Language Variation and Change, **31** (2019), 69–89.
© Cambridge University Press, 2019 0954-3945/19 $16.00
doi:10.1017/S0954394519000036

Variable past participles in Portuguese perfect constructions

SCOTT A. SCHWENTER, MARK HOFF, ELENI CHRISTODULELIS, CHELSEA PFLUM AND ASHLEE DAUPHINAIS
The Ohio State University

ABSTRACT

Some Portuguese verbs have two past participle forms: one regular (stem + *-do*), and the other irregular (often identical to the first person singular present indicative). Per grammars, the perfect auxiliaries *ter/haver* take regulars, while irregulars appear with passive/adjectival *ser/estar*. To test these claims, we analyzed naturally-occurring data from Brazil (twentieth century) and Portugal (nineteenth and twentieth). We coded 1077 tokens from 21 verbs for ten predictors and performed mixed-effects logistic regressions in R. Irregulars appear with *ter/haver* 54% overall and in 68% of cases from Portugal. Our results demonstrate that past participle choice is determined by the interaction of several linguistic factors. While lexical verb is the most significant predictor of participle selection, verbs with irregular participles identical to the first person singular present indicative occur in the irregular significantly more than other verbs. We conclude that analogical processes underpin the variation. This conclusion allows us to adjudicate between competing accounts of past participle choice in Portuguese.

A select group of verbs in Portuguese have competing forms for their past participle, giving speakers the choice between a regular and an irregular form. Together, these forms are called *particípios duplos* 'double participles' or "short-form" participles (Lobato, 1999) when using these *verbos abundantes* ('abundant verbs'). The regular past participle simply adds *-do* to the verb stem as in *pagar* 'to pay' > *pagado* 'paid' or *entregar* 'to deliver' > *entregado* 'delivered.' By contrast, the irregular participle in many cases is equivalent to the first person singular (hereafter referred to as 1SG) present indicative form of the verb as in *pagar* > *pago* (which is the same form as the 1SG present indicative *Eu pago* 'I pay'). Alternatively, it may be a form that is the same as another verb form, such as the first person singular and third person singular present subjunctive form as in *entregar* > *entregue*.

Although there is general agreement among grammarians regarding the existence of these verbs and their competing participles (Lima, 1972; Mateus,

We are grateful to the other participants in Scott Schwenter's Portuguese 5611 class at The Ohio State University in Spring 2016 for their invaluable help on this project. Thanks as well to the audiences at NWAV 45 in Vancouver (2016) and ICLaVE 9 in Málaga (2017), and to Patrícia Amaral, Ana Carvalho, Janayna Carvalho, Luana Lamberti, Lachlan Mackenzie, Marie-Eve Ritz, Malte Rosemeyer, Kim Schulte, Nigel Vincent, and three anonymous LVC reviewers for helpful feedback.

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

2003; Perini, 2002, 2010; Raposo, Bacelar do Nascimento, Coelho da Mota, Segura, & Mendes, 2013), there is little consensus about exactly which verbs constitute the group of *verbos abundantes*, that is, those which provide the option of *particípios duplos*. Additionally, not all authors agree on the distribution of participles by auxiliary. Most scholars claim that the morphologically regular *-do* forms are used strictly or primarily to form the perfect tenses, combining with the perfect auxiliaries *ter* and *haver* (both meaning 'to have'), while the irregular forms are used in the passive and as adjectives that can be inflected for gender and number with the copulas *ser* and *estar* (both meaning 'to be') (Carvalho & Bagno, 2015; Cunha & Cintra, 2001; Ganho & McGovern, 2004; Prista, 1966; Thomas, 1969; *inter alia*).[1] Typical statements to this effect can be found in the classic grammars by Lima (1972:171): "the regular participle is only used with the verb *ter*; and the irregular, either with isolated adjective value, or with the verb *ser*"[2]; or Bechara (2006:230): "In general the regular form is employed, which remains invariable with the auxiliaries *ter* and *haver*, in the active voice, and the irregular form … with the auxiliaries *ser*, *estar* and *ficar*, in the passive voice."[3] Other authors recognize, however, that the situation is considerably more nuanced. Simões (2008:356) and Perini (2002:153–4), for example, documented the possibility of using the irregular forms with *ter* and *haver* in modern Brazilian Portuguese.[4] They also implied that the form of the irregular verb itself plays an important role in participle selection, namely, that higher rates of irregular forms occur if the irregulars match the 1SG present indicative form (presumably via an analogical process), even following *ter* and *haver*. Perini (2010:178) simply stated that, "It is … a tendency, with great variation" without providing any further explanation for the variable patterns.[5] Although the perspectives of these grammars provide a valuable starting point for further research, a comprehensive description of the variation in or explanation(s) for participle selection have yet to be proposed.

In this study, our goal is to reconcile via empirical means the differing views that both grammarians and linguists hold on the use of *particípios duplos* by looking at naturally-occurring corpus data from both Brazilian and European Portuguese. We assess these descriptions of participle selection with a primary focus on the linguistic factors governing speaker choice of regular versus irregular forms. Furthermore, we examine both dialectal and diachronic effects on these constraints, both of which have been overlooked in prior scholarship on this topic. Our study is inspired in part by Poplack and associates (e.g., Poplack & Dion, 2009; Poplack, Lealess, & Dion, 2013), whose recent work on French has shown that real usage of grammatical structures, or what they call "praxis," bears only a slight resemblance, if any, to the representation of such structures in grammars, that is, in "prescription." As Poplack, Lealess, and Dion (2013) pointed out, even grammars and linguistic analyses that purport to be "descriptive" in nature evince this fatal flaw, at least for variable grammatical phenomena. Poplack and Malvar (2007) have already shown this to hold true for one particular phenomenon in Brazilian Portuguese, the choice between synthetic and analytic (*ir* 'go' + infinitive) futures. While grammars still

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

promote the synthetic form as the standard and most frequently-used form, Poplack and Malvar (2007) showed that it has been replaced by the analytic form in both speech and non-formal written styles. In this paper, we show decisively that a similar conclusion can be drawn for the choice of past participle form in perfect constructions in Portuguese. Furthermore, we demonstrate that there is evidence of analogical change toward an even greater use of the "non-standard" irregular participle in such constructions.

PREVIOUS LITERATURE

In an attempt to better understand the diachronic trajectory of Portuguese past participles, Chagas de Souza (2007, 2011, 2015) traced the development of the thematic (regular) and athematic (irregular) participles from Latin to modern Portuguese. He claimed that a three-phase evolution has occurred, yielding the present forms. First, Portuguese inherited and maintained the doublets already found in Latin, such as *aceitado* - *aceito* ('accepted'). Second, this pattern was extended, via analogy, to other first conjugation verbs where the 1SG present indicative form was adopted as the irregular participle form (and as an adjective with copular *ser/estar*), as, for example, *chegado* - *chego* ('arrived'). Finally, innovative forms such as *pedido* - *peço* ('requested') have appeared recently with other verbs that normatively are not considered to be members of the class of *verbos abundantes*.[6] Simões (2008:356) also recognized, at least implicitly, the importance of the 1SG present indicative form to the formation of the irregular participle.

