Isogloss
Open Journal of Romance Linguistics

Received: 21-09-2024
Accepted: 26-06-2025
Published: 14-07-2025

Variation of the inflected infinitive in Portuguese

Jill McLendon
University of Georgia
jsm50968@uga.edu

Vanessa Revheim
University of Georgia
vanessa.revheim@uga.edu

Allonah Ezro-Christy
University of Georgia
aezroc@uga.edu

How to cite: McLendon, Jill, Revheim, Vanessa, & Allonah Ezro-Christy. 2025. Variation of the inflected infinitive in Portuguese. RLLT 25, eds. Adam McBride, James Law & Willis Fails. Special Issue of *Isogloss. Open Journal of Romance Linguistics* 11(5)/10, 1–21.
DOI: https://doi.org/10.5565/rev/isogloss.501

Abstract

This study uses corpus methods to examine a controversial topic in Portuguese syntax (Cunha & Cintra 2016): the variation of the inflected infinitive, an uncommon verbal form that marks the infinitive for person and number but not tense. Previous work by grammarians about the use, structure, and distribution of the inflected infinitive is inconsistent. Pires and Rothman (2009) argue that the inflected infinitive has been lost in colloquial Brazilian Portuguese varieties, only being acquired through schooling and/or through the media. Cunha and Cintra (2016) state that there is variation when the subject of the infinitive is an oblique pronoun. This study uses SketchEngine's ptTenTen20 corpus (Kilgarriff et al. 2014) to investigate theoretical claims from previous literature. The advantage of web-scrapped corpora like the ptTenTen20 lies in large datasets and meta-annotation for variables of interest. Data were queried from SketchEngine, processed using R script, and analyzed with a fixed-effects logistic regression, using variety, genre, and structure as variables. The inflected infinitive was favored in European Portuguese and with clitic pronouns, with significant but contradictory effects seen for register. This study adds directly to the body of literature

about the inflected infinitive in Portuguese, using empirical methods to investigate previous theoretical claims.

Keywords: corpus linguistics, Romance languages, Portuguese, inflected infinitive variation, logistic regression

1. Introduction

The inflected infinitive is an uncommon verbal form found in Portuguese and in Galician. Although other authors present data from other languages, such as Hungarian and Sardinian, suggesting that these languages have conjugated infinitives as well, these languages use it in different environments than we find in Portuguese and Galician (Martins & Carrilho 2016). In terms of verbal endings, the inflected infinitive lacks overt specification for tense; however, it presents the same inflection of person and number found in the future subjunctive, where inflection is overtly marked for 1PL (*falarmos* 'to speak-INF-1PL'), 2SG (*falares* 'to speak-INF-2SG'), 2PL (*falardes* 'to speak-INF-2PL'), and 3PL (*falarem* 'to speak-INF-3PL').

As for its properties, the inflected infinitive depends primarily on the speaker's need or intention to demonstrate the agent of the action, and is commonly used after causative verbs where the agent is not the subject of the sentence, as shown in example (1), below, from the corpus data:

(1) Portuguese

Ela	é	uma	atleta	incrível	que	inspira	
She	is	an	athlete	incredible	who	inspires	
para	nós		podermos		chegar	aos	nossos
for	us.PRO.1PL		be able to.INF.1PL		get	to	our
próprios	sonhos						
own	dreams						

'She is an incredible athlete that inspires us to achieve our dreams.'

Despite being a well-established structure in both European Portuguese (EP) and Brazilian Portuguese (BP), the distribution of the inflected infinitive remains a controversial topic. Cunha and Cintra (2016) defend that previous attempts to establish rules have failed to comprehensively account for the observed variability in the literature. Additionally, variation in use can also be attributed to differences among various Portuguese varieties. Pires and Rothman (2009) claim that the inflected infinitive has been lost in BP varieties, where only educated speakers demonstrate competence in its usage.

Our study was motivated by the controversy pointed out by Cunha and Cintra (2016) and the experimental evidence from Pires and Rothman (2009). Considering that the inflected infinitive is a low-frequency structure in Portuguese, the use of large corpora is a promising methodological alternative for investigating linguistic questions and validating results that would be difficult to examine experimentally due to rarity. Furthermore, the preexisting metadata available in annotated corpora make it possible to explore the variations through the effects of registers, dialectal varieties, and fine-grained syntactic detail. Given these advantages, we use corpus data to empirically

evidence if EP and BP are different in their use of the inflected infinitive, whether the inflected infinitive is only used in formal registers, and to evaluate the theoretical claims about its distributional variation. To narrow the environments where the inflected infinitive occurs, this study focuses on causative constructions in which the matrix verb is followed by *para* 'for'[1] or its informal variant *pra*.

In section 2 we introduce the properties of the inflected infinitive and what is known from previous studies. In section 3, we present our research questions and hypotheses and then outline the methodology in section 4. Finally, we present our findings in section 5, followed by concluding remarks in section 6.

2. Previous Literature

As mentioned in the previous section, the inflected infinitive depends primarily on the speaker's need or intention to demonstrate the agent of the action in causative constructions, as shown in examples (2a) and (3a). However, according to Cunha and Cintra (2016), there is variation in cases when the pronoun is a clitic, as indicated below in example sentences (2b) and (3b):

(2) Portuguese, Authors

a. Eu vi eles partirem
I see.PST.1SG they.PRO.3PL leave.INF.3PL
'I saw them leave.'

b. Eu os vi partir
I them.CL.3PL see.PST.1SG leave.INF
'I saw them leave.'

(3) Portuguese, Authors

a. Eu pedi para eles saírem
I ask.PST.1SG for they.PRO.3PL leave.INF.3PL
'I asked them to leave.'

b. Eu os pedi para saírem
I them.CL.3PL ask.PST.1SG for leave.INF.3PL
'I asked them to leave.'

[1] We would like to thank the reviewer for highlighting the syntactic differences between the *para*-infinitive sentences included in our analysis. We recognize that our study does not differentiate between adjunct and argument clauses, which represent distinct syntactic structures. This limitation arises from the scope of our search, which included all contexts without distinguishing their syntactic roles or the associated constraints on the use of inflected and uninflected infinitives. While refining the search parameters to account for these differences would enhance the accuracy of the results, such modifications are beyond the scope of this current study. We acknowledge this as a limitation and recommend that future research can address these distinctions to provide a more granular analysis of *para*-infinitive variation.

