

# Register effects and the Spanish adjectival construction *sin* + INF in historical corpus data

Aaron Yamada  
Creighton University  
[aaronyamada@creighton.edu](mailto:aaronyamada@creighton.edu)



Received: 10-06-21  
Accepted: 29-09-21  
Published: 21-01-22

**How to cite:** Yamada, Aaron. 2021. Register effects and the Spanish adjectival construction *sin* + INF in historical corpus data. *Isogloss. Open Journal of Romance Linguistics* 8(1) /1, 1-29.  
DOI: <https://doi.org/10.5565/rev/isogloss.147>

## Abstract

---

This study analyzes the usage of the Spanish adjectival *sin* + infinitive verb construction (*un libro sin terminar* ‘an unfinished book’) and the *no* + past participle construction (*un libro no terminado* ‘an unfinished book’) in historical corpus data, with the objective of quantitatively assessing Pountain’s (1993) analysis of the emergence of adjectival *sin* + INF as motivated in part by register formality. A logistic regression analysis finds that the usage of adjectival *sin* + INF over *no* + PP is significantly favored by the text register of Prose Fiction, and significantly disfavored by the register of Legal Texts. Furthermore, this preference increases with time in the register of Prose Fiction. These findings support Pountain’s claim, in showing that register effects significantly influence the usage of the novel construction. Ultimately, this study stresses the importance of measuring register effects in the analysis of language change in corpus data.

**Keywords:** Spanish diachrony, adjectival *sin* + INF, register variation, historical syntax

---

## 1. Introduction

Negation continues to generate interest in linguistic research, especially with respect to its diachrony and the role it plays in discussions of language change (see Horn 1989, Hansen and Visconti 2014, Van der Auwera 2010). In this paper, we measure the diachronic usage of two analytic negative adjectival constructions in Spanish: *sin* ‘without’ + infinitive verb (*sin* + INF) and *no* ‘no’ + past participle (*no* + PP). Specifically, we use the *Corpus diacrónico del español* (CORDE) to test the proposal of Pountain (1993), who argued that the emergence of adjectival *sin* + INF was motivated by a predisposition for *no* + PP to appear in more formal linguistic registers. In total, this paper takes a quantitative approach to the study of historical linguistics, in testing claims regarding language change using and measuring evidence found in corpus data (cf. Jensen and McGillivray 2017).

## 2. Background

### 2.1. The adjectival *sin* + INF construction

In Spanish, the preposition *sin* can be placed before an infinitive verb to create a prepositional phrase denoting the negation of a particular action, as in (1a). Casalicchio (2019) describes this *sin* + INF construction as a prepositional infinitive that can provide the negation of a gerund. Additionally, the *sin* + INF construction can be used to describe a noun that is interpreted as the direct object of the infinitive verb, as shown in (1b) and (1c).

- (1) a. Juan salió sin comer.  
‘Juan left without eating.’
- b. Una casa sin pintar  
‘An unpainted house’
- c. La casa está sin pintar.  
‘The house is unpainted.’

As shown in (1b), the placement of the infinitive verb *pintar* ‘to paint’ after the preposition *sin* indicates its negation. As seen in (1c), it is also possible to find this construction as a complement to a verb, such as the copula *estar* ‘to be’, where it plays the same role. In the three examples, there is an observable coreference relationship that links the *sin* + INF construction and a particular noun. Importantly, though, (1a) differs from (1b) and (1c) because in (1a) we do not interpret the noun *Juan* as being the direct object of the verb *comer* ‘to eat’, whereas in (1b) and (1c) we interpret the noun *casa* ‘house’ as being the direct object of the infinitive verb. In cases such as (1b) and (1c), we shall refer to the *sin* + INF construction as an adjectival *sin* + INF construction, following analyses of this construction as a kind of

adjective. Indeed, as pointed out in the grammar of the *Real Academia Española* (RAE 2009), it can alternate with the *no* + PP construction, as shown in (2).<sup>1</sup>

- (2) a. un libro *sin* terminar  
       ‘an unfinished book’
- b. un libro *no* terminado  
       ‘an unfinished book’

Example (2b) illustrates how the Spanish negator *no* can be placed in front of a past participle to negate the past participle and describe a noun. In both cases (2a) and (2b), the negative constructions act as adjectives that indicate that a given action has not taken place. Thus, rather than considering these cases of *sin* + INF to be prepositional phrases, Bosque (1990) analyzes them as a kind of prefixed adjective, where *sin* plays the role of a prefix indicating the negation of a participle, and where the noun *libro* ‘book’ acts as the internal argument of the verb *terminar* ‘to finish’. In this sense, Bosque (1990) notes that an important feature of categorizing the adjectival *sin* + INF construction as an adjective is that it helps to explain the appearance of these transitive verbs without any kind of direct object in their traditional positions. Among Bosque’s arguments that *sin* is best analyzed as a kind of prefix in this construction, we find that (a) *sin* already exists as a prefix in lexical items such as *sinvergüenza* ‘scoundrel’ or *sinrazón* ‘injustice’, (b) in the adjectival *sin* + INF construction intervening material is not found between *sin* and the infinitive verb that follows as it, as in *\*sin siquiera secar* ‘not even dried’, and (c) the infinitive verb does not take complements, as in *\*La botella está aún sin llenar de vino* ‘The bottle is still unfilled with wine’.

Bosque (1990) also argues that the infinitive verbs in the adjectival *sin* + INF construction must indicate a change of state. In other words, while the examples of *sin* + INF in (3a) below contain change of state verbs and are therefore well-formed, those in (3b) do not contain such verbs and thus are rejected by Bosque.

- (3) a. Un camión *sin* llenar                   ‘An unfilled truck’  
       Ropa *sin* secar                       ‘Undried clothes’  
       Uvas *sin* madurar                   ‘Unripe grapes’  
       Cuerdas *sin* tensar                   ‘Untaut ropes’
- b. \*Gente *sin* respetar               ‘Unrespected people’  
       \*Acciones bancarias *sin* tener   ‘Unheld bank stocks’  
       \*Un cuadro *sin* mirar             ‘An unseen portrait’  
       \*Un coche nuevo *sin* ponderar   ‘An unpraised new car’

(from Bosque 1990: 197)<sup>2</sup>

<sup>1</sup> An anonymous reviewer points out that not all *sin* + INF and *no* + PP alternates seem to be acceptable: *un camión sin llenar* vs. ? *un camión no llenado*, for example. Thus, we refer to *no* + PP as a possible alternative to adjectival *sin* + INF, rather than a concrete equivalent.

<sup>2</sup> As an anonymous reviewer points out, certain well-defined contexts may indeed render these phrases acceptable to some speakers.



Pountain's (1993) paper is the only to offer a proposal with respect to the origin of the adjectival use of *sin* + INF. To explain the emergence of this construction, Pountain points to the historical development of the range of possible complements of the verb *estar* 'to be'. Pountain argues that chronologically, *estar* accepted prepositional phrases before later accepting past participles. Thus, as *estar* accepted prepositional phrases, it would have formed constructions with the preposition *sin*, and constructions with the active value of *sin* + INF. Then, as *estar* began to take past participles as complements, it would have been inclined to produce the adjectival use of *sin* + INF, as adjectival *sin* + INF is semantically equivalent to a negated past participle. Once adjectival *sin* + INF came into use following the verb *estar*, a natural next step would be the elision of the verb *estar* to form noun + *sin* + INF constructions such as those in example (1b). In short, we schematize Pountain's claim regarding the development of adjectival *sin* + INF with examples in (6) below, chronologically progressing from top to bottom:

- (6)
- |  |                              |
|--|------------------------------|
| <i>estar</i> + prepositional phrase                | <i>estar sin sospecha</i>    |
| > <i>estar</i> + <i>sin</i> + INF (active use)     | <i>estar sin comer</i>       |
| > <i>estar</i> + past participle                   | <i>estar terminado</i>       |
| > <i>estar</i> + <i>sin</i> + INF (adjectival use) | <i>estar sin terminar</i>    |
| > noun + <i>sin</i> + INF (adjectival use)         | <i>un libro sin terminar</i> |

