The Origin of the Portuguese Inflected Infinitive through a Corpus Analysis

Michael J. Harris

1. Introduction

The Portuguese inflected infinitive (PII) is a unique form of infinitive. An (uninflected) infinitive is not marked for person or tense ([-TENSE, -AGR]), but an inflected infinitive has orthographic inflection that allows for person (and number) agreement, and is marked [-TENSE, +AGR]. The inflected infinitive also allows for the optional inclusion of a lexical subject in the infinitival phrase. The inflected infinitive is relatively rare. It has been found in a handful of other languages; according to Scida (2005), the inflected infinitive exists or existed in Galician, Sardinian, Old Neapolitan, Old Leonese, and Mirandese. Additionally, the inflected infinitive exists in Hungarian (Kiss 2002) and existed in Old English (Lass 1994).

The origin of the PII has long been a subject of debate. Several theories have been defended; the creative theory maintains that the PII developed in Portuguese from the uninflected infinitive. The analogy theory maintains that it derived from the Portuguese future subjunctive, whose orthographic form is very similar to the PII. The composite theory combines elements from the preceding two theories, maintaining that when the uninflected infinitive occurred with a lexical subject, a verb inflection corresponding formally to the future subjunctive was added to allow personal agreement. Finally, the imperfect subjunctive theory maintains that the PII developed from the Latin imperfect subjunctive.

This paper sheds light on the probable origin of the PII by assessing its occurrence and contexts in a corpus of Old Portuguese. Specifically, it evaluates 1) if the PII occurs in syntactic environments similar to those where the Latin imperfect subjunctive occurred in Late Latin or 2) in contexts that would permit an analogy with the Portuguese future subjunctive, and 3) whether or not the PII occurs frequently with lexical subjects in the infinitival phrase. Accordingly, this study first introduces the Portuguese infinitive, describes its uses, and then presents the four theories of the PII’s origins. Next, the corpus study and statistical analysis are presented. The results suggest that the origin of the PII is the Latin imperfect subjunctive due to the frequency of the PII in prepositional contexts, and especially purpose clauses analogous to the use of the imperfect subjunctive in Late Latin.

2. Infinitives: Classification and Origins of the PII in Previous Studies

Verbs without temporal marking, [-TENSE], belong to the non-finite verbal category, which can be divided into gerunds, past participles, and infinitives. The uninflected infinitive has neither tense nor agreement, [-TENSE, -AGR], and is marked by –r in Portuguese.

(1) [P] É correcto ignorar isso.

‘It is correct to ignore [-TENSE, -AGR] this.’

Michael Harris, University of California, Santa Barbara, michaelharris@umail.ucsb.edu. Thank you to Drs. Stefan Gries, Viola Miglio, and Eduardo Raposo (University of California, Santa Barbara) for their valuable feedback and comments; special thanks to Dr. Stefan Gries for his help in writing the code to search the corpus, and to Raquel Santana Paixão for help in translating Old Portuguese data. This work was presented at the Hispanic Linguistics Symposium 2012 at UFL, Gainesville, FL. The usual disclaimers apply.

In contrast to most other languages worldwide, an [+AGR] infinitive is allowed in Portuguese. The inflected infinitive does not possess temporal specifications, but it does allow for number and person agreement with the subject of the infinitival clause. This is known as the Portuguese inflected infinitive, or the personal infinitive, and is marked [-Tense, +AGR].

(2) [P] É correcto ignorarmos isso.
   ‘It is correct to ignore [-Tense, +AGR] this.’

A complete paradigm of the PII is as follows (exemplified with the verb comer (to eat)):

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p.</td>
<td>eu comer + Ø</td>
<td>nós comer + mos</td>
</tr>
<tr>
<td>2p.</td>
<td>tu comer + es</td>
<td>vós comer + des</td>
</tr>
<tr>
<td>3p.</td>
<td>ele comer + Ø</td>
<td>eles comer + em</td>
</tr>
</tbody>
</table>

The PII permits a lexical subject in the infinitival phrase, nós (we) in the following example. However, an overt subject is ungrammatical with an uninflected infinitive (Quicoli 1995).