The variation in example (3b) is also theoretically unexpected, as the clitic *os* does not trigger agreement with the verb. Therefore, the expected form in example (3b) would be the uninflected infinitive.[2]

Over the years, syntacticians have documented contexts where the inflected infinitive may appear, as well as contexts where the uninflected infinitive is the only acceptable form. Raposo (1987) and Madeira (1994) both concur that the inflected infinitive can be found as compliments to factive predicates, as well as epistemic and declarative predicates, although Raposo (1987) stipulates that this is only possible when the infinitive in question precedes a lexical subject, a restriction that does not apply to factive clauses. Maderia (1994), on the other hand, specifies that factive clauses disallow the inflected infinitive in cases of subject-verb inversion unless there is an auxiliary. Both authors also note that the inflected infinitive can be found in subject clauses as well as adjunct clauses introduced by a preposition. Madeira (1994) lists several additional contexts in which the inflected infinitive appears, such as compliments to perception verbs, compliments to causative predicates, compliments to object control predicates, and compliments to transitive subject control predicates. Raposo (1987) and Madeira (1994) also stipulate that the inflected infinitive cannot appear in compliments to volitional predicates. While Madeira (1994) adds interrogative and relative clauses to the list of contexts where the inflected infinitive cannot appear, Raposo (1987) says it cannot occur with a wh-phrase or a null operator as the subject of the infinitive.

According to Cunha and Cintra (2016), the use of the inflected infinitive is a topic of controversy in Portuguese syntax, where attempts to explain its use have resulted in insufficient or unrealistic rules. The authors propose that its use cannot be explained by grammatical rules only, since pragmatic factors, such as avoiding ambiguity, and stylistic choices also play a role. Cunha and Cintra (2016) suggest observing usage trends of the inflected infinitive instead of explaining its distribution with grammatical rules.

Although the main trigger for the inflected infinitive is when the subject of the matrix verb is different from the agent of the infinitive verb, Maurer (1968) argues that such a rule is irrelevant. Therefore, another possible factor for variation is when the agent of the infinite verb is also the subject of the matrix verb. In fact, whenever the infinitive has its own subject, regardless of whether it is the same subject as the main verb, it will always be inflected.

Cunha and Cintra (2016) also note that the use of the inflected infinitive may be more common in EP because of the use of *tu*, which allows for the possibility of the inflected infinitive, unlike *você* commonly used in BP, which takes a third-person singular verb ending, which for the inflected infinitive is null making it impossible to identify whether or not the inflected infinitive is being employed. Additionally, Pires and Rothman's (2009) study on language acquisition and competence divergence across the heritage grammars of EP and BP states that the inflected infinitive is no longer present in colloquial BP. The authors argue that the inflected infinitive is only acquired through schooling and/or through the media, meaning that it should only be

[2] We would like to thank the reviewer for suggesting an exploration of unexpected cooccurrences of accusative clitics with inflected infinitives, as in the context of causative and perception verbs (Hornstein, Martins & Nunes 2008, Barbosa, Flores & Pereira 2018). As indicated by the reviewer, sentences in which accusative clitics are inflected with both causative and perception verbs are accepted by speakers, particularly with third-person clitics.

found in formal registers of BP. Similarly, Rodrigues and Hornstein (2013: 308) propose that "Loss of verbal morphology in BP is also observed in the scarcity of inflected infinitives among native speakers, who do not master these verbal forms naturally but as a result of a deliberate effort toward learning what is called 'Standard' Portuguese."

Contrary to the claims made by Pires and Rothman (2009) and Rodrigues and Hornstein (2013), Canever and Beline Mendes (2019) present multiple studies that suggest an expansion of the inflected infinitive is in progress. Their study focuses on the perception of competence dealing with hypercorrection of the inflected infinitive after a modal verb, a prescriptively ungrammatical construction, and other syntactic contexts. They find that younger listeners are more neutral in their grammatical judgments, even in non-standard/hypercorrective contexts. The authors cite several corpus-based studies on both EP (Bossaglia 2013) and BP (Ladeira 1986, Cabral 2006, Canever 2012). Based on editorials from *Jornal do Brasil* and a translated version of the Jerusalem Bible, Ladeira (1986) finds several factors commonly observed with the inflected infinitive including the first-person plural, the infinitive in an adverbial clause, and the presence of reflexive pronouns, among others. Ladeira (1986) also finds that the inflected infinitive occurs less when the subject of the infinitive is an oblique pronoun (clitic pronoun), as shown in our example (3b). Cabral (2006) also uses news sources as a corpus, pulling from *Jornal do Brasil* and *O Globo*. Despite predicting that there would be a higher rate of the inflected infinitive in standard language such as written news, Cabral (2006) found a high rate of the uninflected infinitive; she did however find a high rate of the inflected infinitive with the passive voice and non-canonical types of reflexive clauses. Finally, Canever (2012) uses a corpus of theses and dissertations from the University of São Paulo and finds a high rate of the inflected infinitive in adverbial clauses (75%), as complements of adjectives (53.5%), and as complements of nouns (94.5%), all of which are considered by prescriptivists to be cases where use of the inflected infinitive is optional, as well as some nonstandard uses of the inflected infinitive. In line with previous findings, Canever (2012) also notes a preference for the inflected infinitive with the first-person plural (90%) and passive voice (92%).

In the past, due to the relatively infrequent nature of the inflected infinitive, acquiring sufficient data for a variationist study has proven difficult. Researchers have compiled their own corpora consisting of a single genre, with newspaper sources being the most common. Only in the last few decades have large corpora of Portuguese data become available, such as the CETEMPúblico for EP in 2001, the PorPopular for BP released in 2012 and the Corpus do Português, released in 2016. Whereas previous corpus studies have taken a taken a bottom-up approach, exploring the use of the inflected infinitive in a variety of syntactic contexts, such as adverbial clauses (Ladeira 1986, Canever 2012), reflexive pronouns and clauses (Ladeira 1986, Cabral 2006), passive voice (Cabral 2006, Canever 2012), and compliment clauses (Canever 2012), within a specific register or genre, fewer studies have examined factors such as variety or compared the rate of the inflected infinitive across different registers. While grammarians may argue the circumstances under which the inflected infinitive may or may not be used, increased access to larger data sets and the use of statistical software allow variationists to provide a more accurate account of language based on its use.

This study adopts a macro-level approach, supported by a mixed-methods quantitative analysis, while using a larger and more diverse corpus. The use of a

logistic regression analysis supplemented with descriptive statistics allows for an intuitive visualization of patterns while providing statistically validated findings. Section 4 of this paper will discuss the methodology used for this study, including the corpus we chose and the motivations for our decision, and section 6.1 presents some difficulties of working with large corpora.

3. Research Questions and Hypotheses

The inconsistency of rules emphasized by Maurer (1968) and Cunha and Cintra (2016), coupled with the experimental observations by Pires and Rothman (2009) encountered in previous research on BP prompted the following research questions:

1. How do varieties of Portuguese differ in their use of the inflected infinitive?
2. Is the inflected infinitive only found in formal registers of BP?
3. Is there variation in cases when the pronoun is a clitic?