An important question that Pountain aims to answer is why adjectival *sin* + INF emerged when other constructions that expressed adjectival negation already existed. Thus, they compare adjectival *sin* + INF to other negative adjectival constructions, which include the prefixes *des-* (*desfavorable* 'unfavorable') and *in-* (*incontaminado* 'uncontaminated'), and the analytic constructions *poco* + adjective (*poco seguro* 'not very safe') and *no* + adjective (*no terminado* 'unfinished'). Pountain claims that the use of *sin* + INF as an adjectival construction provided a solution to several issues associated with the alternatives mentioned above. First, with respect to the synthetic forms, they argue that the *in-* prefix declined in productivity and did not frequently attach to past participles, instead preferring adjective formation with suffixes such as *-able* or *-ible*. Furthermore, while the *des-* prefix is relatively productive, it can convey a privative or reversative meaning rather than being strictly negative. As far as the analytic constructions are concerned, Pountain notes that the *poco* construction does not always convey only negation when combined with past participles, in that *poco* at its base is a degree adverb and, rather, indicates the negation of *muy* 'very'. Finally, Pountain argues that the *no* + adjective construction is most utilized in registers of high formality. Pountain writes that "In neutral register, *Traiga una ensalada sin aderezar* would be preferred to *Traiga una ensalada no aderezada* 'Bring a salad without dressing'." (Pountain 1993: 112) An additional syntactic restraint is that it is unacceptable for *no* + adjective constructions to follow the verb *estar*, as shown in (7).

- (7) \*El libro está no terminado  
'The book is unfinished'

Because of these limitations, Pountain argues that adjectival *sin* + INF may have developed to satisfy a need not met by the alternative constructions.<sup>4</sup> In summary, the main points regarding the emergence of adjectival *sin* + INF found in Pountain (1993) that must be highlighted are the following: (a) adjectival *sin* + INF is productive as an attribute to a noun and as a complement to the verb *estar*, which makes it more syntactically available than *no* + adjective constructions, which are unable to follow the verb *estar*, and (b) Pountain argues that *no* + adjective is limited to registers of greater formality, which may have contributed to the adjectival usage of *sin* + INF. Following this proposal, it would seem that adjectival *sin* + INF benefited from both intra- and extralinguistic factors working in favor of its emergence.<sup>5</sup>

### 3. Register and historical linguistics

In this section we discuss the notions of register, which as mentioned above, play an important role in the diachronic analysis of adjectival *sin* + INF. An issue in historical linguistic research is that terms such as ‘register’, ‘genre’, and several others are sometimes used interchangeably. Furthermore, Kytö (2019) points out that studies in the Romance tradition generally do not use the term ‘register’ at all. Here we briefly distinguish register and genre as follows. Biber and Conrad (2019: 5) define register as “a variety associated with a particular situation of use (including particular communicative purposes)” and emphasize that certain registers typically exhibit frequently occurring linguistic features because they are well-suited to the function, or purpose, of that register. Genres, on the other hand, are text groupings based upon the structural conventions that certain texts follow. Formal letters, for example, viewed as genres, are expected to include structural elements such as date, greeting, and signature. Importantly, genres are seen as being subordinate to register (Biber and Conrad 2019, Dorgeloh and Wanner 2010, Kytö 2019), in that genre is viewed as a relatively more specialized approach while register is applicable to all text varieties. Thus, regarding the present study, we choose to conduct our analysis in terms of register, being the broader term, and aim to avoid any potential confusion in terminologies.

Register plays a critical role in analyzing and explaining linguistic variation in corpora (Biber 2012, Biber et al. 2016, Goulart et al. 2020). The research literature in historical linguistics also supports the importance of register, although individual studies may vary with respect to terminologies used and approaches taken (see Kytö 2019 for a review). Biber and Gray (2013), for example, showed how differing sub-registers are a mediating factor for measuring grammatical change in news reportage and academic research and asserted that “Historical analyses that disregard these differences would confound the description of linguistic change with patterns that in

<sup>4</sup> The alternative constructions discussed by Pountain are certainly not an exhaustive list of alternatives to adjectival *sin* + INF. Indeed, as suggested by an anonymous reviewer, a natural alternative may be one that includes a noun: *Traiga una ensalada sin aderezar* vs. *Traiga una ensalada sin aderezos*, for example.

<sup>5</sup> An anonymous reviewer suggests that an additional possible motivation for the emergence of adjectival *sin* + INF is the fact that Latin constructions *esse* + *sine* + nominal ‘to be without + nominal’ to express an adjectival function are well attested.

fact reflect register differences” (Biber and Gray 2013: 130). In this section we will review recent corpus-based studies in historical language change, with an eye on their methodological approach to register, as they will bear certain influence on the methodology of the present study.

Recently, Yáñez-Bouza (2014) examined corpus data to show how prescriptivist writings of the 18<sup>th</sup> century affected the usage of preposition stranding and pied-piping constructions in English. Essentially, Yáñez-Bouza made and defended decisions with respect to where to place certain registers along a continuum of formality. Their continuum is copied below in Table 1.

**Table 1.** Yáñez-Bouza’s continuum of registers

		<b>Informal Setting</b>	↔	<b>Formal Setting</b>
<b>Medium: Speech- related</b>	<b>Speech-like</b>	Letters (private) Diaries		Letters (official)
	<b>Speech- purposed</b>	Plays		Sermons
<b>Medium: Writing- related</b>			Journals	Science, Medicine, & Legal Texts

**Source:** Yáñez-Bouza (2014: 33)

As shown in Table 1, Yáñez-Bouza created a continuum with informal and formal settings situated at either pole. They classified the registers of their corpus data, then, as being of a speech-related or a writing-related medium, which can be seen vertically oriented. Within the speech-related medium, there is a split between registers that are speech-like and those that are speech-purposed. Plays and sermons, for example, are speech-purposed, in that texts belonging to these registers have the purpose of being read aloud, often to establish a kind of dialogue between the reader and audience. However, they are placed on opposite ends of the formality continuum. Private letters and diaries, in turn, were placed on the informal end of the continuum, legal texts on the opposite formal end, and journals in between, since journals are informal in that they are like diaries, but the topics narrated in these journals tended to report non-private matters such as journeys or war campaigns. The results of this study indicated that from the years 1500 to 1900, preposition stranding occurred more frequently in informal than in formal registers. Private letters, diaries, and play texts, which are more closely associated with informal spoken registers, were more likely to show preposition stranding. More formal text registers, such as sermons, scientific texts, medical texts, legal texts, and official letters were less likely to predict uses of preposition stranding. Legal texts, in particular, showed the lowest frequency of preposition stranding. In short, Yáñez-Bouza established a continuum of registers, using the registers available in their corpus, in order to assess the degree of formality of different registers and their effect on the usage of certain linguistic constructions.

In studies of Spanish diachrony, the effect of textual varieties in corpus analysis has been examined in several studies within the paradigm known as “discourse traditions” (Koch 1997, Oesterreicher 1997, Kabatek 2005, Kabatek 2008), which approaches the production of an utterance as being influenced by previously produced utterances within the tradition of a certain discourse. Within this

line of research, for example, Rosemeyer (2015) studied differences in auxiliary selection (*ser* vs. *haber*) among three different discourse traditions in 16<sup>th</sup> century Spanish texts: historiographical texts, administrative documents, and private letters. They found that in private letters, the use of change-of-location predicates significantly predicts the selection of the *ser* auxiliary more than change-of-state predicates, a result not found in historiographical texts nor administrative documents. This finding highlights the notion that private letters are more likely to contain references to change-of-location events, where the use of the declining *ser* auxiliary had been most significantly maintained.

Recent studies in Spanish diachrony have included register, or some approach to text variety, in multi-factor regression analyses of historical Spanish grammatical change. Blas Arroyo and Schulte (2017), for example, examined the history of Spanish modal periphrases using data from a corpus of 16<sup>th</sup> and 18<sup>th</sup> century personal correspondence, following the work of Koch and Oesterreicher (1985), Oesterreicher (1996), and Oesterreicher (2004) who argued that personal letters are a textual resource of high communicative proximity and are thus more likely to display features of oral language. Their study measured the use of the modal constructions *haber de* + INF, *deber (de)* + INF, and *tener que* + INF, with one of the extralinguistic factors coded as formality, based on the main topic of the written letters and the closeness of the relationship between the sender and the addressee. The study found that between the 16<sup>th</sup> and 18<sup>th</sup> centuries, *haber de* + INF lowers in usage frequency relative to the other two variants. Interestingly, however, they found that in both time periods, less formal contexts that would be more representative of oral language favored the use of the older variant, *haber de* + INF. The authors argued that this result is due to an entrenchment effect, wherein *haber de* + INF was a highly frequent construction and therefore more resistant to change.