(3) a. [P] É correcto nós ignorarmos isso.
   ‘It is correct we to ignore [-Tense, +AGR] this.’

b. [P] * É correcto nós ignorar isso.
   ‘It is correct we to ignore [-Tense, -AGR] this.’

The origin of the PII has long been a topic of interest to linguists, and there are four major theories attempting to explain it. The first is the creative theory, which claims that the PII developed verbal inflections to allow the infinitive to agree with a lexical subject within the infinitival clause. “There has been a broad consensus amongst scholars” as to the soundness of the creative theory ever since it was defended by Maurer in 1968 (Martins 2001: 209). According to the creative theory, at some point in Old Portuguese, a lexical subject must have been permitted within infinitival clauses, as is sometimes the case in some dialects of Spanish:

(4) [S] Para Rita ver la película, se fue al centro.
   ‘For Rita to see [-TENSE, -AGR] the film, she went downtown.’

The creative theory would posit he following stages of development for the PII:

(5) [P] É necessário terminar a tarefa.
   ‘It is necessary to finish [-TENSE, -AGR] the work.’
   uninflected infinitival clause without lexical subject.

(6) [P] É necessário nós terminar a tarefa.
   ‘It is necessary we to finish [-TENSE, -AGR] the work.’
   uninflected infinitival clause with lexical subject.

(7) [P] É necessário (nós) terminarmos a tarefa.
   ‘It is necessary (we) to finish [-TENSE, +AGR] the work.’
   inflected infinitival clause

While the creative theory has been widely defended, the low frequency of occurrence of the PII with a lexical subject in Old Portuguese makes this development questionable. For instance, Wireback (1994) analyzed 153 sentences of Old Portuguese (dated 1400 and earlier) and found that the inflected infinitive had a lexical subject less than 10% of the time. Had the PII in fact developed from the presence of a lexical subject with a regular infinitive, it is reasonable to expect that the PII would often co-occur with a lexical subject in Old Portuguese. Additionally, as previously mentioned, constructions such as (6) are not permissible in modern Portuguese.

The analogy theory, on the other hand, posits that the PII developed from the Portuguese future subjunctive. In Portuguese, the future subjunctive and the inflected infinitive are formally identical in regular verbs:
Because of this formal identity, it has been suggested that the same inflections extended from the Portuguese future subjunctive to the PII to allow for person agreement. However, another step is necessary to account for irregular verbs like querer (Scida 2004:91):

<table>
<thead>
<tr>
<th>Future Subjunctive</th>
<th>Inflected Infinitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg</td>
<td>quiser</td>
</tr>
<tr>
<td>2sg</td>
<td>quiseres</td>
</tr>
<tr>
<td>3sg</td>
<td>quiser</td>
</tr>
<tr>
<td>1pl</td>
<td>quiserem</td>
</tr>
<tr>
<td>3pl</td>
<td>quiserem</td>
</tr>
</tbody>
</table>

For analogy theorists, if the subjunctive stem differed from the infinitive, the infinitival stem would then have extended to the PII by analogy. The central criticism of the analogy theory is the fact that although they are formally similar, the future subjunctive and the inflected infinitive appear not to share the same syntactic environment. The one commonly shared context of the PII and future subjunctive is following depois 'after', and “only one context seems an unlikely origin to such a widespread verb form.” (Wireback 1994: 548).

A third theory, the composite theory, combines the previous two theories. It claims that the presence of a lexical subject with the uninflected infinitive was viewed as personal, so the inflections from the Portuguese future subjunctive were added by analogy. This theory, however, is criticized for the same reasons as the previous two theories, as enumerated by Scida:

… the non-existence of an inflected infinitive in other Romance languages which have an occurrence of an overt nominative subject construction; the low frequency of occurrence of an overt nominative subject with the infinitive; the difference in stems between the future subjunctive and inflected infinitive; the lack of shared syntactic environments in which both the future subjunctive and the inflected infinitive occur ... (2004:92-93).