Based on our research questions, our hypotheses for the present study are as follows:

H1. European Portuguese will favor the inflected infinitive.
H2. Informal registers will disfavor the inflected infinitive.
H3. Constructions with a clitic will show variation between the inflected infinitive and the uninflected form.
H0. Neither variety nor register will have an effect on the choice between the inflected infinitive and the uninflected infinitive. Constructions with a clitic will only take an uninflected infinitive.

Based on previous studies on the inflected infinitive, we first predict that 1) EP and BP will differ in the use of the inflected infinitive, with EP presenting higher numbers of the inflected infinitive, 2) the inflected infinitive will be more commonly used in formal registers, and 3) there will be variation with clitics.

4. Methodology

4.1. Data collection

The research questions in this study focus on the variation of the Portuguese inflected infinitive, which requires corpora that are annotated in such a way as to allow a distinction between formality of register and language variety, as well as being compatible with Corpus Query Language (CQL), a pattern-matching search language. For this reason, we chose to use the Sketch Engine ptTenTen20 corpus, which contains web-scraped Portuguese data from multiple genres and varieties. Other corpora, specifically the Corpus do Portuguese (CDP), were considered as additional sources of data, but as the CDP only distinguishes between ‘general’ and ‘blog’ genres, it was determined that making an arbitrary distinction in the formality of these genres was unjustifiable, and thus SketchEngine was chosen.

The variables shown in Table 1 were chosen based on previous literature as factors suggested to be involved in the realization of the inflected infinitive. The variables and their descriptions are outlined in Table 1, below.

Table 1. Variables and descriptions.

Variable	Description
outcome	dependent variable: uninflected vs inflected infinitive
variety	Brazilian vs. European Portuguese
register (genre)	formal (legal, news, encyclopedia) vs informal (blog)
type	*para* (formal) vs. *pra* (informal)
structure	inflected infinitive agent as tonic pronoun: *Eu pedi para eles saírem*. inflected infinitive agent as clitic: *Eu os pedi para saírem.*
non-clitic	*nós, tu, vós, eles, elas*
clitic	*nos, te, vos, os, as, lhes*

4.2. Data extraction and analysis

Data were collected from each relevant subcorpus of the ptTenTen20 corpus (full size 12,578,775,252 words), based on two strings written in corpus query language (CQL), a grammatical parsing system used to search through large bodies of annotated text to find tokens. We used two different query structures: The first was used to extract cases in which the subject of the embedded clause is realized as a tonic pronoun, referred to as 'non-clitic' (in (2a)and (3a) in section 2). The second was used for cases in which the subject of the embedded clause is realized as a clitic, referred to as 'clitic' structures (see (2b) and (3b) in section 2). Each grammatical structure was further split into *para*, the prescriptive, formal spelling, and its informal variant, *pra*, as we were also aiming to investigate the effect of formality. These queries follow in Table 3:

Table 2. Non-clitic query structures and examples.

'para'	*Query*	[tag="V.*"]	[lemma="para"]	[lemma="nós|tu|vós|vocês|eles|elas"]	[tag="V.N.*"]
	KWIC example	*economizer*	*para*	*nós*	*casarmos*
'pra'	*Query*	[tag="V.*"]	[lemma="pra"]	[lemma="nós|tu|vós|vocês|eles|elas"]	[tag="V.N.*"]
	KWIC example	*pediu*	*pra*	*nós*	*irmos*

These strings search for any verb ("V.*") of indefinite length (*), followed by *para* or *pra* to find infinitives. Then we searched for an option of any one of the listed pronouns separated by (|) ("nós|tu|vós|vocês|eles|elas") to follow, finishing with the potential location of the inflected infinitive: an infinitive verb ("V.N.*"). These queries follow in Table 4:

Table 3. Clitic query structures and examples.

'para'	*Query*	[lemma="as|os|te|nos|vos|lhes"]	[tag="V.*"]	[lemma="para"]	[tag="V.N.*"]
	KWIC example	*nós*	*serviu*	*para*	*lembrar*
'pra'	*Query*	[lemma="as|os|te|nos|vos|lhes"]	[tag="V.N.*"]	[lemma="pra"]	[tag="V.N.*"]
	KWIC example	*te*	*pago*	*pra*	*viver*

The clitic strings have the same functions but in a different order. These strings retrieve one of the listed clitic pronouns ("as|os|te|nos|vos|lhes"), followed by any verb ("V.*"), followed by *para* or *pra*, and then finally the location where the inflected infinitive could emerge: the infinitive verb ("V.N.*").

When these queries were run, they were filtered based on some of the annotations already available in the corpus. The SketchEngine website gives users an intuitive graphical interface to make filtering searches more streamlined. This makes it possible to narrow a given query by genre, website extension, and other variables of interest. During our query process, each round of querying was partitioned into four genres: 'blog', 'news', 'encyclopedia', and 'legal'. At the same time, because SketchEngine creates corpora by scrapping data from internet websites, we also had website country extensions available. These were used simultaneously with genre filters to proxy for variety: '.br' for Brazilian Portuguese (8,010,603,604 tokens) and '.pt' for European Portuguese (893,179,245 tokens) (totals do not add to 100% because this corpus includes other country extensions). The total sizes of genres for each subcorpora can be found in Table 4, below.

Table 4. Subcorpora size by variety and genre.

Genre	Brazilian Portuguese	European Portuguese
Blog	648,643,773	43,366,313
News	1,171,751,189	893,179,245
Encyclopedia	354,599,657	13,161,382
Legal	132,950,188	73,149,935
TOTALS	8,010,603,604	893,179,245

Source: (Kilgarriff et al. 2014)

Following the initial queries, the data were returned in columnar format, with the words corresponding to the queries in a Key-Word-in-Context (KWIC) column, surrounded by the left and right context of the full sentence, along with the website source. These columns were downloaded from SketchEngine as .csv (comma separated values), an excel format with each type of information separated by commas. The large corpus size of the ptTenTen20 offered a large volume of data, (total count $N = 12{,}450$), which would be unfeasible to hand-code. To streamline this process, instead of hand-coding 12,450 sentences manually, we used the Tidyverse package in R (Wickham et al. 2019) to label each point with the variables under investigation. The ultimate result of the coding process was a dataframe, a large table with a column corresponding to each variable of analysis. These were all coded into the final dataframe using iterative blocks of code to populate columns with the necessary

information. This process is outlined below, concluding the section with a chart displaying how the final dataframe looked with all finalized variables.