Nevins and Rodrigues (2014; see also Nevins, 2015) conducted several "wug" (cf., Berko Gleason, 1958) experiments to test how argument structure affected participle choice. In order to avoid "token frequency in the language or other ungodly factors," they invented Portuguese nonce verbs, as well as their corresponding nonce regular and irregular participles, and created contexts that would make each verb unambiguously unaccusative, transitive, or ditransitive (Nevins, 2015:193). Participants were presented with a discourse context and given a sentence with the option of inserting a regular or irregular participle of the nonce verb. The authors found that regular participles are greatly preferred overall, but the 1SG present indicative irregulars are used slightly more with unaccusatives as opposed to the other verb types. It is difficult to know, however, which specific properties of the verb the participants were reacting to in the experiment: the type of verb, as the authors intended, the form that the irregular took, or other factors such as unintended similarity to existing verbs/participles in the language. In the present paper, we therefore have chosen to confront the "ungodly factors" head-on with corpus data.

Powers (1985) investigated the variation between regular and irregular participles via a questionnaire study with 16 native Brazilian Portuguese speakers. With respect to perfect constructions specifically, he found that his respondents chose the irregular forms a mere 18% of the time. Crucially, however, Powers included morphologically irregular participles that do not have

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

a regular counterpart (e.g., *aberto, escrito*); only four of the irregular verbs in perfect constructions that respondents deemed acceptable (*eleito, entregue, pago, gasto*) had a regular counterpart found in the language. Powers concluded, much like we do here, that existing prescriptive and descriptive characterizations of participle choice are inaccurate and unreflective of actual usage, despite the fact that he did not investigate naturally-occurring data. Below, we show that the rate of irregular participles in naturally-occurring spoken/written Portuguese is much higher than reflected in the results of Powers' questionnaire study, and shows clear constraints on variation that it left unexplored.

Additionally, several previous variationist studies have examined naturally-occurring usage of these different participle forms and the factors governing their selection. Miara (2013) examined newspaper articles from the *Diário Catarinense* in Florianópolis in southern Brazil, extracting 843 total tokens. However, in her analysis, Miara included verbs showing no variation between regular and irregular participle forms. Once these invariant tokens are removed from her dataset, 566 tokens showing variation remain, coming from seven different verbs. From these verbs, Miara found a very high percentage (83%) of irregular participle use. Throughout her study, she suggested possible explanations for differences in usage from participle to participle, claiming that the alternation among different auxiliaries is gradient. Our results below confirm this view but with a larger dataset and more comprehensive multivariate analysis.

In another variationist study, Teixeira da Silva (2008) analyzed approximately 670 tokens from a corpus of spoken southern Brazilian Portuguese. Her results show that the regular participle is favored over the irregular form; however, only eight of these tokens, coming from only two verbs, displayed variation between regular and irregular participles with *ter/haver*. As a result, it is impossible to draw any firm conclusions from this study. Lastly, Mendes (2005, 2010) analyzed the overlap between perfect versus progressive forms from a variationist perspective. Looking at synchronic and diachronic use in two corpora of spoken Portuguese from São Paulo and Rio de Janeiro, he found that the perfect form (*ter* + participle) is becoming more restricted both linguistically and socially when compared to the progressive (*estar* + present participle). Mendes did not, however, look at variation between regular and irregular participles within the perfect construction.

The most recent study of variability between regular and irregular participles is Thomas (2018), who focused on European Portuguese. Analyzing the rate of irregular use in 10 different verbs in a corpus of written internet data, she found that seven favor the irregular form compared to three that favor the regular participle. More importantly, however, she also conducted a study of acceptability judgments and found that speakers heavily prefer the irregular forms over the regular forms for six *verbos abundantes*. Our results below suggest that this preference for the irregular forms may also play a role in the analogical spread of the irregulars, based specifically on the 1SG present indicative form.

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

RESEARCH QUESTIONS AND HYPOTHESES

Based on the findings of previous research on participles in Brazilian Portuguese, we formulated a number of research questions to guide our analysis. The overarching questions posed in this study are as follows: (1) What is the distribution of irregular/regular participles in constructions with *ter/haver* + past participle?, and (2) What are the factors that condition the variation between these two participle variants? In this light, we put to the test the normative rule—found in both prescriptive and descriptive grammars, including some authored by linguists—which claims that only regular participles ending in *-do* should occur in perfect constructions with the auxiliaries *ter* or *haver* (and the irregular form appearing exclusively in passive constructions and as adjectives).

We also hope to shed light on possible dialectal differences between Brazil and Portugal, as well as on diachronic differences to see whether there has been a shift toward more irregular participle use. We expect to see dialectal differences between Brazil and Portugal in light of the finding from Thomas' (2018) acceptability questionnaire that speakers of European Portuguese consistently preferred irregular over regular forms, as well as the prescriptive promotion of irregulars in school (Patrícia Amaral, personal communication). Additionally, we predict diachronic shift toward more irregular usage, especially when the irregular is identical to the 1SG present indicative form, given Chagas de Souza's (2007, 2011, 2015) discussion of innovative irregular forms recently attested in Brazilian Portuguese, which found that these new forms are created via analogy with the morphological basis of the 1SG present indicative, even when the indicative forms themselves are irregular (see also Thomas, 1969:223). In view of the prior linguistic accounts summarized in the previous section, the results of our quantitative analysis of the data lead us to side with Chagas de Souza's qualitative arguments that analogy with the 1SG present indicative best explains the variation. We conclude more generally that probabilistic constraints are operating on analogical processes in grammar in both synchrony and diachrony (cf., Albright, 2009).

METHODS

To investigate the variability between past participle forms, we utilized Davies and Ferreira's (2006) *Corpus do Português* (www.corpusdoportugues.org). The *Corpus do Português* is an online corpus consisting of 45 million words from almost 57,000 texts from the Portuguese-speaking world, ranging from the 1300s to the 1900s and containing sources from different genres or registers (spoken, fiction, newspaper, academic) and dialects (European and Brazilian Portuguese). The envelope of variation for the study was determined inductively (cf., Poplack, Lealess, & Dion [2013] who used a similar method to investigate the French subjunctive). We first compiled a master list of *verbos abundantes*, using multiple existing lists taken from Portuguese grammars, textbooks, and

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

websites. Among these different sources, there was widespread disagreement about which participles should be included in such a list.[7] In the end, we conducted an exhaustive search for every verb that showed variation between regular/irregular participles when occurring with the auxiliaries *ter* and *haver* in twentieth century data (Portugal and Brazil). Additionally, we searched for the same verbs in the nineteenth century for purposes of diachronic comparison within the data from Portugal (quite unfortunately, the Brazilian data in the corpus is limited to the twentieth century).[8] These verbs were then searched for using both auxiliary verbs (*ter/haver*) and both forms of the participle (e.g., *salvado, salvo* 'saved'), and tokens were extracted from the corpus for coding and analysis. Examples (1–6) below show the use of both regular and irregular participles (CdP = *Corpus do Português*; year provided where available in the CdP).