Rosemeyer (2017) analyzed the diachrony of the constructions *deber de* + INF and *deber* + INF ‘to must + INF’ in three distinct time periods of the GRADIA corpus, a multi-genre historical corpus of Spanish data. The author divided their analysis into three time periods: Old Spanish (1200-1499), Renaissance Spanish (1500-1699), and Modern Spanish (1700-2005), and then established a binary variable for formality, coding tokens as originating from genres classified either as formal or informal. In their regression analysis they found informal texts to be a significant predictor of *deber de* + INF in the Renaissance Spanish and Modern Spanish periods, and that the epistemic usage of *deber de* + INF (where it expresses a modal of possibility ‘might’ rather than one of necessity ‘must’) emerged in the Renaissance Spanish period in informal texts. In the Modern Spanish period, however, they found that formal texts were a significant predictor of the epistemic usage of *deber de* + INF. They explained these findings in arguing that a change from below occurred, wherein *deber de* + INF first began to exhibit its epistemic modality in informal texts, which later became solidified by prescriptive pressures in Modern Spanish formal text genres.

Lastly, Rosemeyer and Garachana (2019) examined the diachrony of the *lograr/conseguir* + INF ‘to manage to INF’ constructions from the years 1700 to 2005 in the *Corpus del nuevo diccionario histórico del español* (CDH), a corpus of historical Spanish data, and included a factor for *género* ‘genre’ in their regression analyses. In their analyses, they grouped the CDH’s genre labels into broader groupings to facilitate their analysis. One of these analyses compared the usage



during this time period of the *conseguir* + INF construction to the usage of the *lograr* + INF construction, and found that the Historiography genre was significantly less likely to show uses of *conseguir* + INF. However, possible reasons as to why this particular genre favored the use of *conseguir* + INF were not discussed, as the overall focus of the analysis seemed to lie more heavily upon syntactic and discourse-level factors.

In summary, corpus-based research in historical linguistics has underlined the role of text varieties in measuring and explaining language change. However, it is important to note that different studies take different approaches to how terms such as register and genre are used and operationalized. In particular, a difficult task for historical linguists has been showing how registers correspond to poles of formality and informality; in the studies reviewed above there are approaches that operationalize register formality along a continuum and there are those that operationalize it in binary terms. The approach to register taken in the present study will be explained in the sections that follow.

#### 4. Research question

In the previous sections we discussed the adjectival *sin* + INF construction and Pountain's claim that it emerged as an analytic alternative to other forms of negative adjectival constructions. Critically, however, one must note that the previous literature on adjectival *sin* + INF lacks any data-based substantiation of such claims. That is, while it is claimed in Pountain (1993) that factors involving register may have played a role in the emergence of the adjectival *sin* + INF construction, little data is provided that would support such a proposal. Thus, the present study asks the following research question: Is the use of the adjectival *sin* + INF construction in corpus data modulated by differences in register?

To answer this research question, we will measure the usage of two constructions in a historical Spanish corpus: the adjectival *sin* + INF construction and the *no* + PP construction. The *no* + PP construction makes a good candidate for this study because, as noted by the RAE (2009), it can alternate with the *sin* + INF construction, and furthermore, it was one of the negative adjectival variants discussed in Pountain's (1993) study. The rest of the paper is organized as follows. In Section 5, we discuss our methodological approach. In Section 6, we present the results of our analysis, and in Section 7 we discuss these results in relation to the research question at hand. Lastly, in Section 8 we conclude the paper and offer additional comments regarding the limitations of this study and suggestions for future research.

#### 5. Methodology

##### 5.1. CORDE

The present study utilized the *Corpus diacrónico del español* (CORDE). CORDE is a relatively large corpus of mostly written Spanish historical data. According to the CORDE website, CORDE contains a total of 236,709,914 words. The CORDE query site allows users to conduct searches for strings of words, with possible filter criteria

such as author, date, document title, medium, geographic origin, and topic.<sup>6</sup> We limited our analysis to data in the Peninsular Spanish subsection of CORDE (applying the *España* filter to the geographic origin field), since the large majority of the data (196,106,277 words, or 84% of the total word count) come from documents of Peninsular Spanish origin. No other limitations were applied to the other filters. In total, CORDE contains data from the 8<sup>th</sup> century to the 20<sup>th</sup> century. The documents contained in CORDE are tagged with a date referring to the document's date of publication, which is metadata we included in our dataset. In some cases, the date given in CORDE is an approximation between two dates (i.e., 1800–1810). In these cases, we took the first date given (1800, for example) to be coded as the date in our dataset. In the following sections more information will be provided with respect to how CORDE was utilized.

### 5.2. Constructions searched: adjectival *sin* + *INF*

Because CORDE is not part-of-speech tagged, strings were searched using the CORDE query site in order to capture tokens of the adjectival *sin* + *INF* construction. Specifically, we queried the word *sin* followed immediately by a word ending in \*er, \*ar, and \*ir or \*ír, which are the possible endings of infinitive verbs in Spanish. An asterisk denotes a wild card, such that \*er will return all words ending with the string *er*.<sup>7</sup> We then used a free web-scraping application known as Data Miner<sup>8</sup> to copy the search results from the CORDE website into .csv files. Search results from the CORDE website take the form of a list of tokens. Each token contains the queried string, approximately 7 to 8 words of the text to the right and left of the queried string, and the metadata regarding the document in which the token was found.

Once we had these tokens saved into .csv files, we used a Python script to extract only certain tokens from these lists. These were tokens in which the word *sin* was located within two words to the right of a definite or indefinite article with no intervening punctuation marks, and where the word ending in \*er, \*ar, or \*ir/\*ír appeared at the end of a phrase. This was done to specifically locate examples of *sin* + *INF* that served as adjectival complements to a noun, and that were not followed by additional material such as complements. To illustrate, some hypothetical examples of tokens that would have resulted from this search are given below in (8).

---

<sup>6</sup> The use of CORDE is not at all uncontroversial. As clearly expounded in Rodríguez Molina and Octavio de Toledo y Huerta (2017), CORDE suffers from a host of issues including but not limited to dating inaccuracy, cataloging errors, duplicate texts, and unclear criteria used for the selection of text samples. We recognize these shortcomings and sustain that CORDE is one of the most conveniently accessible Spanish historical corpora from which large amounts of data can be freely extracted for analysis. An anonymous reviewer also points to the comparability problem inherent in corpus analysis. That is, a historical corpus must be both diverse and uniform at the same time, in order to allow for a wide range of time periods and registers that are evenly distributed and thus comparable. A potential solution to this paradox is the use of parallel corpora, as effectuated in Rosemeyer and Enrique-Arias (2016).

<sup>7</sup> See Appendix 1 for the precise search queries used in this study.

<sup>8</sup> <https://data-miner.io/>

- (8) a. una caja<sub>1</sub> sin abrir.  
‘an unopened box’
- b. las dos<sub>1</sub> cajas<sub>2</sub> sin abrir;  
‘the two unopened boxes’
- c. ¡La caja<sub>1</sub> está<sub>2</sub> sin abrir!  
‘the box is unopened’

The examples in (8) contain subscripted numbers to indicate the distance between a definite or indefinite article and the word *sin*, which was limited to two words. In (8a), for example, the word *sin* is found one word to the right of the indefinite article *una*, and in (8b), the word *sin* is found two words to the right of the definite article *las*. All examples (8a) – (8c) are found at the end of a phrase. The definite or indefinite articles searched were specifically: *el, la, los, las, un, una, unos, unas, al, and del*. We define a word that appears at the ‘end of a phrase’ as a word that appears directly before the punctuation marks of period, comma, question mark, exclamation point, closing parenthesis, quotation mark, colon, and semi-colon. In short, this Python script was designed to only extract *sin* within close vicinity of definite or indefinite articles, in order to target the use of *sin* + INF as adjectival complements of noun phrases, and to find the use of *sin* + INF at the end of phrases, where they would likely be, following Bosque (1990) in that the construction is not modified by complements. Admittedly, this method does not capture all instances of adjectival *sin* + INF, given that nouns can appear without articles or at distances from *sin* greater than two words. However, this procedure was executed in order to find the clearest examples of adjectival *sin* + INF in a relatively large corpus that is not POS-tagged.