Finally, the imperfect subjunctive theory proposes that the PII developed from Latin rather than from within Portuguese. Latin had two past forms of the subjunctive, the imperfect subjunctive and the pluperfect subjunctive. The pluperfect subjunctive developed into the Portuguese imperfect subjunctive and, according to this theory, the Latin imperfect subjunctive developed into the PII. Consider the following forms (Martins 2001: 209):

<table>
<thead>
<tr>
<th>Latin Pluperfect Subjunctive</th>
<th>Portuguese Imperfect Subjunctive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 sg</td>
<td>prendissem</td>
</tr>
<tr>
<td>2 sg</td>
<td>prendisses</td>
</tr>
<tr>
<td>3 sg</td>
<td>prendisset</td>
</tr>
<tr>
<td>1 pl</td>
<td>prendissemus</td>
</tr>
<tr>
<td>2 pl</td>
<td>predissetis</td>
</tr>
<tr>
<td>3 pl</td>
<td>prendissent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Latin Imperfect Subjunctive</th>
<th>Portuguese Inflected Infinitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 sg</td>
<td>prenderem</td>
</tr>
<tr>
<td>2 sg</td>
<td>prenderes</td>
</tr>
<tr>
<td>3 sg</td>
<td>prenderet</td>
</tr>
<tr>
<td>1 pl</td>
<td>prenderemus</td>
</tr>
<tr>
<td>2 pl</td>
<td>prederetis</td>
</tr>
<tr>
<td>3 pl</td>
<td>prenderent</td>
</tr>
</tbody>
</table>
The Latin imperfect subjunctive appears to have survived in Late Latin. While the Latin pluperfect subjunctive was the primary form of past-tense subjunctive by the third century A.D, textual evidence shows that the Latin imperfect subjunctive survived, at least in purpose and result clauses, (Wireback 1994: 550), making its development into the PII plausible.

The imperfect subjunctive theory does have one major inconsistency. Proponents of the creative approaches such as Maurer & Henrique (1968) pointed out that the coalescence of the inflected infinitive and the imperfect subjunctive should occur in volitional contexts, like those where the Latin imperfect subjunctive occurred. However, the opposite is true: in volitional constructions, the PII was not permitted. Thus, the imperfect subjunctive theory appears to fail to provide a link with the volitional contexts that should have originated the morphological extension from the Latin imperfect subjunctive to the uninflected infinitive. However, prepositional contexts may provide a different environment for this morphological extension. Late Latin employed *ut* plus the infinitive or imperfect subjunctive in purpose clauses. At a certain stage in the history of Luso-Romance, *ut* plus the imperfect subjunctive or an infinitive indicated a purpose relationship. Because of regular phonological development and frequent *ut* deletion, the imperfect subjunctive and the infinitive resembled one another and could be “reanalyzed as an infinitive with inflectional morphology.” (Wireback 1994:550).

Admittedly, none of the theories of the origin of the PII can definitively account for its development. Thus, a more in-depth exploration of its contexts in Old Portuguese, employing both historical corpus data and modern statistical methodology is beneficial in providing perspective of the origin of the PII.

3. The Present Study

3.1. Data treatment and methodology

In light of the debate of the origin of the PII, it is necessary to investigate a corpus of Old Portuguese sentences in order to identify the context in which the PII occurred. The current study analyzes the Tycho Brahe Parsed Corpus of Historical Portuguese, a publicly available partially annotated corpus of written Portuguese (Galves & Faria 2010), to investigate three main questions:

1) To what extent do inflected infinitives occur after prepositions, especially *por* and *para*?
2) To what extent do inflected infinitives occur in contexts analogous to the Portuguese future subjunctive?
3) To what extent do inflected infinitives occur with overt lexical subjects?

Four authors’ works were analyzed in the current study: a Portuguese grammar written by João de Barros (born 1497, 53,573 words), a historical account by Duarte Galvão (born 1435, 50,732 words), a historical account by Fernão Lopes (born 1380, 49,171 words), and a historical account by Ruy de Pina (born 1440, 50,937 words). The total of 204,413 words was searched for matches to all occurrences of Portuguese inflected infinitives as tagged in the Corpus do Português (Davies 2006) for the same time period using a script in R (R Core Development Team 2013). That is, any sentence in the Tycho Brahe Corpus (Galves & Faria 2010) containing an inflected infinitive tagged in the Old Portuguese files of the Corpus do Português (Davies 2006) was returned by the script. These sentences were then manually inspected for 1) the context in which they occur, and 2) whether they occur with an overt lexical subject.