(1) *Era Alcoforado quem me tinha **salvado** a vida.*
'It was Alcoforado who had **saved** my life.'
(CdP, Title: *Leonor de Mendonça*, Author: Gonçalves Dias)

(2) *Esfreguei por fim os olhos e verifiquei que o Sanches me tinha **salvo**.*
'I finally rubbed my eyes and saw that Sanches had **saved** me.'
(CdP, Title: *O Ateneu*, Author: Raul Pompéia)

(3) *Este homem deve ter **gastado** muito dinheiro.*
'This man must have **spent** a lot of money.'
(CdP, Title: *Fritzmac*, Author: Aluísio Azevedo)

(4) *Lá continua. Tem **gasto** tudo em remédios.*
'He's still there. He's **spent** everything on medicine.'
(CdP, Title: *A Criação do Mundo: O Terceiro Dia* (1948), Author: Miguel Torga)

(5) *Minutos depois, arrependeu-se, devia ter **entregado** a carta, sentiu remorsos, pensou em mandá-la à casa do médico.*
'Minutes later, he changed his mind, he should've **delivered** the letter, he felt remorse, he considered sending it to the doctor's house.'
(CdP, Title: *Quincas Borba*, Author: Machado de Assis)

(6) *Após ter **entregue** a carteira e mais algum objeto pessoal, o executivo tenta reagir e é baleado.*
'After having **handed** over the wallet and some other personal object, the executive tries to fight back and is shot.'
(CdP, Title: *O Silêncio da Chuva* (1996), Author: Luiz Alfredo Garcia-Roza)

Following the variationist method (Tagliamonte, 2006, 2011), the tokens were then extracted and coded for the following linguistic and extralinguistic predictors in order to examine the constraints on the variation in participle selection:

- Lexeme (lexical verb)
- Type of irregular participle (same as 1SG present indicative or different)
- Auxiliary verb (*ter*, *haver*)
- Person (1st, 2nd, 3rd)
- Number (singular, plural)
- Polarity (positive, negative)

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

- Dialect (Brazil, Portugal)
- Century (19^{th}, 20^{th})

We coded for lexical verb because preliminary observations showed appreciable variation regarding rates of irregular participle usage from one verb to another. Our consideration of type of irregular participle was motivated by Chagas de Souza's (2007, 2011, 2015) discussion of innovative forms in Brazil, which, together with Amaral's (personal communication) comments on irregulars in Portugal (cf., also Thomas, 2018), also highlighted the need to account for possible dialectal differences. The inclusion of the auxiliary verb was intended to capture possible diachronic and stylistic differences since *haver* seems to be increasingly relegated to formal/written language. Finally, given that participles appear in verb constructions that could presumably be affected by frequency, clause type, or the discourse function of the verb, we coded for the person and number of the auxiliary, as well as polarity.[9] Mixed-effects logistic regression was carried out using the *lme4* package (Bates, Maechler, Bolker, & Walker 2015) in R, with the predictor LEXEME included in the model as a random effect.[10]

OVERALL RESULTS: BRAZIL AND PORTUGAL

A total of 1077 tokens were included in the analysis, with all of them coming from verb lemmas showing variability with regard to participle form in the *Corpus do Português*. The initial results are contrary to the claim made in most grammars that the auxiliaries *ter* and *haver* combine with regular participles only. In fact, more than half of the examples in our data contain irregular participle forms in perfect constructions, as shown in Table 1.

TABLE 1. *Overall frequencies of Portuguese participle form*

	n	%
Irregular Participle	584	54%
Regular Participle	493	46%
Total	*1077*	*100%*

Furthermore, dialectal differences show that there is an even greater preference for irregular participles in Portugal than in Brazil, as illustrated in Table 2. The use of irregular participles in Portugal is 15% higher than that found in Brazil, a difference that is significant by a chi-square test.

TABLE 2. *Frequencies of participle form across dialects in the 20^{th} century data* (n = *711*)

	Brazil	Portugal
Irregular Participle	164 (53%)	274 (68%)
Regular Participle	145 (47%)	128 (32%)
Total	*309*	*402*

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

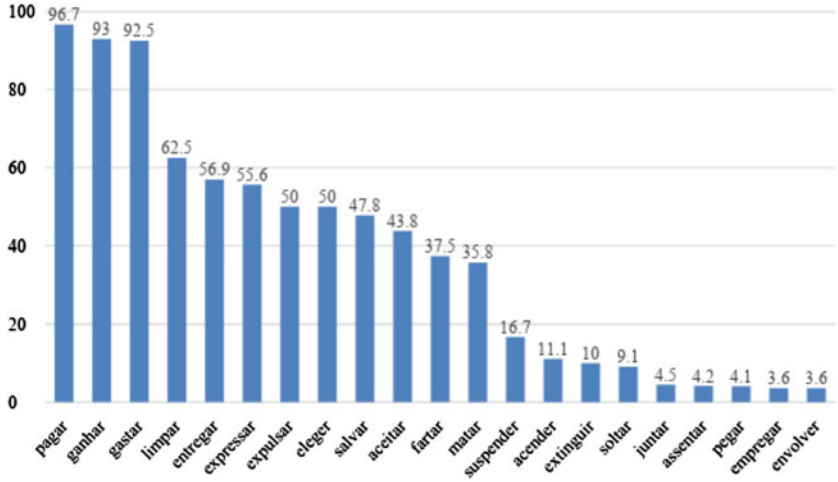


FIGURE 1. The 21 variable verbs in data set (listed by infinitive) with rates of irregular participle selection, from highest to lowest.

Figure 1 below shows the 21 verbs (of the grand total of 58 that we searched for, based on lists from grammars and language textbooks) that showed variation in participle form within the twentieth century data in the *Corpus do Português*.

As Figure 1 illustrates, there are enormous differences in the relative frequencies of irregular participle choice depending on the particular verb in question. Indeed, the rates of irregular participles span nearly the whole range of possibilities from a mere 4% for *empregar* 'to employ' and *envolver* 'to wrap' to a near-categorical 97% in the case of *pagar* 'to pay.' Perhaps most notable is that the three verbs with the highest percentage of irregular participles, *pagar, ganhar,* and *gastar*, display a combined rate of 94% irregular participles (377/401). Moreover, these three verbs actually provide 64% (377/584) of all of the irregular tokens in the full dataset. This means that the other 36% (207/584) of the total set of irregular tokens are distributed over 18 different verbs. But even in this latter set of tokens there is a decided lack of balance, inasmuch as three other verbs (*entregar, salvar, matar*) account for 57% (119/207) of the remaining irregular participles, and therefore 15 verbs account for the remaining 88 tokens. A grand total of six verbs (i.e., 28% of the 21 verb types showing variation), therefore, account for 85% (496/584) of all the irregular participles in the dataset. Such large disparities in the overall frequencies of verbs—not only in terms of irregular versus regular participle choice, but also of the widely varying frequencies with which they occur in perfect constructions—have gone essentially unnoticed in the prior literature on participle choice.[11] Another point to notice about the three verbs with the highest rate of irregular participle usage is that each has an irregular participle that is identical in form to the 1SG present indicative form of the same verb. This coincidence in form will prove important in the analysis below and for our argument that analogical processes are at work in regulating the variation.

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

In order to see beyond the strong dominating effect of the individual verb in the variation, we performed a mixed-effects regression analysis in R (Bates, 2010), including a random slope for LEXEME by DIALECT.[12] The best-fitting model of the twentieth century data (as determined by a log likelihood comparison of nested models), is presented in Table 3 below.

There is a significant difference between Brazil and Portugal, with Portuguese speakers/writers selecting significantly more irregular participles, as expected given the statistically significant difference between the two dialects reported in Table 2. No other fixed effect was significant. In order to more closely examine the hierarchical rankings of constraints on participle choice, as well as the interactions between those constraints, we generated a conditional inference tree (Figure 2). Here, we follow Tagliamonte and Baayen (2012:164), who argue that "conditional inference tree[s] offer an ideal tool to be used in tandem with the mixed-effects logistic model to come to a full understanding of the quantitative structure of a dataset and, as a result, an optimal interpretation of the variation."

As Figure 2 shows, the form of the irregular participle emerges as the top-ranked constraint: irregular participles are significantly more probable when they match the form of the 1SG present indicative form of the verb. Note that, in Figure 2, we include TYPE OF IRREGULAR PARTICIPLE rather than LEXEME for ease of presentation, since the former is binary whereas the latter consists of 21 different verbs. In the mixed-effects regression reported in Table 3, however, LEXEME was included as a random effect and therefore TYPE OF IRREGULAR PARTICIPLE could not be included without issues of collinearity.