One of the issues with this query process is that while it captures strings such as *sin pintar*, it also gives results such as *sin azúcar* ‘without sugar’, a *sin* + noun construction, because searching for *sin* followed by the string *\*ar* does not distinguish between verbs and nouns. It would also produce results with *sin* + INF that have active values, such as those in (1a), rather than adjectival *sin* + INF constructions. For this reason, hand-coding was necessary at this point of the process. After running the Python script to limit our dataset to *sin* within two words of articles and at the end of phrases, the data were then hand-coded to identify the adjectival *sin* + INF construction, according to the following conditions: the construction *sin* + INF was coded as being adjectival if the infinitive verb was transitive, and the noun to its left was interpreted as its direct object, as in (9a) and (9b), or if the infinitive verb was an intransitive unaccusative, and the noun to its left was interpreted as its subject, as in (9c) and (9d).

- (9) a. Atravesaron una plaza sin adoquinar,  
‘They crossed an uncobbled plaza’  
(Caballero Bonald, José Manuel 1962; *Dos días de setiembre*)
- b. Esta es una cuestión sin resolver.  
‘This is an unresolved question’  
(Caro Baroja, Julio 1946; *Los Pueblos de España*)

- c. Y el Mariano sin venir.  
‘And Mariano (who has) not arrived.’  
(Delibes, Miguel 1948-1963; *La mortaja*)
- d. Desvela a un niño sin nacer, su historia múltiple, en la unidad de un cuerpo.  
‘(They) keep awake an unborn child, its multiple history, in the unit of a body.’  
(Prados, Emilio 1961; *La piedra escrita*)

To briefly summarize the process of locating adjectival *sin* + INF in CORDE, from start to finish: (1) we scraped all tokens of *sin* + \*ar, \*er, and \*ir/\*ír from the CORDE website and saved them to a .csv file, (2) we used a Python script to extract only tokens where *sin* was at most two words to the right of a definite or indefinite article and the infinitive verb was at the end of a phrase, and (3) we took these extracted tokens and hand-coded them to locate and verify examples of adjectival *sin* + INF that took as their internal arguments the nouns to their immediate left.

#### 5.4. Constructions searched: *no* + PP

To locate the *no* + PP construction, we followed a similar process to the one outlined in the previous section. We began by querying the CORDE website for all tokens of *no* followed immediately by \*ado and \*ido, which are the regular past participle endings. Feminine and plural forms and irregular past participles were also searched. Once again, a Python script was used to take these results and limit them to tokens where *no* was located within two words to the right of the same definite and indefinite articles used in the search for adjectival *sin* + INF, with no intervening punctuation marks. Tokens of *no* + PP were also limited to those that appeared at the end of a phrase, following the same definition for the end of a phrase as above. Because this query returns results that do not actually contain a past participle, the resulting tokens were then hand-coded to verify examples of *no* + PP. Specifically, these constructions were coded as being *no* + PP if the word following *no* was indeed a past participle originating from some verb, and if the construction as a whole modified the noun located to its immediate left. Examples of tokens that resulted from this process are shown in (10).

- (10) a. ...hallaron cantidades imponderables de un alcaloide no  
identificado, capaz de dilatar la pupila ocular.  
‘...they found imponderable quantities of an unidentified alkaloid,  
capable of dilating the ocular pupil.’  
(Font Quer, Pío 1962; *Plantas Medicinales: El Dioscorides  
Renovado*)
- b. ...poner el dedo sobre las heridas aún no cerradas.  
‘put the finger on the still unclosed wounds.’  
(Pérez Galdós, Benito 1899; *La estafeta romántica*)

Example (10a) shows the use of *no* + PP within one word of the indefinite article *un* and example (10b) shows *no* + PP within two words of the definite article

*las*. Both examples appear at the end of phrases. To briefly summarize the data collection process for the *no* + PP construction: (1) we scraped all tokens of *no* + past participle endings from CORDE and saved them to a .csv file, (2) we used a Python script to extract only tokens where *no* was at most two words to the right of a definite or indefinite article and where the past participle appeared at the end of a phrase, and (3) we took these extracted tokens and hand-coded them to locate and verify tokens of *no* + PP modifying some noun located to its immediate left.

#### 5.4. Additional classifications: Intervening verb

The above sections 5.2 and 5.3 describe the process of locating the tokens of adjectival *sin* + INF and *no* + PP to be compared in this study. These tokens were also coded for the following three factors: intervening verb, verb class, and register. In this subsection we describe the process of coding tokens for an intervening verb. The adjectival *sin* + INF and *no* + PP tokens collected were binarily hand-coded for whether or not they appeared as a complement to some verb whose subject was the noun that appeared in the construction. We call this verb an ‘intervening verb’, in that they were only coded when they appeared between the noun and the *sin* + INF or *no* + PP construction. The most frequent example of this intervening verb was the verb *estar* ‘to be’, but other examples included *quedar* ‘to remain’, and *seguir* ‘to continue’. The tokens in (11) exemplify cases in the dataset of adjectival *sin* + INF with intervening verbs.

- (11) a. La cama estaba sin hacer, todavía con la huella del cuerpo en las sábanas, arrugadas y ligeramente sucias.  
 ‘The bed was unmade, still with the imprint of the body on the sheets, wrinkled and lightly dirty.’  
 (Aldecoa, Ignacio 1954; *El fulgor y la sangre*)
- b. ...hasta el punto que algunos años las aceitunas quedaron sin recoger.  
 ‘...until the point where some years the olives remained unpicked.’  
 (Villalonga, Lorenzo 1956; *Bearn, o la sala de las muñecas*)

As shown in (11a), the verb *estar* intervenes between the noun *cama* ‘bed’ and the adjectival *sin* + INF construction, and in (11b), the verb *quedar* intervenes between the noun *aceitunas* ‘olives’ and the adjectival *sin* + INF construction. In both examples, the adjectival *sin* + INF construction is a complement to the intervening verbs. These tokens, then, were coded as containing intervening verbs. If adjectival *sin* + INF appeared directly to the right of the noun without any intervening verb, it was coded as not containing an intervening verb. The previously presented examples in (9) and (10), for example, would be coded as not containing an intervening verb.

#### 5.5. Additional classifications: Verb class

In this subsection we describe the process of coding the tokens for verb class. As noted by Bosque (1990), adjectival *sin* + INF occurs with verbs that indicate a change of state. It was therefore necessary to identify the aspectual class of the verbs used in the adjectival *sin* + INF and *no* + PP constructions found in this study. We

therefore classified each infinitive verb that appeared in the adjectival *sin* + INF construction tokens, and the infinitive verb from which the past participle was derived in the *no* + PP construction tokens. In order to use a marginally less subjective and more replicable method of classifying verbs than hand-coding each verb, we used the doctoral thesis of Sánchez Marco (2012: 189-194), who classified a relatively large number of verbs according to their aspectual class. Thus, Sánchez Marco's categories of "Accomplishments", "Achievements", "Degree Achievements", "Extent predicates", "Location and locatum verbs", and "Other change of state and change of location verbs" were considered to be change of state verbs, following that these verbs were the ones most likely to indicate a change of state. Sánchez Marco's aspectual classes of "States" and "Object experiencer psychological verbs" were considered to be not change of state verbs, following that these verbs were the ones least likely to indicate a change of state. Verbs produced in the present dataset that did not fall into one of these two lists were coded as "Unclassified". Thus, in total, there were three factor levels in coding verb class: "Change of State", "Not Change of State", and "Unclassified".