3.2. Results

226 Portuguese inflected infinitives were positively identified1 in the relevant files of the Tycho Brahe Parsed Corpus of Historical Portuguese (Galves & Faria 2010), and categorized according to the context in which they occurred (following each context is an example from the corpus):

---

1 Some matches returned by R were deemed either false positives or too ambiguous in form and/or orthography to include in the results of the study.
1) Por/Para– PII in purpose clauses with *por, para, pera* (for, in order to).
   (8) [P] por serem substantivos que não recebem em si outros.
   *‘for to be [3pl] nouns that do not themselves receive others.’*

2) Other Prepositions/ Relative Pronouns- PII in clauses introduced by prepositions such as *sem* (without), *com* (with), *em* (in, at), etc.
   (9) [P] serem mortos e feridos muitos mais sem se poderem remediar
   *‘to be [3pl] dead and wounded many more without to be able [3pl] to repair themselves’*

3) De- PII in clauses introduced by preposition *de* (from) or equivalent contractions.
   (10) [P] me espanto de a trazeres... decorada
   *‘It surprises me to bring it [2s] decorated.’*

4) Temporal clauses- PII introduced by prepositions such as *antes de* (before)/ *depois de* (after), etc.
   (11) el rey d. diniz chegou .... ante de darem, e pronunciarem [as] dictas.
   *‘King Diniz arrived...before to give [3pl], and pronounce [3pl] the maxims.’*

5) Subject Clauses- PII used as the subject of a subject clause.
   (12) [P] irmos plantando estas flores de gentilidade para recreação
   *‘to go [1pl] planting these flowers of kindness for amusement.’*

6) Complement Clauses– PII following causative and perception verbs, and *verba dicendi*.
   (13) [P] podemos também dizer serem nomes verbais todos os infinitivos do presente tempo
   *‘we can also say to be [3pl] unfinites all present tense infinitives.’*

7) Gerund- PII introduced by a gerund.
   (14) [P] fugindo perdermos
   *‘fleeing to lose [1pl].’*

Additionally, for each observed PII, it was recorded whether or not the PII had an overt lexical subject and, if so, what that subject was.

1) Inflected infinitive without lexical subject
   (15) [P] por terem outros signficados
   *‘for to have [3pl] other meanings.’*

2) Inflected infinitive with lexical subject
   (16) [P] Disse os nomes terem dois números
   *‘said the nouns to have [3pl] two numbers.’*

The following total occurrences of the contexts of the PII were observed:

<table>
<thead>
<tr>
<th>Context</th>
<th>Occurrence</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) pera/ para/ por</td>
<td>99</td>
<td>44%</td>
</tr>
<tr>
<td>2) other prepositions</td>
<td>36</td>
<td>16%</td>
</tr>
<tr>
<td>3) de</td>
<td>25</td>
<td>11%</td>
</tr>
<tr>
<td>4) temporal clauses</td>
<td>14</td>
<td>6%</td>
</tr>
<tr>
<td>5) subject clauses</td>
<td>27</td>
<td>12%</td>
</tr>
<tr>
<td>6) complement clauses</td>
<td>24</td>
<td>11%</td>
</tr>
<tr>
<td>7) gerund</td>
<td>1</td>
<td>.004%</td>
</tr>
</tbody>
</table>

Table 1: Observed contexts of the PII in Old Portuguese.
According to a chi-squared goodness of fit test, the distribution of the occurrences of 1-4 and 6-7 with the PII deviated significantly from the distribution of these same categories elsewhere in the corpus (χ-squared = 260.27, df = 5, p-value < 2.2e-16). That is to say that the distribution of these categories in conjunction with the PII was highly significantly different from their distribution within the rest of the corpus. Not enough information about the distribution of subject clauses in the corpus was available in order to perform a test for dissimilarity with subject clauses (6). The following occurrences of PII and overt lexical subjects were observed:

<table>
<thead>
<tr>
<th>Context</th>
<th>Occurrence</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>no lexical subject</td>
<td>204</td>
<td>90.3%</td>
</tr>
<tr>
<td>with lexical subject</td>
<td>22</td>
<td>9.7%</td>
</tr>
<tr>
<td>total</td>
<td>226</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Observed occurrences of the PII with lexical subjects in Old Portuguese.