TABLE 3. *Results from the mixed-effects logistic regression model (20th century data, both dialects) predicting irregular Portuguese participle forms* (n = *711). Also included: Auxiliary verb, Person, Number, Polarity*

Fixed effects:	Estimate	Std. Error	*z* value	*p*-value	*n*, % Irregular
(Intercept)	0.84	0.51	1.66	0.10	
DIALECT (Reference Level: Brazil, *n* = 309)					164, 53%
Portugal (*n* = 402)	1.12	0.50	2.25	0.02 *	274, 68%
Random effects:	**Variance**	**Std. Dev.**	**Corr.**		
Lexeme (Intercept)	3.68	1.92			
DialectPortugal	2.45	1.57	0.06		

On the right-hand first branch of the conditional inference tree, where the verb's irregular participle form is distinct from the 1SG present indicative, DIALECT emerges as a significant predictor, with significantly higher rates of regular participles in Brazil than in Portugal. Put another way, Portugal shows greater use of irregular participles even when the form of the irregular participle is not identical to that of the 1SG present indicative. Regarding the selection of regular or irregular participle forms based on the irregular participle form's shape

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

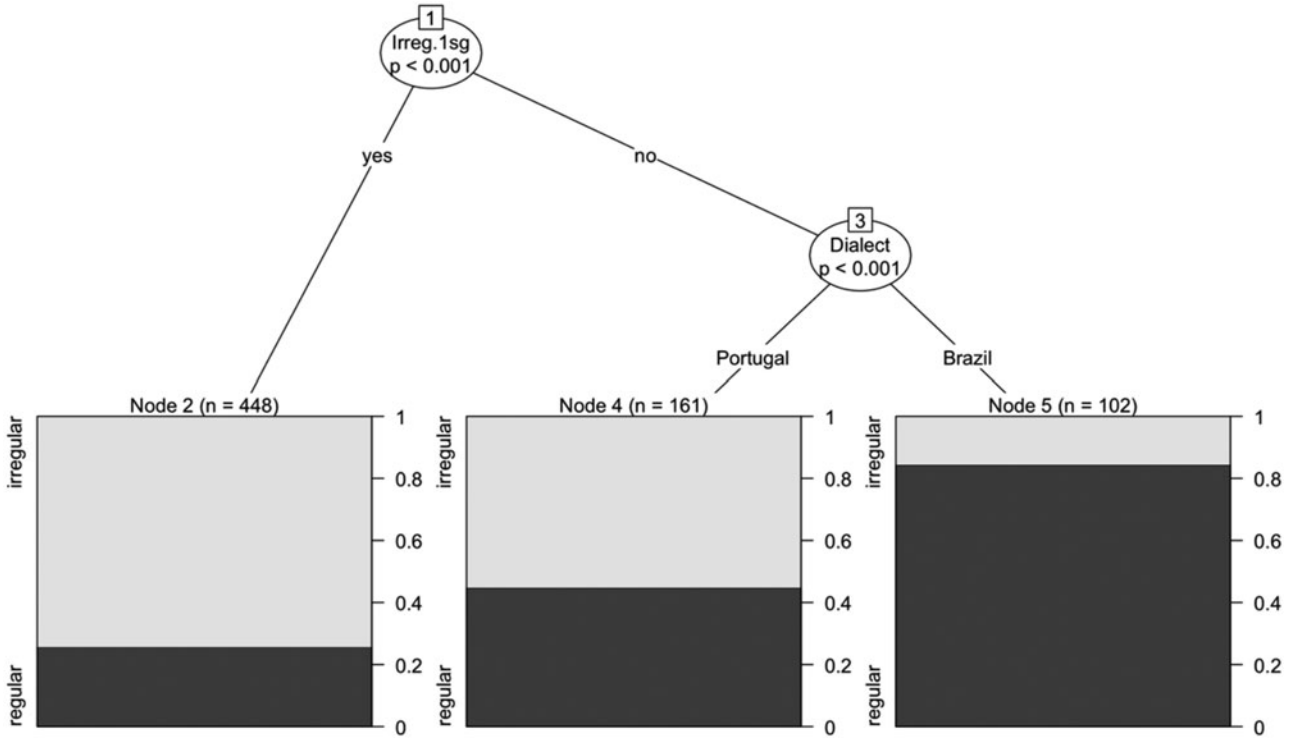


FIGURE 2. Conditional inference tree showing the constraint hierarchy for participle choice (20th century data) ($n = 711$).

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

(i.e., whether or not the participle coincides with the 1SG present indicative), the data are distributed as follows (Table 4):

TABLE 4. *Distribution of Portuguese irregular and regular participles based on the irregular participle's shape*

	Irregular SAME as 1st SG Present Indicative	Irregular DIFFERENT from 1st SG Present Indicative
Irregular Participle	444 (69%)	140 (32%)
Regular Participle	198 (31%)	295 (68%)
Total	*642*	*435*

These results reveal that, when the form of the irregular matches that of the 1SG present indicative, the distribution of irregular and regular participles is the exact opposite of when the irregular form's shape is of some other type. Thus, when the irregular form shares the shape of the 1SG present indicative, it has an overall usage rate of 69%, but when it is does not share that shape it only occurs 32% of the time. The actual morphological shape of the participle, then, exerts a strong influence on whether speakers/writers will choose the irregular form.

DIACHRONIC RESULTS: PORTUGAL

Turning to our diachronic analysis of European Portuguese across the nineteenth and twentieth centuries, we provide the results for the overall distribution of the regular and irregular participles in Table 5. Here we only examined the participles that occurred variably in the twentieth century data (see Figure 1) and compared them with their rates in the nineteenth century.[13]

As these numbers show, there is a clear and statistically significant shift toward higher rates (28% higher) of irregular participle use from the nineteenth to the twentieth century. We see a more fine-grained reflection of this change when we examine the individual verbs, as illustrated in Figure 3, which shows the seven most frequent variable verbs in the European Portuguese data.

As Figure 3 shows, some of the verbs remained essentially unchanged from the nineteenth to the twentieth centuries, as in, for example, *gastar* 'to spend' and *pagar* 'to pay,' both of which already displayed high rates of irregular participles in the nineteenth century. Other verbs, however, showed drastic increases in the overall usage of the irregular form (e.g., *aceitar* 'to accept' and *salvar* 'to save'). Just as impressively, **none** of the verbs showed an increase in the use of the regular participle across the two centuries. There was either an increase in the irregular and corresponding decrease in the regular form, or there was little to no change, as in the case of those verbs which already had high relative rates of the irregular participle in the nineteenth century. The overall trend, then, is clearly toward greater use of the irregular form no matter which verb is being employed.[14]

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

TABLE 5. *Distribution of irregular and regular participles across the 19th and 20th centuries (European Portuguese)*

	19th Century	20th Century
Irregular Participle	146 (39.9%)	274 (68.2%)
Regular Participle	220 (60.1%)	128 (31.8%)
Total	*366*	*402*

When we included these predictors in the fixed-effects model (see Table 6 in the appendix), we found that both CENTURY and TYPE OF IRREGULAR PARTICIPLE were significant predictors of choice between regular and irregular participles in the European Portuguese data.[15] The conditional inference tree (Figure 4) provides a visual representation of the hierarchical relationship among these independent variables.