### 5.6. Additional classifications: Register

In this subsection we explain how CORDE organizes its data into different text varieties, and how we chose to incorporate those varieties into our analysis. CORDE classifies its documents into a total of thirteen main *temas* 'topics', and at times, more specific sub-*temas*. For example, Miguel Delibes' *Cinco horas con Mario* is classified under the main *tema* of *Narrativa (Prosa)* 'Narrative Prose', and also under the sub-*tema* of *Novela y otras formas similares* 'Novel and other similar forms'. In the present study we took a methodological approach similar to that of Yáñez-Bouza (2014) and Rosemeyer and Garachana (2019). That is, (1) the thirteen CORDE *temas* were reclassified into larger register groupings and (2) these larger register groupings were placed on a continuum of formality. These measures were executed in order to achieve a more balanced dataset (some *temas* were not represented well enough in the corpus to warrant their own groupings) and thus draw meaningful statistical conclusions. This approach is described as follows.

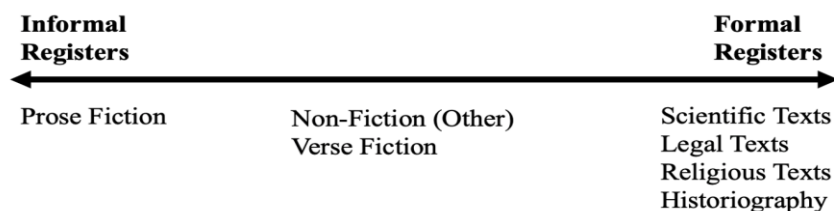
We compiled CORDE's thirteen main *temas* into seven larger registers, which we call: Prose Fiction, Verse Fiction, Non-Fiction (Other), Scientific Texts, Religious Texts, Legal Texts, and Historiography. In Table 2, the first column shows the thirteen main *temas* given in the CORDE database, with their names in Spanish as they appear in CORDE, and the author of this paper's English translations. The second column shows how these thirteen *temas* were collapsed into the seven registers used for this study.

**Table 2.** Classification of registers

<b>CORDE Tema</b>	<b>Register in Present Study</b>
<i>Lírica (Prosa)</i> ‘Poetry (Prose)’ <sup>9</sup>	Prose Fiction
<i>Narrativa (Prosa)</i> ‘Narrative (Prose)’	
<i>Teatro (Prosa)</i> ‘Theater (Prose)’	
<i>Lírica (Verso)</i> ‘Poetry (Verse)’	Verse Fiction
<i>Narrativa (Verso)</i> ‘Narrative (Verse)’	
<i>Dramática (Verso)</i> ‘Dramatic (Verse)’	
<i>Prensa</i> ‘Press’	Non-Fiction (Other)
<i>Sociedad</i> ‘Society’	
<i>Didáctica</i> ‘Didactic’	
<i>Ciencia y técnica</i> ‘Science and Technology’	Scientific Texts
<i>Religión</i> ‘Religion’	Religious Texts
<i>Derecho</i> ‘Law’	Legal Texts
<i>Historia y documentos</i> ‘History and Documents’	Historiography

As shown in Table 2 above, the decisions with respect to compilation were made largely based on topic. For example, the three prose-based fiction registers in CORDE were condensed into ‘Prose Fiction’, based on their shared overall register of Fiction. In a similar vein, the verse-based fiction groups were also combined under the register of ‘Verse Fiction’. The CORDE groupings of ‘Science and Technology’, ‘Religion’, ‘Law’, and ‘History and Documents’ were each thought to represent comparatively unique enough topics to merit their individual registers. The ‘Didactic’, ‘Society’, and ‘Press’ texts were combined into the register group ‘Non-Fiction (Other)’. To elaborate on the kind of texts that would appear in these topics, the ‘Didactic’ topic is a largely academic register containing academic essays such as Unamuno’s *En torno al casticismo*, transcribed speeches such as *Discurso de recepción en la Real Academia de Medicina*, and textbooks. The ‘Society’ topic covers texts on art, music, sociology, and travel, among a variety of general cultural themes. Some of these texts are academic in nature, such as Martín Gaité’s *Usos amorosos del dieciocho en España*, which originated from her doctoral thesis, whereas others were directed to a wider public, such as Ortega’s cookbook *1080 Recetas de cocina*. The ‘Press’ topic contains various press texts such as articles from the Spanish newspaper ABC.

In order to relate these registers to scales of formality and oral language, we followed an approach similar to that of Yáñez-Bouza (2014), in establishing the continuum of formality which will be utilized in the present study, shown below in Figure 1.

**Figure 1.** Registers according to formality

<sup>9</sup> “Prose poetry”, as it is known in English, is prose writing that maintains certain poetic qualities.

The following assumptions were made in the continuum shown in Figure 1. First, we followed Short (1996: 91) in assuming that “writers often create special effects by writing in ways which borrow characteristics associated with speech”. The register most likely to do so is Prose Fiction, wherein characters of novels and plays are given textual dialogue authored to borrow from speech. Thus, we placed Prose Fiction towards the informal end of the continuum.<sup>10</sup> We placed the registers of Scientific Texts, Legal Texts, Religious Texts, and Historiography towards the formal end, in assuming that the overall character of these texts is monologic, conservative, and more inclined to use archaic linguistic features. In between these two poles, we placed the registers of Non-Fiction (Other) and Verse Fiction. We argue that these two registers trend towards informality, in that Non-Fiction (Other) contains newspaper texts and writings of interest to a wider readership, and Verse Fiction contains plays and narrative texts more representative of oral speech. At the same time, these registers also trend towards formality, in that Non-Fiction (Other) includes texts for learned audiences, and Verse Fiction contains poetry texts that are structured in ways that do not reflect informal spoken language.

## 6. Results

### 6.1. Descriptive data

We first present a descriptive overview of the dataset that resulted from the data collection process outlined in Section 5. This dataset contains a total of 1,159 tokens, 240 of which contain the adjectival *sin* + INF construction and 919 of which contain the *no* + PP construction. The adjectival *sin* + INF construction was therefore the less frequently occurring construction of the two, accounting for 20.7% of the total dataset. Below, this dataset is presented according to chronological date in Section 6.1.1, according to intervening verb status in Section 6.1.2, according to verb class in Section 6.1.3, and according to register in Section 6.1.4. A mixed-effects logistic regression then analyzes this dataset in Section 6.2.

#### 6.1.1. Organization of data by date

Table 3 below outlines the token counts of the two constructions in the dataset, divided by century. In addition, the token frequencies of the single words *sin* and *no* from each century, found by searching the terms *sin* and *no* in the CORDE query site, are also included in Table 3 under the column headings of “Tokens *sin*” and “Tokens *no*”, in order to give some context with respect to the relative frequency of the target constructions. These relative frequencies appear under the column headings of “*sin* + INF per 1,000” and “*no* + PP per 1,000”. The relative frequency of adjectival *sin* + INF, for example, is the token count of adjectival *sin* + INF for every 1,000 tokens of the word *sin* found in CORDE. Similarly, the relative

---

<sup>10</sup> It is not our aim to say that all Prose Fiction is informal. Indeed, we only defend that Prose Fiction is more likely to reflect informal patterns of language when compared to the other registers measured here.



frequency of *no* + PP is the token count of *no* + PP for every 1,000 tokens of the word *no* found in CORDE.<sup>11</sup>