Four of the meta-variables ultimately used for analysis were already available from the filters used during the querying stage: variety, genre, structure, and type. Variety was already pre-defined by which website extension was used, either 'Brazilian' (.br) or 'Portuguese' (.pt) for European Portuguese. Genre was coded as a four-level choice containing 'legal', 'news', 'encyclopedia', and 'blog' (more below for why we ended up with five genres). Three of these original genres were coded as formal: 'legal', 'encyclopedia', and 'news', whereas 'blog' was used to represent informal data.

During the process of coding for genre, some data cleaning was found to be needed. Manual examination of some of the websites in the 'blog' genre revealed that some websites were providing unexpectedly formal results, as judged by the native Brazilian Portuguese author. The sources of these formal data turned out to be Bible quotes from religion-focused blogs (espada.eti.br), from government websites like the Brazilian supreme court (stf.jus.br), or blogs providing legal advice (jus.com.br). The unusual formality of these websites was deemed to be a miscategorization by the SketchEngine web-scraper and thus entries from any religious or legal blog were switched to be 'formal' under the register column. However, this resulted in just over a third of the blog data being switched to formal (1415 formal blog entries out of total 4154 blog entries). To retain the distinction between informal and formal registers, these re-coded formal blog websites were specified as 'BlogFormal' under the genre column to allow for analysis, providing a fifth option under the genre column. The resulting number of individual hits for the final iteration of the genre variable are shown below (Table 5):

Table 5. Hits per register and genre.

Register	**Informal**	**Formal**			
Genre	Blog	BlogFormal	News	Encyclopedia	Legal
Hits	2739	1415	7286	834	176

Also available straight from the querying process, structure and type were known based on the different pronouns used in each query, clitic or non-clitic. These four variables, Brazilian or European variety, formal (news, legal, encyclopedia, formal blog) or informal register (blog), structure (non-clitic or clitic), and type (*pra* or *para*), were all added to the correct dataframes as hard-coded variables. The remaining variables, the pronoun, the infinitive verb itself, and the conjugation of the infinitive verb, all had to be extracted from the information given in the KWIC column. These were all handled using regular expressions and if loops to match search for strings of information and populate each variable's column if the strings matched our given criteria.

Pronouns were extracted from the KWIC column using the regular expressions presented in Table 6:

Table 6. Regular expressions used to extract pronouns.

Clitic status	*para*	*pra*
Clitic	.{2,4}(?=\\s)	.{2,4}(?=\\s)
Non-clitic	(?<=\\spara\\s).*?(?=\\s)	(?<=\\spra\\s).*?(?=\\s)

The clitic expressions tell the computer to find a word in the location of the period (.), with a length from 2-4 letters that is followed by a space. This word corresponds to the first entry in each clitic KWIC, which is always one of the pronouns designated in the query. The non-clitic expressions search for a word (.) immediately after the word *para* or *pra* and a space (\\s), followed by another space (\\s). This extracted the third word in the non-clitic KWIC, which is always one of the designated pronouns. Table 7 shows the resulting number of each type of pronoun divided across structure type. No 3PL pronouns were found.

Table 7. Hits per pronoun, subdivided by pronoun type.

Structure	Non-clitics			Formal		
Pronoun	nós	tu	vós	nos	te	vos
Hits	428	59	4	10017	1809	133

Similar to how pronouns were extracted, the infinitive verb that could be an inflected infinitive also had to be copied and placed in its own column. This was similarly done with regular expressions as seen below:

Table 8. Regular expressions used to extract verbs.

Clitic status	*para*	*pra*
Clitic	(?<=\\bpara\\s).*	(?<=\\bpra\\s).*
Non-clitic	(?<=nós|tu|vós|eles|las|nos|vos).*	(?<=nós|tu|vós|eles|las|nos|vos).*

The clitic searches looked for the word (.) of any length (*) immediately following the word *para* or *pra*. The non-clitic search looks for a word of any length immediately after any of the pronouns. Once this verb was extracted (column called realization), an *if* loop was used to look at the ending of the verb to find conjugations. If the ending of the infinitive verb in the realization column matched a specific conjugation (inflected infinitive), it was populated in another column called person_of_infinitive, containing the matching person and number. The results of the person_of_infinitive are listed in Table 9:

Table 9. Person_of_Infinitive.

Outcome	Uninflected infinitive	Inflected infinitive		
Conjugation	No conjugation	1PL	2SG	2PL
Hits	10566	1710	143	31

Finally, once this was coded, the presence or absence of the inflected infinitive could be defined with *if* logic: if the conjugation column was populated by a conjugation, it was labeled in another column as an inflected infinitive. If not, it was

labeled an uninflected infinitive. This final column was called the outcome and could now be used for statistical analysis.

The tables below (10A and 10B) show an example of data in the final dataframe with all coded variables. The reference column contains the website from SketchEngine containing the hit returned by our search. KWIC (Key Word in Context) corresponds to the exact four words that were extracted from a sentence using the CQL queries, with the wider context of the sentence in the Left and Right columns.

Table 10.A. Final Dataframe.

Reference	Left	KWIC	Right
xr.pro.br	se torna cada vez mais impraticável…	nos equipou para competirmos	pela nossa perpetuação, mas as atitudes…

Table 10.B. Final Dataframe, including coded variables.

Variety	Genre	Struc-ture	Type	Realization	MatrixVerb	Person_of _Infinitive	Register	Outcome
Brazilian	Blog	Clitic	*Para*	*competirmos*	*equipou*	1PL	Informal	Inflected Infinitive

The final number of observations totaled $N = 12{,}450$, now coded for the three independent predictor variables (Tables 5, 7, and 9): pronoun, genre, and presence of the inflected infinitive. At this point, a fixed-effects logistic regression model was applied to the data using the glm function in the Lme4 R package (Bates et. al 2015). AIC (Akaike Information Criterion) values were then compared to select the optimal model with the lowest AIC, found by taking the variation of the dataset that the model does not explain plus a penalty value that increases with more variables. Comparing different sets of variables in this way for the lowest AIC (the model with the best explanatory power with the lowest penalty) ensures that the model is parsimonious by not over-describing the data with too many variables.

5. Results

The goal of the logistic regression was to answer these research questions:

Q1. Variety: How do varieties of Portuguese differ in their use of the inflected infinitive?
Q2. Register: Is the inflected infinitive only found in formal registers of BP?
Q3. Structure: Is there variation in cases when the pronoun is a clitic?

The results of the logistic regression are shown in Table 11. Of interest here are the estimate and *p*-value coefficient numbers reported for each variable down the far left and far right columns, respectively. The estimate column shows the effect of each variable as a predictor of the outcome, as compared to a reference or baseline value. Estimates are interpreted as an increase (positive estimates) or decrease (negative estimates) in the likelihood of the inflected infinitive being present, as compared to the reference.