As was the case with the occurrence of subject clauses, there was not enough information regarding the distribution of nouns and pronouns within the corpus in order to see whether or not the distribution of lexical subjects within inflected infinitival clauses differed from their distribution in the rest of the corpus.

Figure 1: Observed contexts of the Portuguese Inflected Infinitive in Old Portuguese from the Tycho Brahe Parsed Corpus of Historical Portuguese (Galves & Faria, 2010): dark grey bars represent prepositional contexts while light grey bars represent non-prepositional contexts.

With regards to the first of our main questions, a strong preference for prepositional contexts was found in the corpus data. In particular, por/para (in order to) purpose clauses were the most frequent contexts of the PII. As shown in Figure 1, the prepositional phrases represented 77% of the contexts in which the PII was used, 44% of which were por/para phrases. Because the propensity of Latin to employ purpose clauses with the imperfect or infinitive, and possible ut deletion, the imperfect subjunctive theory seems to be favored as the origin of the PII. Following is an example of a purpose clause from the corpus and an example of a Latin purpose clause with the Latin imperfect subjunctive from Scida (2004:124-125).

(17) [L] Caesar equos removit ut spem fugae tolleret.
   ‘Caesar removed the horses that he might take away the hope of flight.’

[P] não irem juntos, e que para irem mais folgados
   ‘not to go [3pl] together, and so to go [3pl] more easily.’
Other prepositions found in the corpus also provide analogy between the Latin imperfect subjunctive and the PII. For example, consider the following example of a temporal clause from the corpus as compared to the Latin imperfect subjunctive from Scida (2004:124-125).

(18) [L] Pervenit, priusquam Pompeius sentire posset.
   ‘He arrived before Pompey could become aware of his approach.’
   [P] el rey d. diniz chegou... ante de darem, e pronunciarem [as] dictas.
   ‘King Diniz arrived...before to give [3pl], and pronounce [3pl] the maxims.’

In addition to prepositional forms, complement clauses employed in the corpus provide shared syntactical context between the PII and the Latin imperfect subjunctive (Scida 2004:124-125).

(19) [L] huic imperat quas posit adeat civitates.
   ‘He orders him to visit such communities as he can.
   [P] ordenou …porem nas cartas.
   ‘(He) ordered to put [3pl] in the cards

Thus, the corpus data reflects that the PII appeared in Old Portuguese in contexts that allow analogy with the Latin imperfect subjunctive. The purpose clauses involving por/para are the most common, but there are several other contexts that may have allowed transfer from the Latin imperfect subjunctive to the PII as well. The overall occurrences of the PII in the corpus reveal a high correlation to the Latin imperfect subjunctive in Late Latin, strengthening the argument of the Latin imperfect subjunctive theory of the origin of the PII.

The second main question concerned the shared context between the PII and the Portuguese future subjunctive. As previously mentioned, the analogy theory had been criticized due to the scarcity of analogous contexts between the PII and the future subjunctive. For instance, Wireback found only one such context in 153 sentences examined (1994), namely depois (after). The current study found seven instances of the PII following depois (2.9%). Additionally, the PII was found following se (if).

(20) [P] se quererem... dizer, antes no meyo das mayores chamas se diz
   ‘if to want [3pl] to say, before in the middle of the biggest flames they said it.’

Note that if (20) were the Portuguese future subjunctive, it would take the irregular root as in quiserem. However, se also allows the future subjunctive, as seen in (21) from the corpus.

(21) [P] para nós maiormente se dermos [future subjunctive] em eles de sobressalto
   ‘better for us if [we] encounter them in fright.’

Although Wireback criticizes the future subjunctive theory of the origin of the PII because “the future subjunctive and the inflected infinitive do not share the same syntactic environment” (1994:548), the current data provides both additional shared contexts between the two forms and more examples of the PII following depois. While this in itself is not enough to definitively sway the debate in favor of the future subjunctive or analogy theories, it does demonstrate the need to further explore shared syntactical contexts between the PII and the Portuguese future subjunctive.