**Table 3.** Frequency of adjectival *sin* + INF and *no* + PP by date

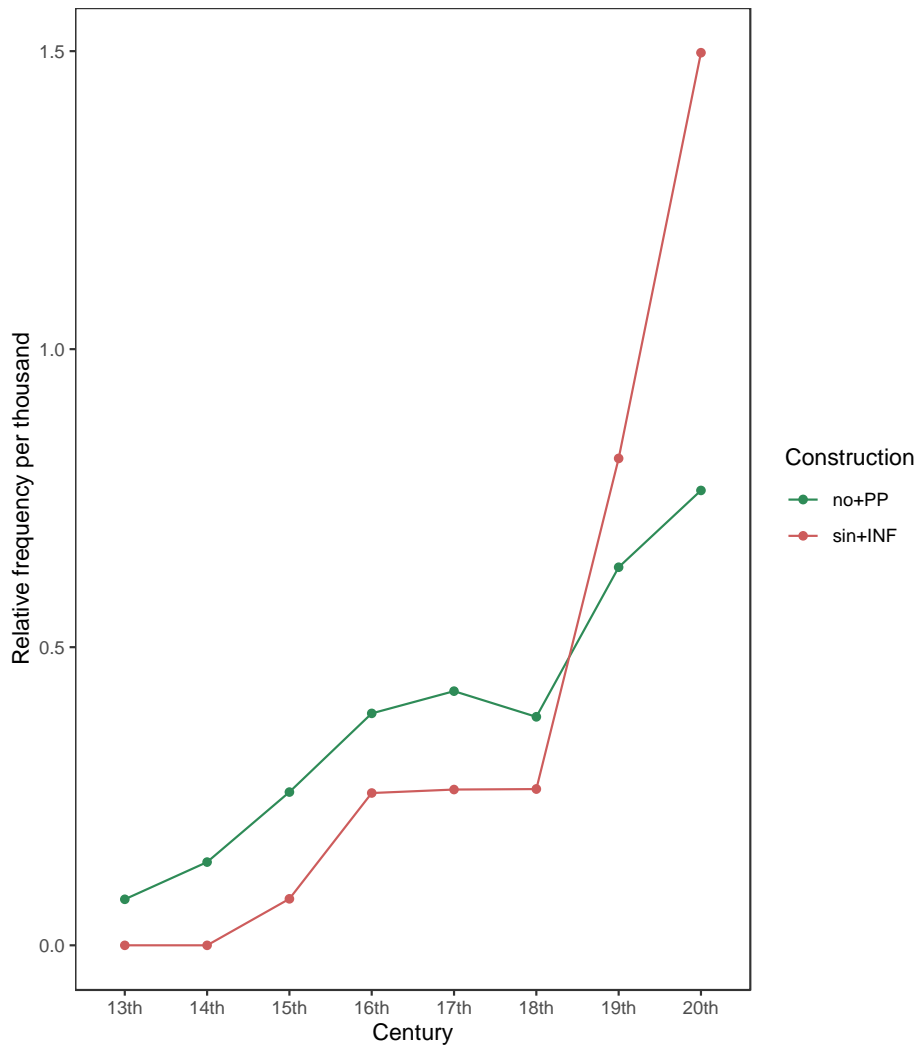
Century	Tokens <i>sin</i> + INF	Tokens <i>no</i> + PP	Tokens <i>sin</i>	Tokens <i>no</i>	<i>sin</i> + INF per 1,000	<i>no</i> + PP per 1,000
1201 – 1300	0	2	7,197	25,955	0	0.08
1301 – 1400	0	4	7,507	28,667	0	0.14
1401 – 1500	2	30	25,691	116,779	0.08	0.26
1501 – 1600	21	195	82,197	501,184	0.26	0.39
1601 – 1700	17	136	65,067	318,933	0.26	0.43
1701 – 1800	6	38	22,887	99,134	0.26	0.38
1801 – 1900	69	228	84,476	359,490	0.82	0.63
1901 – 2000	125	286	83,473	374,773	1.50	0.76

Indeed, a novel finding of the present study is that adjectival *sin* + INF appears as early as the 15<sup>th</sup> century in these Peninsular Spanish data. In the present dataset, the earliest two examples of adjectival *sin* + INF to appear are the following.

- (12) a. nos podría llevar Dios al hijo sin confesar,  
‘God could take away our unconfessed son’  
(Manrique, Gómez 1474; *Poesías*)
- b. mande quedar la dicha yglesia sin derribar,  
‘that they order to keep the said church undemolished’  
(Anonymous author 1491; *Fernando ordena a los obispos de Badajoz y Astorga consigan del Papa una disposición prohibiendo el ...*)

In other words, it seems that adjectival *sin* + INF has existed in Spanish since at least the 15<sup>th</sup> century, somewhat earlier than suggested in Pountain (1993). The relative frequency data in Table 3 is plotted in Figure 2 below, where century groupings are plotted on the x-axis, and the relative frequency rates of the two constructions are plotted on the y-axis.

<sup>11</sup> It is not possible to retrieve total word counts for chronological subsections of CORDE. We thus use relative frequency as shown in Table 3, as we were able to retrieve overall token counts of *sin* and *no* and apply the chronological filter to limit data to certain centuries. Since *sin* and *no* are relatively frequent words in Spanish, we see them as relatively reliable proxies for overall word count.

**Figure 2.** Relative frequency of adjectival *sin* + INF vs. *no* + PP

As visualized in Figure 2, adjectival *sin* + INF appears in CORDE in the 15<sup>th</sup> century, but does not eclipse *no* + PP in relative frequency until after the 18<sup>th</sup> century. In other words, after the 18<sup>th</sup> century, adjectival *sin* + INF appears more frequently relative to the lone word *sin* than *no* + PP relative to the lone word *no*. Despite this, the *no* + PP construction does not fall into disuse as the language enters its contemporary era.

#### 6.1.2. Organization of data by intervening verb

As discussed in Section 5.4 above, tokens were coded for whether or not the adjectival *sin* + INF and *no* + PP constructions appeared as complements to some verb that intervened directly between the noun and the respective construction. Table 4 below shows the number of tokens collected per construction, organized according to whether or not they appeared as complements to an intervening verb.

**Table 4.** Frequency of adjectival *sin* + INF and *no* + PP by intervening verb status

Intervening Verb Status	<i>sin</i> + INF	<i>no</i> + PP	% <i>sin</i> + INF	% <i>no</i> + PP
Verb Present	29	5	85.3	14.7
Verb Not Present	211	914	18.8	81.2

As shown in Table 4, intervening verbs were rare, appearing in 34 total tokens of the 1,159 tokens in the dataset. When they appeared, they occurred primarily with the adjectival *sin* + INF construction. Examples of such tokens can be seen above in (11a) and (11b).

### 6.1.3. Organization of data by verb class

As discussed in Section 5.5 above, each token in the present dataset was associated with a verb which was coded as being “Change of State”, “Not Change of State”, or “Unclassified”, according to the aspectual classifications of verbs compiled by Sánchez Marco (2012). Table 5 below summarizes how the tokens in this dataset were coded according to verb class.

**Table 5.** Frequency of adjectival *sin* + INF and *no* + PP by verb class

Verb Class	<i>sin</i> + INF	<i>no</i> + PP	% <i>sin</i> + INF	% <i>no</i> + PP
Change of State	61	94	39.4	60.6
Not Change of State	3	154	1.9	98.1
Unclassified	176	671	20.8	79.2

As Table 5 shows, adjectival *sin* + INF is more likely to appear with verbs classified as Change of State verbs than verbs classified as Not Change of State verbs.

### 6.1.4. Organization of data by register

Table 6 below outlines the number of tokens collected per construction and per register grouping, following the register groupings outlined in Section 5.6.

**Table 6.** Frequency of adjectival *sin* + INF and *no* + PP by register

Register	<i>sin</i> + INF	<i>no</i> + PP	% <i>sin</i> + INF	% <i>no</i> + PP
Prose Fiction	101	155	39.5	60.5
Non-Fiction (Other)	56	169	24.9	75.1
Religious Texts	12	70	14.6	85.4
Scientific Texts	38	195	16.3	83.7
Legal Texts	4	83	4.6	95.4
Historiography	16	83	16.2	83.8
Verse Fiction	13	164	7.3	92.7

As shown in Table 6, of the seven registers, adjectival *sin* + INF tokens are most likely to occur in the register of Prose Fiction, and least likely to occur in the register of Legal Texts.

