſem ordem dos ſeus officiaes*, ... (1714. *Relaçam da campanha de Alem-Tejo no Outono de 1712*; p. 45) http://purl.pt/26483

for-ACC.3.M.PL NEG AUX see.PST.PTCP as ART other without order of-DEF.M.PL POSS.3.M.PL commander.PL

'without having seen them as the remaining ones with no order of their commanders'

d. Clarimundo: *E quanto déra pelo não haver ſido*? (1804. Manuel de Figueiredo *Escola da Mocidade* Comedia Acto III, Scene VII; p. 219) http://purl.pt/11977

and how-much give.PPRF.1.SG for-ACC.3.M.SG NEG to-be.INF

'What I would not give for it not to have happened'

The last attestation in our corpus dates back to as late as 1804. It occurs in a theatre play which clearly imitates spoken language (lively conversation). Yet, in present-day standard EP, the sequence has lost its allomorphic realization, even though interpolation is still found. The examples below have been retrieved from corpusdoportugues.org (the 'Web / Dialects' sub-corpus).

(24) a. ... *como o inquilino pode fazer uma queixa de si nas finanças por o não ter feito.* https://forumdacasa.com/discussion/29746/como-colocar-um-apartamento-para-alugar/

'since the tenant can submit a claim against you in fiscal bodies for having failed to do that'

b. ... *verter o ódio que sentem de si próprios **por o não conseguirem** fazer de viva voz.*
   http://metoscano.blogspot.com/2011/04/de-que-ajuda-precisamos.html
   '… ooze with the hatred they feel towards themselves because they do not succeed in doing it aloud'

Concluding remarks: Allomorphy-avoidance and its side-effects in clitic grammar

The line of argument followed in the present paper is strongly functionalist in its flavour in that it plays on the usefulness of certain quantitative restructurings in grammar for communicative and/or language processing purposes (Waltereit 2012: 14–15). To circumvent the difficulties posed by a given structure, speakers might privilege an alternative one. Yet, usefulness does not equal intent here. Thus, interpolation cannot be viewed simply as a remedy whose aim was to reduce allomorphic variation in preverbal position or do away with the variation altogether.

Moreover, our analysis consistently shows a gradual pace of grammatical change. The gradual pace implies that at least two grammars must be available to the speakers simultaneously. This does not imply, however, that particular grammatical solutions have to be mutually exclusive. Instead, both allomorphic variation in proclisis and interpolation seem to have served their own functions, with each model overriding the other in a separate domain. These separate domains are associated here with the speakers' needs for a determined method of language processing. No matter what pattern is chosen, the same meaning is encoded throughout. Thus, preverbal clitic allomorphy meets the need for the uniqueness of realization (Thornton 2011: 360–361) of a single linguistic sign across different structural environments. If the [n]-onset is added consistently, both in preverbal and postverbal position, 3rd person direct object pronouns are given a homogeneous treatment. Thus, with respect to their host, pronouns become category-blind. The only type of data they remain sensitive to is phonological information.

(25) a. *Pobrezinha, anda sem ter qu**em na** aconselhe* 'Poor thing, walking around with no one to give her advice'

    b. *Quer**em-na** ver* '(They) want to see her'

By contrast, allomorphy-avoidance in preverbal position (i.e. interpolation) helps establish a rigorous distinction between an exclusively phrasal attachment of proclitics and an exclusively morphological attachment of enclitics. This enables speakers to dispense with the need to recognize and deal with the allomorphy-inducing expressions preceding the verb. This is an undeniable advantage given that phonological change affecting preverbal pronouns is not tied to the syllabic structure of proclisis-triggers (as we have seen above, *sem*, *em*, *bem*, etc. do not produce clitic allomorphy). As a result, in present-day standard EP, empty onset 3rd person direct object pronouns are normative in preverbal position. Interpolation itself is less normative, because it violates an important constraint of EP clitic grammar (clitic-verb adjacency). All in all, interpolation's frequency has been seriously reduced in comparison to Phase II.

One last thing that remains to be seen is whether there are any present outcomes of such a historical treatment. To answer this question, we conducted another search in corpusdoportugues.org. In its currently available version, only the 19th and 20th centuries are represented in the 'Genre / Historical' sub-corpus. Chronologically, they correspond to Phase III of interpolation as discussed above.