Note that some variables do not occur in Table 11, including the informal blog genre, the type *para*, the Brazilian variety, and the pronoun *nos*. The logistic regression assesses each variable in comparison to a base or reference value for that category to set a relative benchmark for comparison. The variables that are shown in the table describe the change in outcome with respect to these reference values. For example, for variety, Brazilian Portuguese is the reference variable, and the estimate for European Portuguese is positive, meaning that the inflected infinitive is more likely to be found in EP than in BP. On the far right, the *p*-value column shows significant results, using an alpha of 0.05. Only four predictors, the 'encyclopedia' and 'legal' genres, and the pronouns *vos* and *vós* did not significantly predict the variation of the inflected infinitive.

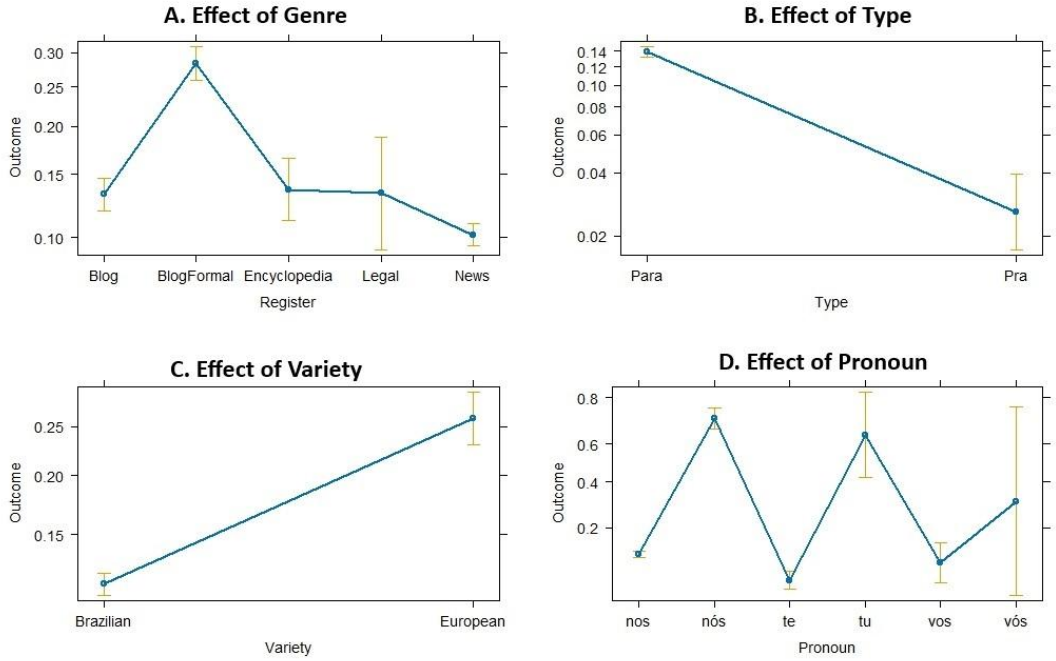
Table 11. Logistic regression model results for glm(formula = Outcome ~ Genre + Variety + Pronoun, family = "binomial", data = OutcomeBoth)

Variable	Estimate	Std. Error	*z*-value	odds-ration	*p*-value
Intercept	−1.8534096	0.0620604	−29.865	0.1567020	< 2e−16 ***
Genre BlogFormal	0.9411410	0.0861643	10.923	2.5629040	< 2e−16 ***
Genre Encyclopedia	0.0386357	0.1257303	0.307	1.0393918	0.759
Genre Legal	0.0001885	0.2185294	0.001	1.0001885	0.999
Genre News	−0.2940687	0.0704685	−4.173	0.7452253	3.01e−05 ***
TypePra	−1.6495627	0.1671095	−9.871	0.1921339	< 2e−16 ***
Variety European	0.9676826	0.0862557	11.219	2.6318385	< 2e−16 ***
Pronoun nós	2.9703608	0.1177321	25.230	19.4989545	< 2e−16 ***
Pronoun te	−0.5644247	0.0988443	−5.710	0.5686872	1.13e−08 ***
Pronoun tu	2.2454812	0.3149996	7.129	9.4449594	1.01e−12 ***
Pronoun vos	−0.1653174	0.2167697	−0.763	0.8476246	0.446
Pronoun vós	1.1411341	1.0212987	1.117	3.1303165	0.264
Signif codes: 0 '***' 0.001 '**' '*' 0.05 '.' 0.1 ' ' 1					

As noticeable from the *p*-values, all the variables of interest for the research questions showed statistical significance. The inflected infinitive is more likely to be used when the variety is European Portuguese ($p < 2e{-}16$), in formally written blogs ($p < 2e{-}16$), and with the non-clitics *nós* ($p < 2e{-}16$) and *tu* ($p < 1.01e{-}12$). The remaining predictors turn out to condition against finding the inflected infinitive, with the news genre showing a negative estimate ($p < 3.01e{-}5$), as well as the clitic *te* ($p < 1.13e{-}8$). Additionally, in accordance with the second hypothesis that the inflected infinitive would be less likely in informal register, we also found that contexts with *pra*, the informal variant, disfavor the inflected infinitive ($p < 2e{-}16$).

Figure 1 visualizes effect sizes. Figure 1.A shows the likelihood of finding the inflected infinitive in each of the genre, with 'BlogFormal' showing the largest effect size. Figure 1.B graphs the effect of *para* vs. *pra*, with para much more likely to coexist with the inflected infinitive, which also accords with the second hypothesis regarding formality. Figure 1.C shows the positive effect of European Portuguese. Finally, Figure 1.D shows spikes with *nós* and *tu*, the non-clitic pronouns, as much likely than the clitic-pronouns to condition the inflected infinitive.

Figure 1. Effect size by variable.



5. Conclusions

The goal for this study was to explore how a relatively uncommon feature of the Portuguese language is used by its speakers using a large data sample. For this reason, our research questions and hypotheses focused on variety, register, and the presence of variation of the inflected infinitive with clitic vs non-clitic structures. The first two research questions dealt with variety and register. Based on the observation made by Cunha and Cintra (2016) and the claims made by Pires and Rothman (1999), discussed in sections 1 and 2, we hypothesized that the inflected infinitive would be more common in EP and formal registers. Finally, based on observations made by Cunha and Cintra (2016), we predicted that there would be variation between the inflected and uninflected infinitive with clitic structures. The results presented above are insufficient to uphold the null hypothesis, and therefore provide evidence, of varying degrees, to the hypotheses regarding variety, register, and structure.