## 6.2. Logistic regression analysis

While the descriptive data presented in the sections above give an overview of the dataset, further analysis is necessary to weigh these different factors together and measure which conditions favor the usage of the construction under study. We begin this analysis by recalling the research question, which sought to find out whether the use of adjectival *sin* + INF over *no* + PP is modulated by differences in register. To answer this question, this dataset was analyzed using a general mixed-effects logistic regression (*glmer*) in the *lme4* package in R (Bates et al. 2015). The 1,159 individual tokens that form the dataset were counted as individual observations. The model was

built as follows. The production of adjectival *sin* + INF versus *no* + PP was coded as a binary dependent variable, where a token with adjectival *sin* + INF was coded as 1 and a token with *no* + PP was coded as 0. This is to say that we aim to measure the probability that adjectival *sin* + INF will be chosen over *no* + PP. Fixed effects were the date of the document in which the token appeared (coded as a centered continuous variable), its intervening verb status (coded as “True” for containing an intervening verb or “False” for not containing an intervening verb), its verb class (coded as “Change of State”, “Not Change of State”, or “Unclassified”), and its register (coded as one of the seven registers outlined in Section 5.6). A random effect was included for the title of the document in which the token was located. Sum treatment contrasts were applied to the factor group of verb class (with baseline level as “Unclassified”) and the factor group of register (with baseline level as “Non-Fiction (Other)”). Since the intervening verb status factor group only contained two levels, default treatment contrasts were applied, with the intervening verb status of “False” used as the baseline level. The resulting regression model was calculated by entering the following formula in R:

```
glmer(sininf ~ date*verbclass + date*intervening + date*register + (1|document),
control=glmerControl(optimizer='bobyqa'), family = binomial)
```

The above formula, written in R syntax, takes the effects outlined above and produces a regression model in which the dependent variable of adjectival *sin* + INF (labeled as “sininf”) is regressed on the independent variables of chronological date (labeled as “date”), intervening verb status (labeled as “intervening”), verb class (labeled as “verbclass”), register (labeled as “register”), and a random variable of the document in which the token appeared (labeled as “document”). Interactions were measured between date and each of the other independent variables. In sum, this formula shows that the effects of chronological date, intervening verb status, verb class, register, and document are being weighed together in assessing the probability that a particular observation will contain the adjectival *sin* + INF construction, and not the *no* + PP construction. The output of this regression analysis is summarized in Table 7 below, with statistically significant effects ( $p < 0.05$ ) in bold.

**Table 7.** Results of Mixed-Effects Logistic Regression Analysis

Effect	Estimate	Std. Error	z value	p value
(Intercept)	-3.80754	0.83815	-4.543	5.55e-06
Date	1.15368	1.02389	1.127	0.259844
<b>Verb Class (Change of State)</b>	<b>2.18275</b>	<b>0.79262</b>	<b>2.754</b>	<b>0.00589</b>
<b>Verb Class (Not Change of State)</b>	<b>-3.42274</b>	<b>1.53701</b>	<b>-2.227</b>	<b>0.025956</b>
<b>Intervening Verb (True)</b>	<b>4.69961</b>	<b>1.11603</b>	<b>4.211</b>	<b>2.54e-05</b>
Register (Verse Fiction)	-0.39544	0.46741	-0.846	0.397542
<b>Register (Legal Texts)</b>	<b>-1.74532</b>	<b>0.70762</b>	<b>-2.466</b>	<b>0.013645</b>
Register (Historiography)	0.10075	0.42089	0.239	0.810815
<b>Register (Prose Fiction)</b>	<b>1.46831</b>	<b>0.32353</b>	<b>4.538</b>	<b>5.67e-06</b>
Register (Religious Texts)	-0.17128	0.53661	-0.319	0.749586
Register (Scientific Texts)	0.03404	0.3375	0.101	0.919669

Date:Verb Class (Change of State)	-0.90348	1.02819	-0.879	0.37956
Date:Verb Class (Not Change of State)	1.82547	2.02317	0.902	0.366906
<b>Date:Intervening Verb (True)</b>	<b>1.99002</b>	<b>0.99965</b>	<b>1.991</b>	<b>0.046511</b>
Date:Register (Verse Fiction)	0.43127	0.43549	0.99	0.322017
<b>Date:Register (Legal Texts)</b>	<b>-1.41646</b>	<b>0.55088</b>	<b>-2.571</b>	<b>0.010133</b>
Date:Register (Historiography)	0.1199	0.40861	0.293	0.769194
<b>Date:Register (Prose Fiction)</b>	<b>1.12759</b>	<b>0.32938</b>	<b>3.423</b>	<b>0.000619</b>
Date:Register (Religious Texts)	-0.46503	0.45117	-1.031	0.302668
Date:Register (Scientific Texts)	-0.00941	0.35848	-0.026	0.979058

As shown in Table 7, the effect of the Change of State verb class was a significant positive predictor of adjectival *sin* + INF relative to *no* + PP, and the Not Change of State verb class was a significant negative predictor of adjectival *sin* + INF relative to *no* + PP. The presence of an intervening verb was also a significant positive predictor of adjectival *sin* + INF. In terms of register, the Prose Fiction register was a significant positive predictor of adjectival *sin* + INF, while the Legal Texts register was a significant negative predictor of adjectival *sin* + INF. Examining the interactions between independent variables, there was a significant interaction of date and the intervening verb status, such that the presence of an intervening verb becomes increasingly likely to predict adjectival *sin* + INF as chronological date increases. There was also a significant interaction between date and the Prose Fiction register, such that the Prose Fiction register is increasingly likely to predict adjectival *sin* + INF with increases in date. Additionally, there was a significant interaction between date and the Legal Texts register, such that the Legal Texts registers is increasingly unlikely to predict adjectival *sin* + INF with increases in date. To visualize the interaction between date and register graphically, Figure 3 below plots mean-centered date on the x-axis, the probability of adjectival *sin* + INF over *no* + PP on the y-axis, according to the moderators of register represented by the different colored lines.

**Figure 3.** Interaction effect of date and register

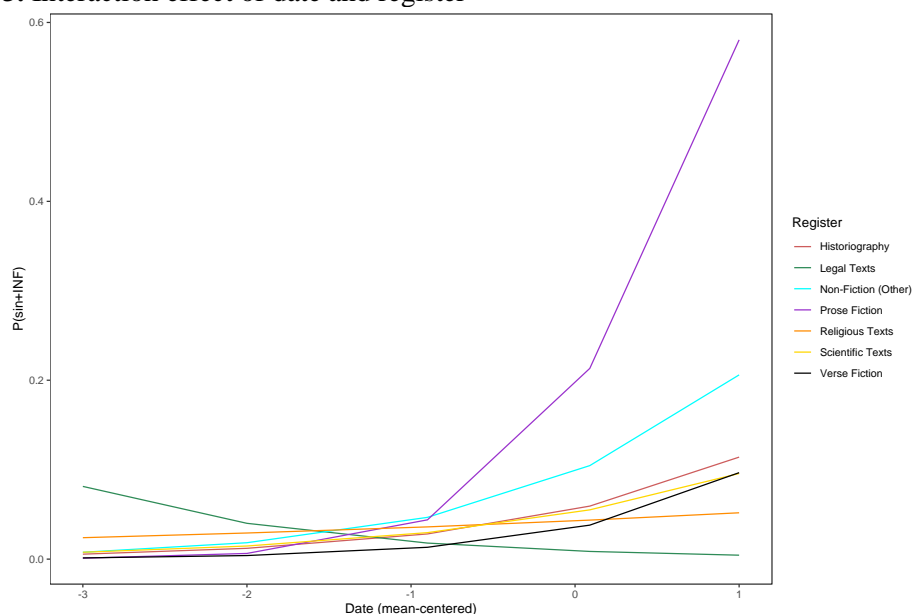


Figure 3 illustrates how the moderators of the Prose Fiction and Legal Texts registers, represented by the purple and green lines, widely differ in slope as date increases, in comparison to the lines that represent the moderators of the other registers. This is to say that with increasing chronological date, the register of Prose Fiction becomes increasingly more likely to contain examples of adjectival *sin* + INF, relative to *no* + PP, and the register of Legal Texts becomes increasingly more likely to contain examples of *no* + PP relative to adjectival *sin* + INF. The results presented in this section and their relevance to the research question will be discussed in the following section.

## 7. Discussion

This study set out to identify whether differences in textual formality, operationalized in terms of register differences, would modulate the usage of adjectival *sin* + INF versus *no* + PP in Spanish historical corpus data. The results show that adjectival *sin* + INF was indeed preferred to *no* + PP in Prose Fiction texts, which in Figure 1 we classified as being relatively more representative of informal speech patterns. Additionally, the register of Legal Texts, classified in Figure 1 as a register most representative of formal speech patterns, significantly disfavors the use of adjectival *sin* + INF, in favor of *no* + PP. Furthermore, the results show that as time went on, these particular effects became increasingly pronounced. Therefore, this quantitative analysis offers support for Pountain's (1993) proposal, which suggested that the propensity of the *no* + PP construction to appear in formal registers may have influenced the rise of the adjectival *sin* + INF construction. Apart from the influence of register, the results also confirmed certain linguistic assumptions regarding the usage of the adjectival *sin* + INF construction. In particular, the data showed a very strong preference for adjectival *sin* + INF over *no* + PP when used with Change of State verbs, which reflects Bosque's (1990, 2014) description of the adjectival *sin* + INF construction as only appearing with such verbs. Furthermore, the data also showed a preference for intervening verbs to appear with the adjectival *sin* + INF construction, which supports Pountain's analysis of adjectival *sin* + INF as having originated in an *estar sin* + INF construction which later led to a noun + *sin* + INF construction. This result also supports the notion that the adjectival *sin* + INF construction increased in usage not only because *no* + PP was more limited to formal registers, but also because it was more limited to certain syntactic contexts. As noted in Pountain (1993), of the two constructions, adjectival *sin* + INF has a wider distribution, in that it can be used as both an attribute to a noun and a complement to a predicate. The *no* + PP construction, on the other hand, seems to staunchly prefer being placed as an attribute to a noun.