The 'Web / Dialects' sub-corpus brought no results of [n]-insertion in preverbal direct object pronouns. The search involved two lexemes – *quem* and *não* – likely to trigger the phonological change in question. In the first stage, the query focused on preverbal allomorphic realizations. Out of the eight conceivable configurations (*quem no*, *quem na*, *quem nos*, *quem nas* and *não no*, *não na*, *não nos*, *não nas*), only six were retained. *Quem nos* and *não nos* were left out of the analysis because of their homonymy with *nos* 1st person direct/indirect objects, and it would be an exceedingly time-consuming task to tell them apart.

The search conducted in the 'Genre / Historical' sub-corpus yielded a total of 28 hits (*quem no* – 6, *quem na* – 3, *quem nas* – 5, *não no* – 5, *não na* – 5, *não nas* – 4). All of the hits were in the 'Fict' section, which included literary production from the 19th and 20th century. By contrast, no hits were achieved in the remaining three sections ('oral', 'acad' and 'news' offering results for the 20th century only). These results are indicative of a diphasic and sociolin-

guistic profile of the preverbal allomorphic variants under discussion. Furthermore, an interesting feature was the predominance of 20th-century examples. Even more interestingly, most of them originated from the post-World War II period, which would show that [n]-insertion is still an extant pattern in preverbal 3rd person pronouns. Moreover, a detailed analysis of these variants across discourse contexts revealed that they represented mainly dialogue scenes or tended to imitate a spoken register. This characteristic is somewhat surprising in light of the fact that allomorphic variants were in no way represented in the 'oral' section of the sub-corpus. One plausible explanation is that the authors were somewhat heavy-handed in their attempts to reconstruct what they perceived to be the flow of 'real-life' conversations.

(26) *Meu rico paizinho, meu santo, que tanto lhe eu queria e não no pude ir ver na sua última hora!* (José Rodrigues Miguéis. ed. 1982. *A Escola do paraíso*)
REL so-much DAT.3.SG I.NOM want.IPFV.IND.1.SG and NEG ACC.3.M.SG can.PFV.1.SG go.INF see.INF in-DEF.F.SG POSS.F.SG last hour
'My beloved daddy, my saint, how I wanted to go and see him but couldn't, even in his last hour'

In the second stage of our quest for modern results of historical interpolation, bare *o*, *a*, *as* realizations following *quem* and *não* were taken into account. Unlike allomorphic sequences, they were found across all sections, thereby covering all registers, degrees of formality and discourse contexts. Likewise, they were not restricted to the 'Genre / Historical' sub-corpus but were also found in the 'Web / Dialects' one. They were also much more numerous than the occurrences of [n]-onset variants. For example, the search for the *não a* sequence in the 'Genre / Historical' sub-corpus for 20th century only has brought 482 hits.

Thus, over the course of the history of EP, allomorphy-avoidance in preverbal position and a rigorous distinction between postverbal morphological and preverbal phrasal attachment prove more compelling than the need for the uniqueness of realization. Curiously, the drift towards this manner of language processing appears to be one of the side-effects of interpolation. The latter, which owes its former prominence to the confluence of numerous factors (semantic properties of *não*, its syllabic structure corresponding to that

of 3rd person direct objects), might have additionally contributed to the relegation of allomorphic variants of preverbal pronouns to the periphery of clitic grammar in contemporary EP.

Corpora

Corpus Informatizado do Português Medieval, [online:] http://cipm.fcsh.unl.pt/

Biblioteca Nacional de Portugal − Biblioteca Digital, [online:] http://purl.pt/index/geral/PT/index.html

Corpus do Português: 1 billion words: Dialects / Genre / Historical, [online:] http://www.corpusdoportugues.org/

References

ANDERSON S.R., 2005, *Aspects of the Theory of Clitics*, Oxford: Oxford University Press.

ARTEAGA D., 2009, On the existence of null complementizers in Old French, In: P.J. Masullo, E. O'Rourke, C-H Huang (eds.), *Romance linguistics 2007: Selected Papers from the 37th Linguistic Symposium on Romance Languages (LSRL), Pittsburgh, 15–18 March 20017*, Amsterdam−Philadelphia: John Benjamins, pp. 19–36, [online:] https://doi.org/10.1075/cilt.304.03art.

BOUZOUITA M., KEMPSON R.M., 2006, Clitic Placement in Old and Modern Spanish: A Dynamic Account, In: O. Nedergaard Thomsen (ed.), *Competing Models of Linguistic Change: Evolution and beyond*, Amsterdam−Philadelphia: John Benjamins, pp. 253–268, [online:] https://doi.org/10.1075/cilt.279.16bou.