Our first research question asks how varieties of Portuguese differ in their use of the inflected infinitive and the first hypothesis states that EP will favor the inflected infinitive. Part of understanding how these varieties differ in their use of the inflected infinitive also entails understanding how they differ in their use of pronouns. While EP uses *tu* for the 2nd person singular, BP also makes use of *você* which takes a 3rd person singular ending. This would seem to give EP an advantage in our dataset; however, for the 59 instances of *tu* in the data, only eight are from EP, 100% of which occur with an inflected infinitive, such as in example (4), below. In comparison to the 14/51 (27.45%) examples of the inflected infinitive with *tu* in BP (cf. example 5). There are only four examples of *vós* in the data, all from BP, half of which occur with the inflected infinitive. Example (6) from the data, seen below, shows three inflected infinitives, meaning that four of the six infinitives that appear with vós show person and number agreement. Interestingly, three of the four examples of *vós* all come from

the same website, rainhamaria.com.br. Curiously, the sentence from this website that does not show agreement, example (7), has an almost identical structure up to the infinitive as example (6), with multiple inflected infinitives.

(4) European Portuguese, (sapo.pt)

houve	um	que	me	disse	“Isto
exist.PST.3SG	one	REL	me	say.PST.3SG	this
é	para	tu		aprenderes	a
be.PRS.3SG	for	you.2SG		learn.INF.2SG	to
não	escrever	como	escreves.”		
NEG	write.INF	as	write.PRS.2SG		

‘There was one who told me, ‘this is for you to learn not to write as you write.’’

(5) Brazilian Portuguese, (overmundo.com.br)

como	é	que	foi	pra	tu
how	be.PRS.3SG	COMP	go.PST.3SG	for	you.2SG
começares	a	desenhar			
begin.INF.2SG	to	draw.INF			

‘What was it like for you to start drawing?’

(6) Brazilian Portuguese, (rainhamaria.com.br)

de	que	adianta	para	vós	seguirdes	
of	what	serve.PRS.3SG	for	you.2PL	follow.INF.2PL	
as	Minhas	Leis	fazerdes	a	Minha	Vontade
the	my	laws	do.INF.2PL	the	my	will
mas	não	praticardes	os	Meus	Ensinamentos	
but	NEG	practice.INF.2PL	the	my	teachings	

‘What is the point of you following My Laws, doing My Will, but not practicing My Teachings?’

(7) Brazilian Portuguese, (rainhamaria.com.br)

de	que	adianta	para	vós		
of	what	serve.PRS.3SG	for	you.2PL		
acumular		tantos		tesouros	na	terra
accumulate.INF		so.many		treasures	on	earth

‘What is the point of you accumulating so many treasures on earth?’

This unexpected contradiction regarding the presence of certain pronouns in the data is likely due to the size difference in the subcorpora, but despite this we find that, just as our hypothesis predicted, EP significantly favors the inflected infinitive. It is worth noting that *vós* and *vos* are no longer used in standard Portuguese.[3] This accounts for the low number of occurrences for these pronouns in the data, which can likely be attributed to quoted archaic usage (for example, many blogs of a religious nature contained quotes from the Bible) and possibly writing from nonstandard varieties of Portuguese. However, since our research question regarding variety was

[3] We would like to acknowledge a reviewer’s observation that *vós* and *vos* no longer exist in standard Portuguese.

specific to regional variety, and not standardization, we did not classify the data according to standard and nonstandard varieties nor consider this as a variable.

In Table 12, below, we see that EP used an inflected infinitive in a formal register 27.49% of the time, compared to 11.86% for BP. Interestingly, despite Pires and Rothan's (2009) claim that the inflected infinitive has been lost in colloquial BP, we see that BP uses the inflected infinitive more than EP in the data for informal register. There is a possibility that the frequency of the inflected infinitive has decreased in colloquial BP, but our results confirm that, as of 2020, it was still in use in informal BP.

Table 12. Total and inflected numbers by variety and register.

Register	Brazilian		European	
	Total	*Inflected (%)*	*Total*	*Inflected (%)*
Formal	7583	899 (11.86)	713	196 (27.49)
Informal	3788	743 (19.61)	366	46 (12.57)

The second part of the claim from Pires and Rothman (2009) is that the inflected infinitive should only be found in formal registers, since it is acquired through school and media. Although the data regarding register supports our hypothesis, they do not provide evidence as straightforward as what we find for variety. The second research question asks if the inflected infinitive is only found in formal BP, with the second hypothesis predicting that informal registers will disfavor the inflected infinitive. The data suggest that this is the case, as the formal register favors the inflected infinitive. Examples (8) and (9), below, as well as (5), above, show three of the numerous examples of the inflected infinitive that come from blog (informal) websites in Brazil.

(8) Brazilian Portuguese, (mesquita.blog.br)
"Um dos nossos clientes... nos pediu
one of.the our clients us ask.PST.3SG
para corrigir**mos** os nossos números",
for correct.INF.1PL the our numbers
acrescentou.
add.PST.3SG
'"One of our clients…asked us to correct our numbers," he/she added.'

(9) Brazilian Portuguese, (unicamp.br)
por isso a família da Maria Lúcia nos
for this the family of Maria Lúcia us
convidou para ir**mos** até a Unicamp
invite.PST.3SG for go.INF.1PL to the Unicamp
'That's why Maria Lúcia's family invited us to go to Unicamp.'

As seen in Table 12, above, in our data there are 3788 hits for informal blogs from Brazilian Portuguese, 743 (19.61%) of which employ the inflected infinitive. Interestingly, the percentage of inflected infinitives from formal BP data is lower than from informal BP, with 899 (11.44%) of 7853 occurrences. However, this makes sense

when considering that only formal blogs significantly favored the inflected infinitive, while the genres 'legal' and 'encyclopedia' were insignificant predictors, and news disfavored the inflected infinitive. When we look at the rate of the inflected infinitive in BP formal blogs, we find it in 409 (29.38%) of the 1392 tokens. As for the hypothesis, however, we tentatively suggest that informal registers disfavor the inflected infinitive, since our model took blog as a reference value. Nor did the formal register as a whole favor the inflected infinitive, given the finding for news.

The nature of the data and limitations of our study may explain why the formal genres fail to support, and even contradict, our hypothesis. Although our queries included 3rd person tonic and clitic pronouns, they yielded no hits, meaning the data we present here is limited to the 1st and 2nd person pronouns *nós*, *nos*, *tu*, *te*, *vós*, and *vos*. These pronouns are less likely to occur in news, legal, and encyclopedia writing due to their reportative, and less personal nature, with certain exceptions, such as direct quotes (see example 4, above)[4]. The lack of 3rd person in our data can be explained by the fact that our query syntax was limited to pronouns, therefore the data does not include examples where a full noun phrase is the subject of the infinitive, something we would expect to see in the genres that we have considered to be part of a more formal register. The significance of the recategorized BlogFormal data supports our hypothesis; however, data that includes full NPs, providing a more accurate depiction of the formal register as represented by 'news', 'encyclopedia', and 'legal' genres, is needed in order to clarify how the formality of register interacts with the frequency of the inflected infinitive. Taking into consideration that Cabral (2006) also reported a low rate of the inflected infinitive in the writings of two Brazilian newspapers, our data still challenges previous assumptions about the rate of use of the inflected infinitive in the 'news' genre, despite it being considered a more formal register.