Just as in the overall dataset above, when the irregular participle shares its form with the 1SG present indicative, there are significantly more irregular participles in the twentieth century than in the nineteenth. However, when the irregular participle shape is of some other type, there are significantly fewer irregular participles used in the nineteenth century.

Taken as a whole, the results presented in this section evince a clear change across the nineteenth and twentieth centuries in Portugal toward greater use of irregular participles in perfect constructions with *ter* and *haver*. Additionally, a comparison of those participles identical in form to the 1SG present indicative with those that are not shows that the increase in use of irregular participles was faster for the latter group; that is, irregulars identical to 1SG increased by only

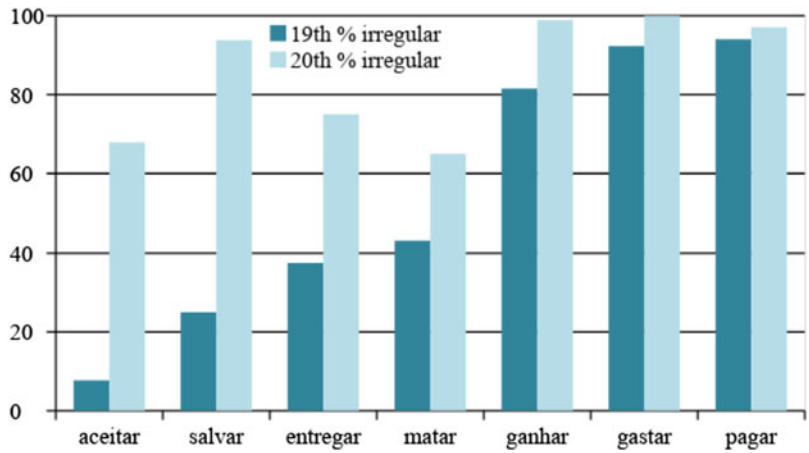


FIGURE 3. Rates of irregular participle usage in 19th and 20th centuries among the seven most frequent verbs.

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

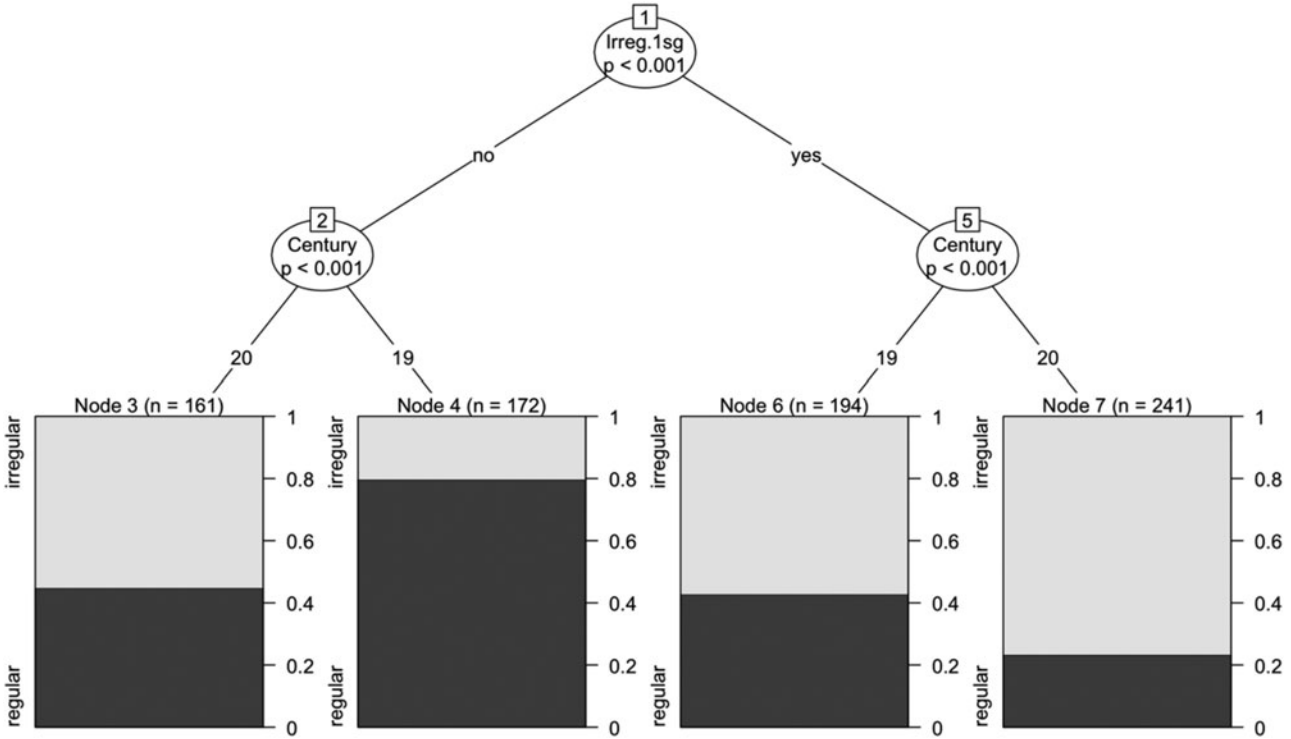


FIGURE 4. Conditional inference tree showing 19th and 20th century participle choice results for Portugal ($n = 768$).

about 20% (comparing the light gray areas in nodes 6 and 7 of Figure 4) from the nineteenth to twentieth centuries, while irregulars not identical to 1SG increased appreciably more (nodes 3 and 4). This could indicate that irregular participles began with forms identical to 1SG and then spread quickly to other verbs without this formal overlap, a typical pattern for actualization/analogical generalization (Malte Rosemeyer, personal communication). More detailed diachronic analysis would be necessary, however, to corroborate this process.

DISCUSSION

As demonstrated in the previous sections, descriptions of Portuguese participle use offered in grammars and usage guides of the language are highly inaccurate when confronted with naturally-occurring corpus data. The present study thus represents a contribution to the literature on Portuguese grammar by showing that normative and even descriptive claims are not borne out in actual usage, while, at the same time, providing the distributional details necessary to refine previous linguistic descriptions. The first example of this comes from the comparison of overall frequencies of regular and irregular participles. Though the majority of grammars postulate that the auxiliaries *ter* and *haver* require or at the very least strongly prefer regular participles, in reality these auxiliaries appear with irregular participles of *verbos abundantes* more than half of the time. This discrepancy between prescription and practice is even more accented in the data from Portugal, where irregular participle use with *ter* or *haver* is shown to reach nearly 70%. Though the majority of grammars relegate irregular participles to the domain of passives with *ser* or *estar*, the data presented here underscore how problematic such a generalization is in light of actual usage.[16]

In addition to responding to problematic elements in grammarians' descriptions, the present study also employed multivariate statistical analysis to uncover constraints on participle use not included in previous accounts. For example, the considerably distinct behavior of individual verbs (coded as LEXEME in our analysis), which was found here to be the most important predictor of the variation observed, goes unreported in nearly all of the grammars we consulted. No distinction is made between, say, *eleger* and *envolver* or *limpar* and *pegar*, when, in fact, these verbs have clearly differing preferences of participle choice.[17] Similarly absent from most accounts (with the exception of the work of Chagas de Souza [2007, 2011, 2015]) is any discussion of the critical importance of a participle's shape, that is, whether the irregular form is identical to that of the 1SG present indicative (e.g., *tinha ganho* 'I had won,' *eu ganho* 'I win') or not (*tinha extinto* 'I had extinguished,' *eu extingo* 'I extinguish'). As we have shown, the coincidence (via analogical process) between the 1SG form and irregular participles is the strongest factor leading to more selection of the irregular form.