BRESNAN J., 2001, *Lexical-Functional Syntax*, Oxford: Blackwell.

CASTRO I., 2006, *Introdução à história do português*, Lisboa: Edições Colibri.

Da Tos M., The Italian FINIRE-type verbs: a case of morphomic attraction, In: S. Cruschina, M. Maiden, J.-Ch. Smith (eds.), *The Boundaries of Pure Morphology*, Oxford: Oxford University Press, pp. 45–67.

Dalrymple M., 2001, *Lexical Functional Grammar*, New York: Academic Press.

Delfitto D., 2002, On the Semantics of Pronominal Clitics and some of its Consequences, *Catalan Journal of Linguistics 1*, pp. 41–69.

Dryer M., 1988, Universals of negative position, In: M. Hammond, E. Moravcsik, J. Wirth (eds.), *Studies in Syntactic Typology*, Amsterdam−Philadelphia: John Benjamins, pp. 93–124, [online:] https://doi.org/10.1075/tsl.17.10dry.

Duarte I., Matos G., 2000, Romance Clitics and the Minimalist Program, In: J. Costa (ed.), *Portuguese Syntax. New Comparative Studies*, Oxford: Oxford University Press, pp. 116–142.

Fiéis M.A., 2001, Interpolação em Português Medieval como Adjunção a XP, In: *Actas do XVI Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, Colibri-APL, pp. 197–211.

Gerlach B., 2002, *Clitics between syntax and lexicon*, Amsterdam−Philadelphia: John Benjamins.

Heap D., Oliviéri M., Palasis K., 2017, Clitic pronouns, In: A. Dufter, E. Stark (eds.), *Manual of Romance Morphosyntax and Syntax*, Berlin−Boston: De Gruyter, pp. 183–229, [online:] https://doi.org/10.1515/9783110377088-005.

Hinzelin M.-O., 2010, L'interpolation dans les langues romanes: aspects diachroniques, In: M. Iliescu, H. Siller-Runggaldier, P. Danler (eds.), *Actes du XXVe Congrès International de Linguistique et Philologie Romanes. Innsbruck, 3–8 septembre 2007*, t. II, Tübingen: Niemeyer, pp. 331–339.

Labelle M., 2001, Trente ans de psycholinguistique, *Revue québécoise de linguistique 30(1)*, pp. 155–176.

Levelt W.J., 1989, *Speaking: from intention to articulation*, Cambridge, Mass.: MIT Press.

Luís A. 2014, On clitic attachment in Ibero-Romance. Evidence from Portuguese and Spanish, In: P. Amaral, A.M. Carvalho (eds.), *Portuguese-Spanish Interfaces: Diachrony, synchrony and contact*, Amsterdam−Philadelphia: John Benjamins, pp. 203–235.

Luís A., Kaiser G., 2016, Clitic Pronouns: Phonology, Morphology, and Syntax, In: L.W. Wetzels, J. Costa, S. Menuzzi (eds.), *The Handbook of Portuguese Linguistics*, West Sussex: John Wiley & Sons, Inc, pp. 471–486.

Magro C., 2010, When corpus analysis refutes common beliefs: The case of interpolation in European Portuguese dialects, *Corpus 9*, pp. 115–135.

Martins A.M., 2000, A Minimalist Approach to Clitic Climbing, In: J. Costa (ed.), *Portuguese Syntax. New Comparative Studies*, Oxford: Oxford University Press, pp. 169–190.

Martins A.M., 2014, Syntactic Change in Portuguese and Spanish. Divergent and parallel patterns of linguistic splitting, In: P. Amaral, A.M. Carvalho (eds.), *Portuguese-Spanish Interfaces: Diachrony, synchrony and contact*, Amsterdam––Philadelphia: John Benjamins, pp. 35–64.

Martins A.M., 2016a, Introdução: O português numa perspetiva diacrónica e comparativa, In: A.M. Martins, E. Carrilho (eds.), *Manual de linguística portuguesa*, Berlin–Boston: De Gruyter, pp. 1–40, [online:] https://doi.org/10.1515/978311036 8840-003.

Martins A.M., 2016b, A colocação dos pronomes clíticos em sincronia e diacronia, In: A.M. Martins, E. Carrilho (eds.), *Manual de linguística portuguesa*, Berlin––Boston: De Gruyter, pp. 410–430, [online:] https://doi.org/10.1515/97831103688 40-017.