The final research question and hypothesis were about whether or not there would be variation in structures with clitics, as Cunha and Cintra (2016) had claimed. As seen in Table 7, 96.05% of the data comes from constructions with clitic pronouns and 13.04% of these use the inflected infinitive. In comparison, 65.98% of the data using tonic pronouns use the inflected infinitive. Table 13, below, shows a breakdown of the data by variety, structure, and individual pronoun. We confirm that there is empirically evidenced variation in constructions with clitics, as seen in (9) above, as well as (10) and (11), below; however, as expected based on Ladeira (1986), there is considerably less variation in these structures when compared to non-clitic structures.

(10) Brazilian Portuguese, (xr.pro.br)

o	Proto-Ciúmes	é	um	dos	mecanismos	
the	Proto-Jealousy	be.PRS.3SG	one	of.the	mechanisms	
que	a natureza	nos	equipou			
REL	the nature	us	equip.PST.3SG			
para	competir**mos**	pela	nossa	perpetuação		
for	compete.INF.1PL	for.the	our	perpetuation		

'Proto-Jealousy is one of the mechanisms that nature equipped us with to compete for our perpetuation.'

[4] We would like to thank a reviewer for observing that 1st and 2nd person pronouns are less likely in the news genre.

(11) European Portuguese, (sapo.pt)

se	alguém	te		disser	
if	someone	you.2SG		say.FUT.SUBJ.3SG	
para	pore**s**	sal	no	teu	café
for	put.INF.2SG	salt	in.the	your	coffee
presumo	que	não	o	irás	fazer
presume.PRS.1SG	that	NEG	it	go.FUT.2SG	do.INF

'If someone tells you to pore salt in your coffee, I assume you won't do it.'

In fact, looking at the results from Table 11 specifically for the clitics *te* and *vos* (*nos* was used as a reference by our model), we see that *te* significantly disfavors the inflected infinitive and *vos*, while not significant, also trends negative. The lack of significance found here is likely due to the small number of hits we found.

Table 13.A. Total and inflected numbers by structure and pronoun.

Structure	**non-clitic**					
Pronoun	*nós*		*tu*		*vós*	
Total/Inflected(%)	428	300 (70.09%)	59	22 (37.29%)	4	2 (50.00%)
Total/Inflected(%) non-clitics	491		324 (65.99%)			
Structure	**clitic**					
Pronoun	*nos*		*te*		*vos*	
Total/Inflected(%)	10017	1401 (13.99%)	1809	130 (7.19%)	133	29 (21.80%)
Total/Inflected(%) clitics	11959		1560 (13.04%)			

Table 13.B. Total and inflected numbers by structure and pronoun for BP.

Brazilian Portuguese						
Structure	**non-clitic**					
Pronoun	*nós*		*tu*		*vós*	
Total/Inflected(%)	390	269 (68.97%)	51	14 (27.45%)	4	2 (50.00%)
Total/Inflected(%) non-clitics	445		285 (64.04%)			
Structure	**Clitic**					
Pronoun	*nos*		*te*		*vos*	
Total/Inflected(%)	9284	1273 (13.71%)	1542	56 (3.63%)	100	28 (28.00%)
Total/Inflected(%) clitics	10926		1357 (12.42%)			

Table 13.C. Total and inflected numbers by structure and pronoun for EP.

European Portuguese						
Structure	**non-clitic**					
Pronoun	*nós*		*tu*		*vós*	
Total/Inflected(%)	38	31 (81.58%)	8	8 (100%)	0	-
Total/Inflected(%) non-clitics	46		39 (84.78%)			

Structure	clitic					
Pronoun	*nos*		*te*		*vos*	
Total/Inflected(%)	733	128 (17.46%)	267	74 (27.72%)	33	1 (3.03%)
Total/Inflected(%) clitics	1033		203 (19.65%)			

As reported in Table 7, and shown again here in Table 13 above, only 4 hits were found in total for *vós*, and only 133 for *vos*, with both categories being two of least-frequent pronouns in our data set. The non-clitic pronoun *tu*, also has a low rate of occurrence in the data, since as mentioned above, it has been replaced by *você* in BP; its clitic counterpart *te* is well represented in the data since it is still commonly used in BP. Relatedly, *vós* and *vos* are primarily found in EP and may be rare in the syntactic structures we used to extract the inflected infinitive. These facts might contribute to their low frequency and explain why no significant effects were found, particularly in the case of *vós*.

Looking back at Table 11 from the results section, although *vos* and *vós* showed no significance, they do trend in the same direction as their singular counterparts *tu* and *te*, which are significant with *tu* favoring the inflected infinitive, and *te* disfavoring it. This suggests a larger picture could be found with clitics in general disfavoring the inflected infinitive but needs further data to explore more thoroughly. In fact, as seen in Table 11, non-clitics have very high odds ratios for the inflected infinitive, with *nós* at 19.4989545 and *tu* at 9.4449594. Odds ratio values between 0 and 1 are negative predictors, making the outcome less likely. Any values above 1 condition for outcomes favorably. Compared to the reference for pronouns (the clitic form *nos*), as calculated from the high odds ratios, the two non-clitics *nós* and *tu* are 3,798 times and 1,700 times more likely to be found with the inflected infinitive, respectively. The high positive estimate and odds ratio for *nós* in particular align with previous studies by Ladeira (1986) and Canever (2012) that noted a preference for the inflected infinitive with a first-person pronoun, with the Canever study reporting 90% of cases of infinitives associated with first-person pronouns showing inflection. Lastly, the higher odds-ratio for non-clitics with the inflected infinitive aligns with the fact that clitics are theoretically unable to trigger agreement with the verb. Prescriptively, the inflected infinitive licenses nominative subjects, which explains the higher odds ratio for non-clitics in this context.[5]

6.1. Limitations and future directions

As noted above, during manual inspection of the results, we decided to modify the initial categorization of certain websites, which were judged to be the opposite of what SketchEngine had originally tagged. This reveals that the automatic tagger used by SketchEngine might have mislabeled or double-labeled websites that were edge cases, such as religious or legal advice blogs. As mentioned previously, corpus data is a powerful tool for examining language variation and usage; however, its reliability

[5] We would like to acknowledge the reviewer's observation that variation involving clitics may occur because, prescriptively, the inflected infinitive should license nominative subjects. However, speakers appear to accept accusative clitics with the inflected infinitive, which could explain the observed variation.

depends heavily on how the corpus was sampled, as well as the accuracy of the tagging and metadata. Although some of the blogs found on SketchEngine contained more formal language, the data required little cleaning.