The lexical differences between verbs, as well as the verb-type distinction just described, are especially interesting because of their implications for acquisition and, indirectly, diachrony. Indeed, the fact that individual verb tendencies are by

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

far the most important predictor overall of regular or irregular participle choice suggests that, during the acquisition process, children must learn participles on a verb-by-verb basis rather than as an overarching rule with several frequent exceptions. Given the decreasing frequency of the perfect construction overall, at least in Brazil (Mendes, 2005, 2010), this is also to be expected.

Regarding verb type, we emphasize again that the vast majority of irregular participles are identical to the 1SG present indicative form. While we have no evidence that the regular *-do* forms are dying out, and indeed novel verbs are overwhelmingly formed this way (cf., *tenho googleado* 'I've been googling'), our results show that participles have a competing analogical formation based on the 1SG form. This analogy is most strongly instantiated by *pagar, ganhar*, and *gastar*, the three highly frequent variable verbs with the highest rates of irregular participle use (all over 90% in Figure 1), whose irregulars are all identical to their 1SG form, thereby providing a robust model for less frequent verbs (cf. Bybee, 2010:57–75, 2015:93–116). It is therefore not surprising that when speakers create innovative participles in Brazilian Portuguese they nearly always base their analogical innovations on this same form (Chagas de Souza, 2007, 2011, 2015).[18] Examples (7–14), consisting of naturally-occurring tokens taken from Twitter users in Brazil, evidence the use of the innovative participles *chego* (normatively *chegado*), *falo* (*falado*), *passo* (*passado*), *peço* (*pedido*), *penso* (*pensado*), *perco* (*perdido*), and *trago* (*trazido*). These forms have already become the subject of much prescriptive commentary online. For example, the use of *chego* and *trago* as participles is labeled a common error on language usage-related websites such as Guia do Estudante, Brasil Escola, and Ciberdúvidas da Língua Portuguesa, among many others. Scher, Lunguinho, and Takahira (2013, 2014) also documented the existence of the innovative participles *falto* (*faltado*), *levo* (*levado*), *quebro* (*quebrado*), and *subo* (*subido*) with *ter/haver* as well as forms such as *compros* (*comprados*), *fechos* (*fechados*), and *mando* (*mandado*) in adjectival/passive constructions. All of these innovative forms are identical to the 1SG present indicative form of the same verb and are suggestive of a strong analogical process at work. In addition, the forms *peço* (11), *perco* (13), and *trago* (14) are all irregular in the 1SG present indicative and conserve this irregularity as participles. If speakers were not basing their innovations on the 1SG form, then there would be no reason why they would choose these irregular forms as the basis for innovative irregular past participles. Instead, they would presumably create unattested forms consisting of the verb stem plus *-o* (e.g., **pedo* and **trazo*).[19]

(7) *Ultimamente as coisas tem **chego** até mim muito fácil.*
'Lately things have come to me really easily.'[20]

(8) *Se eu não tivesse **falo** aquilo hoje, seria bem diferente.*
'If I hadn't said that today, it would be much different.'

(9) *Se eu soubesse teria **fico** em casa limpando o côco do cachorro.*
'If I knew, I would've stayed at home cleaning up the dog's poop.'

(10) *Me sinto como se um caminhão tivesse **passo** por cima de mim.*
'I feel as if a truck had run over me.'

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

(11) *Cheguei em casa e meus pais tinham* ***peço*** *hamburger pra mim.*
'I got home and my parents had ordered me hamburgers.'
(12) *Às vezes eu tenho* ***penso*** *em fazer faculdade de astrologia tb.*
'Sometimes I've thought about going to school for astrology, too.'
(13) *Sonhei que tinha* ***perco*** *a hora pro teste de matemática e acordei surtando.*
'I dreamt that I'd missed my math test and I woke up freaking out.'
(14) *Se eu soubesse, teria* ***trago*** *meu caderno pra disenhar.*
'If I'd known, I would've brought my notebook to draw.'

It is possible, in light of this trend in Brazil, that this same analogical pattern was responsible for the diachronic increase in irregular use in Portugal as well, though empirical confirmation of this claim would require diachronic data from colloquial speech that is currently unavailable to us. Another possible explanation for the notable increase (over 25%) in irregular participle use in Portugal from the nineteenth to twentieth centuries is the normative view espoused in the school system that irregular participles are preferable to their regular counterparts (Thomas, 2018; Patrícia Amaral, personal communication). Indeed, the perceived prestige of irregulars could work in tandem with the preference for participles to match the 1SG present indicative form, thus explaining the notable increase in irregular participle selection from the nineteenth to the twentieth century.

In addition to the linguistic factors just described, the data presented here highlight the importance of dialectal variation to participle choice. No grammars or prior studies have discussed differences in regular versus irregular participle use between Brazilian and European Portuguese (aside from passing qualitative mentions that forms such as *aceite* 'accepted' exist in Portugal but not Brazil, which uses *aceito*). Still, we identified clearly different distributional behaviors between these dialects. The Brazilian data show higher rates of selection of regular forms while Portugal exhibits significantly greater use of the irregulars.

CONCLUSION

The results of the quantitative analysis of participle alternation have provided empirical evidence for the strong effect of analogy in the choice of the irregular participle forms in both Brazilian and European Portuguese. By operationalizing the analogical target as the participle form that is equivalent to the 1SG present indicative form, we have been able to demonstrate statistically significant correlations between the perfect participles and these targets, thereby corroborating in some cases (Chagas de Souza, 2007, 2011, 2015) and refuting in others (Scher, Lunguinho, & Takahira, 2013; Scher, 2014) prior qualitative analyses of the structural motivations for participle form choice. Arguments that innovative irregular participles do not always equate in form to the 1SG (e.g., Scher, Lunguinho, & Takahira, 2013; Scher, 2014), while perhaps interesting from a qualitative point of view, are actually exceptions to the robust patterns found in the data analyzed here. Although these exceptions are interesting in their own right and deserve further research, they should not overshadow the probabilistic

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

grammar of participle choice revealed in our analysis, which is clearly driven by the 1SG present indicative target. The point of greatest theoretical interest here is that we have been able to corroborate the hypothesis of analogical matching using a large dataset and shown how this particular factor interacts with others, but at the same time dominates them. Moreover, this dominant effect of analogy spans both European and Brazilian varieties of Portuguese. The effect of analogy also applies to innovative types in synchrony, such as those found on Twitter, and to the diachronic increase of the irregulars in the data from the nineteenth to twentieth centuries, thereby solidifying the pattern based on the 1SG present indicative form to more and more instances of verbs with *particípios duplos*, and the subsequent reduction of the purported regular pattern with *-do* (cf., Bagno & Casseb-Galvão, 2016). Whether this spread of the 1SG pattern will continue, and result in a reversal of markedness such that *-do* becomes the irregular formation, is a possibility that will only be determined with time.

The present study makes it clear that neither modern grammars (whether prescriptive or descriptive) nor linguists have fully or accurately described how speakers alternate between the regular and irregular past participles in Portuguese. Our quantitative analysis reveals that dialect is an important factor determining participle choice across varieties of Portuguese (cf., Thomas, 2018). European Portuguese uses significantly more irregular participles than does Brazilian Portuguese, and these results conform to EP speakers' positive evaluations of irregular forms (Thomas, 2018). Moreover, the linguistic divide between the participles does not boil down to auxiliary verb choice or sentence/construction type, as most grammars and linguists claim, but is instead determined by linguistic constraints that could not have been uncovered without the statistical analysis performed in this paper. Given how much of the alternation between participles occurs on a verb-by-verb basis, and the degree to which it is sensitive to participle form and its analogical similarity or lack thereof to the 1SG present indicative, it is also clear that choice of past participle in Portuguese is not a strictly grammatical phenomenon, but rather one where both lexical and grammatical constraints interact to create the patterns of variability uncovered in our study.