Matos G., 2003, Aspectos sintáticos da negação, In: M.H.M. Mateus, A.M. Brito, I. Duarte, I.H. Faria (eds.), *Gramática da Língua Portuguesa* (6.ª edição), Lisboa: Caminho, pp. 767–793.

Miller Ph., Monachesi P., 2003, Les pronoms clitiques dans les langues romanes, In: D. Godard, A. Abeillé (eds.), *Les langues romanes, problèmes de la phrase simple*, Paris: CNRS, pp. 53–106.

Piel J.-M., 1989, A flexão verbal do português (Estudo de morfologia histórica), In: J.-M. Piel (eds.), *Estudos de linguística Galego-Portuguesa*, Lisboa: Imprensa Nacional-Casa de Moeda, pp. 213–244.

Raposo E., 2000, Clitic Pronouns and Verb Movement, In: J. Costa (ed.), *Portuguese Syntax. New Comparative Studies*, Oxford: Oxford University Press, pp. 266–297.

Sánchez López C., 2017, Coordination and correlatives, In: A. Dufter, E. Stark (eds.), *Manual of Romance Morphosyntax and Syntax*, Berlin–Boston: De Gruyter, pp. 647–688, [online:] https://doi.org/10.1515/9783110377088-018.

Sánchez-Rei X.M., 1999, Repercusións gramaticais do uso da interpolación pronominal en galego, *Cadernos de Lingua 19*, pp. 85–110.

Spencer A., Luís, A., 2012, *Clitics. An Introduction*. Cambridge: Cambridge University Press.

Thornton A., 2011, Overabundance (Multiple Forms Realizing the Same Cell): A Non-Canonical Phenomenon in Italian Verb Morphology, In: M. Maiden, J.-Ch. Smith, M. Goldbach, M.-O. Hinzelin (eds.), *Morphological Autonomy. Perspectives From Romance Inflectional Morphology*, Oxford: Oxford University Press, pp. 358–381.

Vigário M., 2003, *The Prosodic Word in European Portuguese*, Berlin–New York: De Gruyter.

Waltereit R., 2012, *Reflexive Marking in the History of French*, Amsterdam–Philadelphia: John Benjamins.

Zanuttini R., 2001, Sentential Negation, In: M. Baltin, Ch. Collins (eds.), *The handbook of contemporary syntactic theory*, Oxford–Malden: Blackwell, pp. 511–535.

Zwicky A.M., Pullum, G., 1983, Cliticization vs. Inflection. English *n't*, *Language 59(3)*, pp. 502–513.

Ku nieafiksalnym związkom proklityka-czasownik w europejskiej odmianie języka portugalskiego. Badanie korpusowe nad zmianą w składni wywołaną czynnikami fonologicznymi

(s t r e s z c z e n i e)

Artykuł przybliża powody, dla których portugalskie (europejskie) zaimki atoniczne w pozycji prewerbalnej, dawniej stanowiące część innych wyrażeń, stały się wyrażeniami prozodycznie wprawdzie niesamodzielnymi, lecz składniowo samodzielnymi. Materiał empiryczny pochodzi z utworzonego przez autorów korpusu, opartego na tekstach z XVII, XVIII i XIX wieku, wydobytych z portugalskiej Narodowej Biblioteki Cyfrowej. Ważną rolę w ewolucji klityk odegrał model linearny zwany 'interpolacją', czyli brak styczności między przedczasownikową klityką a formą czasownika. Elementem, który jako jedyny mógł je rozdzielić w analizowanym czasie był wykładnik negacji *não* 'nie'. Jego lokalizacja po zaimku eliminowała możliwość zmian alomorficznych, jakim ulegały w zwykłej proklizie zaimki 3 os. dopełnienia bliższego. To te właśnie zaimki były najczęściej oddzielone od formy czasownikowej przez *não*. Zmiana wyeliminowała bardzo trudne w przetwarzaniu mowy sekwencje sylab powstałych wskutek sąsiedztwa *não*-zaimek 3 os. (stąd popularność sekwencji o odwrotnym szyku: zaimek-*não*). Ponadto dowiedzione zostało istnienie tzw. samoistnej (nie wymuszonej) interpolacji. Analiza materiału językowego wskazuje, że była ona możliwa w kontekstach pozostawiających użytkownikom swobodę wyboru lokalizacji zaimka (przed bądź po czasowniku), a w których w obecnym portugalskim europejskim obowiązkowa stała się pozycja poczasownikowa.