Another variable of interest that can be further explored is the effect of the matrix verb. This data was coded in the final data set under the 'matrix verb' column, extracted from the non-restricted first verb in the query (i.e., it did not have to be an infinitive), but to evaluate the effect of the matrix verb on the inflected infinitive, a subset would be needed due to the high amount of variation. Without sub-setting the verbs, treating the matrix verb as a random effect involved too many variables to be able to account for. We suspect that causative matrix verbs will have a strong preference for the inflected infinitive.

7. Conclusions

This study uses a corpus-based approach to provide a quantitative analysis of the effects of variety, register, and structure on the use of the inflected infinitive in Portuguese. Heavily debated among grammarians, the inflected infinitive can be found in a variety of syntactic structures, and authors suggest that use of the inflected infinitive may be based on personal preference for stylistic choices (Cunha & Cintra 2016). Some studies show that semantics may play a role in determining inflected infinitive variation (Bossaglia 2013).

We found significant results to reject the null hypothesis. Variety has significant effects on the choice between the inflected infinitive and uninflected forms, with the inflected infinitive favored in European varieties. Furthermore, the clitics *nos*, *te*, and *vos*, condition against the inflected infinitive. This confirms our expectation that variation would be seen with clitics, as claimed by Cunha and Cintra (2016), although this structure was previously considered prescriptively ungrammatical.

Previous claims that the inflected infinitive has already been lost in colloquial BP (Pires & Rothman 2009, and previous studies by the same authors) are contradicted by other studies suggesting instead that it is expanding to structures that were previously thought to be ungrammatical (Cabral 2006, Canever 2012). In the context of the clitic and European Portuguese preference, it is important to note that we do find instances of the inflected infinitive in Brazilian-written blogs from data as recent as 2020.

Finally, all the data used in this study were from the ptTenTen20, which contains data from websites obtained from June to November 2020 (Kilgarriff et al. 2014). One advantage of the TenTen corpora series is that they include different subcorpora for different years of sampling, including the ptTenTen11 with data from year 2011 and 2018. This study also provides rich grounds for future work on the data gaps seen here, including the need for more information about *vos* and *vós*. Future studies should consider a diachronic study comparing results from the ptTenTen 11 (2011) and the ptTenTen20 (2020). This represents almost a decade of diachronic difference and has the potential to conclusively support or deny claims that use of the inflected infinitive is in decline or expanding to new syntactic forms.

Acknowledgments

We gratefully thank the members of UGA's Digilab, especially Michael Wolfmann, Keiko Bridwell, and Larissa Do Carmo Inacio for their extensive help with R coding, statistical expertise, and advice on data management. Further thanks go to Dr. Chad Howe and Dr. John Hale for their mentorship and dedicated guidance.

References

Bates, Douglas, Maechler, Martin, Bolker, Ben, & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1): 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Bossaglia, Giulia. 2013. Inflected/non-inflected alternation in causative and perception constructions of contemporary European Portuguese: A corpus-based study. *Procedia: Social and Behavioral Sciences* 95: 220–230. https://doi.org/10.1016/j.sbspro.2013.10.642

Cabral, Grasiela de Sousa. 2006. A concordância variável do infinitivo na estrita padrão. M.A. thesis, Universidade Federal do Rio de Janeiro. https://doi.org/10.1590/S1981-57942012000300013

Canever, Fernanda. 2012. Evidências para um modelo de língua baseado no uso: O infinitivo flexionado em português brasileiro. M.A. thesis, Universidade de São Paulo.

Canever, Fernanda. 2017. Infinitivo flexionado em portugês brasileiro: Fraquência e percepções sociolinguísticas. Ph.D. thesis, Universidade de São Paulo. https://doi.org/10.11606/T.8.2018.tde-04042018-184015

Canever, Fernanda, & Ronald Beline Mendes. 2019. *Infinitive verbs, verbal agreement and perceived competence / Verbos infinitivos, concordância verbal e competência percebida. Revista de Estudos da Linguagem* 27(4): 1671–1700. https://doi.org/10.17851/2237-2083.27.4.1671-1700

Cunha, Celso, & Lindley Cintra. 2016. *Nova gramática do português contemporâneo*. Rio de Janeiro: Lexikon Editora Digital.

Kilgarriff, Adam, Jakubíček, Miloš, Pomikalek, Jakubiček, Sardinha, Tony Berber, & Pete Whitelock. 2014. PtTenTen: a corpus for Portuguese lexicography. In T. B. Sardinha, & T. de Lurdes São Bento Ferreira (eds), *Working with Portuguese Corpora*, 111–130.

Ladeira, José Dionísio. 1986. A flexão do infinitivo em português. Ph.D. thesis, Universidade Federal do Rio do Janeiro.

Madeira, Ana María. 1994. On the Portguese inflected infinitive. In G Borgato (ed.), *Teoria del Linguaggio e Analisis Linguistica: XX Incontro di Grammatica Generativa,* 289–314. Padova: Unipress.

Martins, Ana Maria, & Ernestina Carrilho (eds). 2016. *Manual de Linguística Portuguesa.* Berlin/Boston: De Gruyter, Inc., ProQuest Ebook Central.

Maurer, Theodore Henrique Jr. 1968. *O infinitivo flexionado português: Estudo histórico-descritivo.* São Paulo, SP: Cia. Editora Nacional.

Pires, Acrisio, & Jason Rothman. 2009. Disentangling Sources of Incomplete Acquisition: An explanation for competence divergence across heritage grammars. *Iternational Journal of Biligualism* 13(2): 211–238. https://doi.org/10.1177/1367006909339806

Raposo, Eduardo. 1987. Case theory and Infl-to-Comp: The inflected infinitive in European Portuguese. *Linguistic Inquiry*, 18(1): 85–109.

Rodrigues, Cilene, & Norbert Hornstein. 2013. Epicene Agreement and Inflected infinitives when the data is "Under Control": A reply to Modest (2010). *Syntax*, 16(3): 292–309.

Wickham, Hadley, Averick, Mara, Bryan, Jennifer, Chang, Winston, D'Agostina McGowan, Lucy, François, Romain, Grolemund, Garrett, Hayes, Alex, Henry, Lionel, Hester, Jim, Kuhn, Max, Pedersen, Thomas Lin, Miller, Evan, Bache, Stephan Milton, Müller, Kirill, Ooms, Jeroen, Robinson, David, Seidel, Dana Paige, Spinu, Vitalie, Takahashi, Kohske, Vaughan, Davis, Wilke, Claus, Woo, Kara, & Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43): 1686. https://doi.org/10.21105/joss.01686