As noted by Poplack and Dion (2009) and Poplack, Lealess, and Dion (2013) for different phenomena in French, there is more generally a clear disconnect between "prescription" by grammars (even those purporting to be "descriptive" in nature) and "praxis" or usage by speakers/writers. Grammars, grammar writers, and even many linguists appear to do their best to follow what Poplack and Dion (2009:557) have called the "doctrine of form-function symmetry," whereby differences between distinct forms are purported to correspond to differences in function/meaning. The examination and analysis of Portuguese past participle data in this paper has shown that the same disconnect holds for this language, where mention is rarely if ever made of lexical differences or of the role of the form of the irregular participle in the variability. Even recent linguistic research on participles has focused more on the creation of innovative irregular participles from verbs that are not considered to form part of the group of *verbos abundantes* and ignored the widespread real-life usage of (established) irregular participles in

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

perfect constructions that runs contrary to prescriptive dictates (cf., Chagas de Souza, 2007, 2011, 2015; Scher, Lunguinho, & Takahira 2013, Scher, 2014). The prescription that allocates regular and irregular participles to distinct grammatical structures is one that has little basis in the empirical reality of how Portuguese is actually used, just as Poplack and Malvar (2007) already found for the variation between competing future forms in Brazilian Portuguese. We believe that modern-day descriptions of Portuguese should be updated in order to reflect this reality more accurately, and that we linguists should play a role in promoting this correction.

NOTES

1. Luana Lamberti (personal communication) reports that these same "rules" are taught explicitly in Portuguese language classes in the Brazilian educational system.

2. Original: "O particípio regular somente se usa com o verbo *ter*; e o irregular, ou com valor adjetivo isolado, ou com o verbo *ser*."

3. Original: "Em geral emprega-se a forma regular, que fica invariável com os auxiliares *ter* e *haver*, na voz ativa, e a forma irregular … com os auxiliares *ser, estar* e *ficar*, na voz passiva."

4. Neither of these authors mentioned the state of this variation in European Portuguese, since their textbooks (Simões, 2008) or grammars (Perini, 2002) are restricted to Brazilian Portuguese. We were unable to find any mention of dialect differences between Portugal and Brazil regarding this phenomenon.

5. Original: "Trata-se … de uma tendência, com muita variação."

6. While none of these innovative participles occurred in the data we used for this study from the *Corpus do Português*, we provide naturally occurring examples from Twitter in the discussion section.

7. See Said Ali (2008:149–162) for a detailed discussion of the tremendous historical disagreement among grammars in the verbs they list as having *particípios duplos*, as well as discrepancies between the prescription of grammars and the use of participles by respected authors.

8. However, a few authors from the nineteenth century, such as Gonçalves Dias (1823–1864) and Machado de Assis (1839–1908) are included in the twentieth century portion of the corpus. Rather than confirming the correct period for each token, we elected to trust the classification provided in the corpus.

9. Although we coded for TAM, the results obtained were unclear. Therefore, we leave further exploration of this factor to future research. Genre (written, oral) was also coded for. However, due to the nature of the corpus, only 80 (27 regular [34%], 53 irregular [66%]) of the total 1077 tokens came from the oral genre. Therefore, we do not discuss this factor further, though sources that approximate spoken language such as Twitter indicate that irregular participles are more frequent in speech than in writing.

10. As to including speaker/writer as a random effect (cf., Baayen, Davidson, & Bates [2008] and Johnson [2009] for discussion), in the *Corpus do Português* identifying information for speakers/writers is often unavailable and/or inaccurate, unfortunately. While it would have been prudent to include document as a random effect, there were few documents with more than one token of the construction in question, thereby making this option impossible.

11. Thomas (2018) identified a pattern, albeit imperfect, whereby irregular participles in European Portuguese are more frequent with *-ar* than with *-er/-ir* verbs. However, such a pattern is not clearly borne out in our data, as shown in Figure 1, where individual *-ar* verbs take irregulars at some of the highest and also the lowest rates.

12. We followed Tagliamonte and Baayen's (2012) recommendation of using random forests in conjunction with mixed-effects regression models for unevenly distributed corpus data. A random forest of the twentieth century data ($n = 711$), with both Brazil and Portugal included, indicated that LEXEME and DIALECT were the most important factors predicting participle choice, with LEXEME being the strongest predictor. See James, Witten, Hastie, and Tibshirani (2013) for a more complete discussion of the inner workings of random forests.

13. We are grateful to an anonymous referee for pointing out that consideration of verbs showing variation in the nineteenth, but not in the twentieth century, would be a fruitful area for future research.

14. In order to see the linguistic conditioning behind these raw frequencies, we again employed a random forest, together with the step function, to determine the best predictors for inclusion in the regression model. This random forest showed that LEXEME and CENTURY were both strong candidates for inclusion in the model.

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

15. The model in Table 6 in the appendix contains only fixed effects because inclusion of a random intercept for LEXEME would not allow for the examination of TYPE OF IRREGULAR PARTICIPLE in the same model.
16. Indeed, learners of Portuguese who are taught to follow the rule defended in so many grammars would actually turn out to be "wrong" more often than "right" in terms of native-like participle selection if they followed this generalization.
17. However, Said Ali (1964) does highlight differences in the historical participial behavior of individual verbs, and Villalva and Almeida's (2004) experimental findings for European Portuguese show great variability by verb. Móia (2004) proposes four subgroups of verbs in EP based on the frequency of the irregular form, using corpus data, but does not offer any explanation for these groupings.
18. Scher, Lunguinho, and Takahira (2013, 2014) provide three examples of innovative participles in passive constructions from European Portuguese: *foram… compras* (*compradas*), *cortas* (*cortadas*), *falas* (*faladas*).
19. The innovative participle *perdo* (instead of regular *perdido*) does occur on Twitter, though seemingly less frequently than *perco*. But note as well that Brazilians often regularize the 1SG present indicative of *perder* to *perdo*, so it is not clear that this is a case where speakers/writers are not using the 1SG form as a target of the analogy.
20. From our Twitter searches for *chego* as a participle, we have found that these innovative irregular participles are also variably number marked, as in *Dormiria mais se meus pais não tivessem chegos* 'I would've slept more if my parents hadn't arrived' or *Os livros não tinham chegos* 'The books hadn't arrived.'