Finally, regarding the presence of overt lexical subjects with the PII, the current study found a low co-occurrence of overt lexical subjects and PII. Only 9.7% of observed instances of the PII had overt lexical subjects within the infinitival clauses. This does call into question the validity of the creative and composite theories, which claim that the presence the overt lexical subject with the uninflected infinitive led to the addition of verbal inflection, leading to the PII. However, while one might expect that the PII would occur with more lexical subjects if this were the case, this is not necessarily a foregone conclusion. Consider that an inflected infinitive may make the use of an overt subject unnecessary, especially in the case that the subject of the main clause and the infinitival clause differ, as in the following ad hoc possibilities:
(22) [P] A criança, por ter chegado tarde, chorou.

*The child because to have arrived [-TENSE, -AGR] late, cried.* Grammatical but assumed co-reference between subject of both clauses, leading to ambiguous interpretation.

(23) [P] *A criança, por os pais chegado tarde, chorou.

*The child because the parents to have arrived [-TENSE, -AGR] late, cried.*

Ungrammatical use of uninflected infinitive + subject to clarify difference between subjects of clauses.


*The child to have arrived [-TENSE, +AGR] late, cried.* Grammatical with clear distinction between subjects of clauses.

As (23) is ungrammatical, (24) may indeed be a plausible solution to avoid ambiguity in the phrase. By adding verbal inflection, the infinitival clause is preserved but the distinction between the subjects of both clauses is no longer ambiguous. This is in agreement with Vanderschueren and Diependaelle’s (2013) work regarding the choice to use the inflected infinitive as opposed to the uninflected infinitive in modern Portuguese: “the utterer adapts his message in order to optimize the processing ease of the receiver” (2013:181). It may be that the PII does not commonly appear with an overt subject due to the fact the verbal inflection makes it unnecessary to include the subject. This notwithstanding, the purpose for inclusion of the lexical subject was not the subject of the current study and further investigation into the matter is necessary before any conclusions can be drawn.

In general, the data from the Tycho Brahe Parse Corpus of Historical Portuguese lend support to the imperfect subjunctive theory of the origin of the PII. The occurrence of the PII in prepositional phrases, especially por/para purpose clauses provide a syntactical environment where analogy with the Latin imperfect subjunctive is possible. Additionally, other prepositional phrases, which are possibly analogous to contexts available to the Latin imperfect subjunctive, as well as complement clauses of purpose, add further contexts where transfer was possible. As previously noted, the current data do not completely discount the theory that the PII may have arisen via analogy with the Portuguese future subjunctive.

4. Conclusions

The present study has investigated the origin of the PII using corpus-based methodology in order to make an educated inference as to its origins. The corpus data suggest that the PII most likely developed from the Latin imperfect subjunctive as it most frequently occurs in contexts similar to those where the imperfect subjunctive occurred in Late Latin, although the data also suggest that analogy with the Portuguese future subjunctive may have led to the PII, as it also occurs in contexts where the future subjunctive is possible.

Although previous studies investigating the origin of the PII have yielded somewhat similar results (cf. Wireback 1994), this study utilized a much larger corpus on which to test existing hypotheses than in previous studies, and it has validated its results using sound statistical methods. There are some limitations to this study (little genre diversity, only four authors from a later time period than would be ideal), but in the end, a classification of the PII according to context provides true insight into the origin of the PII and its use in Old Portuguese, even if a unique, definitive answer is still not possible.

For further study, it would be desirable to use a corpus encompassing more genres and the ability to ascertain whether the distribution of nouns and pronouns with the PII was significantly different from their distribution elsewhere in the corpus. It would also be of interest to investigate in what ways the distribution of the PII in Old Portuguese resembles the distribution of the uninflected infinitive in Old Portuguese, and in what particular contexts overt lexical subjects are used with the Portuguese infinitive. As is often the case in historical linguistics, a definitive answer as to the origin of the
inflected infinitive still eludes us. The investigation of the preceding questions will further enlighten us as to the evolution of the PII.

References