REFERENCES

Albright, Adam. (2009). Modeling analogy as probabilistic grammar. In J. P. Blevins & J. Blevins (eds.), *Analogy in grammar: form and acquisition*, 185–213. Oxford: Oxford University Press.
Baayen, Harald, Davidson, Doug, & Bates, Douglas. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59:390–412.
Bagno, Marcos & Casseb-Galvão, Vânia Cristina. (2016). Hipercorreção e analogia: o caso dos particípios passados. In M. Rios de Oliveira & M. M. Cezario (eds.), *Funcionalismo linguístico: diálogos e vertentes*, 141–156. Niterói: Eduff.
Bates, Douglas. (2010). lme4: Mixed-effects modeling with R. Unpublished manuscript.
Bates, Douglas, Maechler, Martin, Bolker, Ben, & Walker, Steve. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67,1:1–48.
Bechara, Evanildo. (2006). *Moderna gramática portuguesa, 37ª edição*. Rio de Janeiro: Editora Lucerna.
Berko Gleason, Jean. (1958). The child's learning of English morphology. *Word* 14,2–3:150–177.
Bybee, Joan. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
_____. (2015). *Language change*. Cambridge: Cambridge University Press.
Carvalho, Orlene Lúcia & Bagno, Marcos. (2015). *Gramática brasileña para hablantes de español*. São Paulo: Parábola Editorial.
Chagas de Souza, Paulo. (2007). Athematic participles in Brazilian Portuguese: a syncretism in the making. *Acta Linguistica Hungarica* 54:119–128.
_____. (2011). Particípios atemáticos no Português Brasileiro - um processo paradigmático. *ReVEL*, special edition 5.
_____. (2015). Particípios atemáticos do Português Brasileiro: paradigmas, sincretismo e diacronia. In I. M. Alves & A. M. Ribeiro de Jesus (eds.), *Os estudos lexicais em diferentes perspectivas*, vol. 5, 29–58. São Paulo: Universidade de São Paulo.
Cunha, Celso & Cintra, Lindley. (2001). *Nova gramática do português contemporâneo, terceira edição*. Rio de Janeiro: Nova Fronteira.
Davies, Mark & Ferreira, Michael. (2006). Corpus do português: 45 million words, 1300s-1900s. http://www.corpusdoportugues.org.
Ganho, Ana Sofia & McGovern, Timothy. (2004). *Using Portuguese: A guide to contemporary usage*. Cambridge, UK: Cambridge University Press.
James, Gareth, Witten, Daniela, Hastie, Trevor, & Tibshirani, Robert. 2013. *An introduction to statistical learning with applications in R*. New York: Springer.
Johnson, Daniel Ezra. (2009). Getting off the GoldVarb Standard: introducing Rbrul for mixed-effect variable rule analysis. *Language and Linguistics Compass* 3:359–383.
Lima, Carlos Henrique da Rocha. (1972). *Gramática normativa da língua portuguesa*. Rio de Janeiro: José Olympio.

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

Lobato, Lucia. (1999). Sobre a forma do particípio português e o estatuto dos traços formais. *DELTA* 15. Online at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44501999000100005&lng=en&nrm=iso&tlng=pt#12not.
Mateus, Maria Helena Mira. (2003). *Gramática da língua portuguesa*. Lisbon: Editorial Caminho.
Mendes, Ronald Beline. (2005). *Estar + gerúndio e Ter + particípio: Aspecto verbal e variação em português*. PhD dissertation, UNICAMP.
_____. (2010). Aspectual periphrases and syntactic variation in Brazilian Portuguese. In J. A. Walker (ed.), *Aspect in grammatical variation*. Amsterdam: John Benjamins Publishing Company. 27–47.
Miara, Fernanda Lima Jardim. (2013). *Particípios duplos: Usos, desusos e alguns "intrusos."* MA thesis, Universidade Federal de Santa Catarina.
Móia, Telmo. 2004. Algumas áreas problemáticas para a normalização linguística–disparidades entre o uso e os instrumentos de normalização. *Actos do XX Encontro Nacional da Associação Portuguesa de Linguística*. Lisbon: APL. 109–125.
Nevins, Andrew. (2015). Productivity and Portuguese morphology: How experiments enable hypothesis-testing. In E. O. Aboh, J. C. Schaeffer, & P. Sleeman (eds.), *Romance Languages and Linguistic Theory 2013*. Amsterdam: John Benjamins Publishing Company. 175–201.
Nevins, Andrew & Rodrigues, Cilene. (2014). The L-syntax of athematic past participles: Evidence from wug experiments. *Revista da ABRALIN* 13:225–262.
Perini, Mário. (2002). *Modern Portuguese: a reference grammar*. New Haven, CT: Yale University Press.
Perini, Mário. (2010). *Gramática do português brasileiro*. São Paulo: Parábola Editorial.
Poplack, Shana & Dion, Nathalie. (2009). Prescription vs. praxis: the evolution of future temporal reference in French. *Language* 85:557–587.
Poplack, Shana, Lealess, Allison, & Dion, Nathalie. (2013). The evolving grammar of the French subjunctive. *Probus* 25:139–196.
Poplack, Shana & Malvar, Elisabete. (2007). Elucidating the transition period in language change: The expression of the future in Brazilian Portuguese. *Probus* 19:121–169.
Powers, Michael D. (1985). Variação dos particípios duplos no português e as suas implicações pedagógicas. *Hispania* 68:412–417.
Prista, Alexander da R. (1966). *Essential Portuguese grammar*. Mineola, NY: Dover Publications.
Raposo, Eduardo Buzaglo Paiva, Bacelar do Nascimento, Maria Fernanda, Coelho da Mota, Maria Antónia, Segura, Luísa, & Mendes, Amália. (2013). *Gramática do português I, volume II*. Lisbon: Fundação Calouste Gulbenkian.
Said Ali, M. (1964). *Gramática Histórica da Lingua Portuguesa, 5ª edição melhorada e aumentada*. São Paulo: Edições Melhoramento.
_____ (2008). *Dificuldades da Lingua Portuguesa: Estudos e observações, 7ª edição*. Rio de Janiero: Biblioteca Nacional.
Scher, Ana Paula, Lunguinho, Marcus Vinicius, & Takahira, Aline Garcia Rodero. (2013). Innovative participles in Brazilian Portuguese. *Revista LinguíStica* 9,1:92–111.
_____. (2014). Voice (a)symmetries and innovative participles in Brazilian Portuguese. *Cadernos de Estudos Lingüísticos* 56,1:45–64.
Simões, Antônio Roberto Monteiro. (2008). *Pois não: Brazilian Portuguese Course for Spanish Speakers with Basic Reference Grammar*. Austin, TX: University of Texas Press.
Tagliamonte, Sali. (2006). *Analysing sociolinguistic variation*. Cambridge, UK: Cambridge University Press.
_____. (2011). *Variationist sociolinguistics: change, interpretation, observation*. Oxford: Basil Blackwell.
Tagliamonte, Sali, & Baayen, Harald. (2012). Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24:135–178.
Teixeira da Silva, Inaciane. (2008). O uso do particípio em formações verbais no português do sul do Brasil. MA thesis, Universidade Federal do Rio Grande do Sul.
Thomas, Amanda. (2018). Accounting for double past participle forms in Romance. MA thesis, University of Oxford.
Thomas, Earl. (1969). *The syntax of spoken Brazilian Portuguese*. Nashville, TN: Vanderbilt University Press.
Villalva, Alina & Almeida, Marta. (2004). Verbos abundantes: usos, desusos e alguns 'abusos.' Paper presented at the XX Encontro da Associação Portuguesa de Lingüística, Lisbon.

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press

APPENDIX

*Results from the fixed-effects logistic regression model (Portugal data) predicting irregular participle forms (*n = *768). Also included: Auxiliary verb, Person, Number, Polarity*

	Estimate	Std. Error	*z* value	*p*-value	*n*, % Irregular
(Intercept)	−1.15	0.15	−7.74	1.01e-14 ***	
TYPE OF IRREGULAR PARTICIPLE (Reference Level: Not identical to 1SG, *n* = 333)					124, 37%
Identical to 1SG (*n* = 435)	1.30	0.16	8.12	4.52e-16 ***	296, 68%
CENTURY (Reference Level: 19th, *n* = 366)					146, 40%
20th (*n* = 402)	1.20	0.16	7.51	6.15e-14 ***	274, 68%

https://doi.org/10.1017/S0954394519000036 Published online by Cambridge University Press