A novel finding in these results is that the chronological date, as a main effect alone, did not predict increases in adjectival *sin* + INF relative to *no* + PP. This seems to reflect the descriptive data in Section 6.1.1, which show that although adjectival *sin* + INF greatly increases in relative frequency as Spanish moves towards the 20<sup>th</sup> century, *no* + PP does not completely disappear from usage. Thus, it is not entirely accurate to say that adjectival *sin* + INF simply increased in frequency relative to *no* + PP only as a result of increased time. As the interaction effect shows, chronological date does matter, but specifically within the registers of Prose Fiction

and Legal Texts. In other words, it is more accurate to describe increases in adjectival *sin* + INF as being modulated by increases in time, but only within a certain register of the language. That is, adjectival *sin* + INF increased in usage relative to *no* + PP as time went on, but only in registers of the language more representative of informal speech patterns.

Importantly, then, we must ask ourselves why adjectival *sin* + INF was preferred in more informal text registers. Is there some property of adjectival *sin* + INF that seems to make it more “informal” than *no* + PP? Indeed, the aforementioned limitation of *no* + PP to a smaller syntactic distribution may offer some clues. That is, the significant increase in frequency of adjectival *sin* + INF in informal text registers, compared to other registers, may actually be a reflection of the syntactic reality that *no* + PP was unable to be selected in as many sentence structures as adjectival *sin* + INF. Since adjectival *sin* + INF was available to follow both nouns and verbs, but *no* + PP was only available to follow nouns, it may be the case that over time adjectival *sin* + INF increasingly became the preferred variant in either syntactic position, encroaching on contexts previously occupied by *no* + PP. This change would have been most pronounced in informal texts, which are more likely to represent *au courant* speech patterns, and less pronounced in formal texts that conserved more archaic patterns of language. We leave this particular point of interest open to scrutiny in future research studies. To close this discussion, we highlight the importance of measuring register in corpus studies of historical linguistics, in order to capture the most accurate picture of diachronic change.

## 8. Conclusion

The primary goal of this study was to provide a quantitative test of a proposal regarding the usage of the adjectival *sin* + INF construction. The results provide evidence to support the proposal of Pountain (1993), whose analysis described adjectival *sin* + INF as a potential response to the unavailability of alternative negative adjectival constructions. In particular, the usage of adjectival *sin* + INF relative to *no* + PP in historical corpus data is significantly influenced by register formality. In this sense, that *no* + PP was restricted to more formal registers may have indeed provided the motivation for the increasing frequency of an adjectival use of the already existing *sin* + INF construction.

This study is not without certain limitations. One of the key limitations of the present study is that it did not measure the entire range of Spanish negative adjectival constructions, which would include prefixes such as *in-* and *des-*. Additionally, this study did not aim to ascertain the genesis of the adjectival use of the *sin* + INF construction, or how the prepositional infinitive *sin* + INF adopted a new adjectival function. Rather, this study only focused on analyzing the increasing usage of this construction in relation to register categories of textual data. We believe that an analysis of the precise origin of adjectival *sin* + INF would require a discussion of the typological implications of the construction and its counterparts in other Romance languages spoken in the Iberian Peninsula and beyond, in addition to the potential geographic variation that exists among different regions of the world where Spanish is spoken. It is interesting to note that equivalent adjectival constructions descendent from Latin SINE and infinitives are not available in certain other

Romance languages such as French. A potential piece to this entire puzzle, as an anonymous reviewer points out, is the fact that Spanish past participles and infinitives find their etymological origin in Latin perfect passive participles and present active infinitives, which differ in voice (passive versus active). This central voice distinction was not explored in the present study and should be discussed in future studies of the constructions under question. In addition, in classifying the semantic nature of verbs we relied on classifications already rendered by Sánchez Marco their study. While we chose to do this in order for this research to be carried out similarly by other researchers, it did indeed leave many verbs under the “Unclassified” grouping. Future studies may wish to consider the advantages and disadvantages of hand-coding each verb for aspectual verb class. Finally, the data analyzed in this study was limited to written data. Future corpus research on this topic may very well uncover novel findings in oral data.

In all, we conclude by noting that the results of the present study reflect the opinions of Biber and Gray (2013), who argued for the importance of considering register in historical linguistic analyses. Indeed, without having measured the effects of register, the diachronic usage of adjectival *sin* + INF would have been left only partially explained. We thus reassert that future investigations of linguistic phenomena in corpus data must make efforts to account for register differences in their analyses.



## Appendix 1

### Locating adjectival *sin* + INF using the CORDE query website

The fields of the CORDE query website were filled in as follows.

- Field *Consulta*: The following strings were searched: *sin \*ar*, *sin \*er*, *sin \*ir*, *sin \*ír*
- Field *Autor*: Leave blank (default option)
- Field *Obra*: Leave blank (default option)
- Field *Cronológico*: CORDE will only allow the viewing of results for queries that return less than 1,000 hits. It is therefore not possible to simply search *sin \*ar* from the year 0 to 2000 and scrape the results, which are much more than 1,000. Thus, the search and token collection process was broken up into many smaller searches. To do this, the *Cronológico* field, which filters date, was utilized such that data from 1801-1810 (for example) were searched and scraped, followed by data from 1811-1814, then 1815-1820, and so on and so forth.
- Field *Medio*: Leave on *Todos* (default option)
- Field *Geográfico*: Select the option *España* (this limits the data to Spain)
- Field *Tema*: Leave on *Todos* (default option)

### Locating *no* + PP using the CORDE query website

The fields of the CORDE query website were filled in as follows.

- Field *Consulta*: The following strings were searched one at a time: ‘no’ \*ado, ‘no’ \*ido, ‘no’ \*ído, ‘no’ \*ada, ‘no’ \*ida, ‘no’ \*ída, ‘no’ \*ados, ‘no’ \*idos, ‘no’ \*ídos, ‘no’ \*adas, ‘no’ \*idas, ‘no’ \*ídas
- Irregular past participles (in masculine, feminine, singular, and plural forms) were also searched: ‘no’ \*abierto, ‘no’ \*muerto, ‘no’ \*absuelto, ‘no’ \*resuelto, ‘no’ \*puesto, ‘no’ \*cubierto, ‘no’ \*dicho, ‘no’ \*roto, ‘no’ \*escrito, ‘no’ \*inscrito, ‘no’ \*satisfecho, ‘no’ \*hecho, ‘no’ \*visto, ‘no’ \*vuelto
- All other fields were set to the same settings as in the query for adjectival *sin* + INF

## Corpus Resources

*Corpus diacrónico del español*. REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) [en línea]. <<http://www.rae.es>> [Accessed June 18<sup>th</sup>, 2020]

*Peter Boyd-Bowman's Léxico hispanoamericano 1493-1993*. Eds. Ray Harris-Northall and John J. Nitti. New York: Hispanic Seminary of Medieval Studies, 2003-2007. Version 2.10 April 2007. Web version: General editor Ivy A. Corfis, with the assistance of Pablo Ancos and Fernando Tejedo and technical assistance by Philip Tibbetts. Online at [textred.spanport.wisc.edu](http://textred.spanport.wisc.edu) [Accessed March 20<sup>th</sup>, 2021]

## Acknowledgments

This work was supported by a Creighton University CURAS Faculty Research Fund. I thank Samantha Mora and Bretta Nienow for their assistance in this work. I would also like to thank the anonymous reviewers for their contributions. All errors are my own.

## References

- Blas Arroyo, José Luis & Kim Schulte. 2017. Competing modal periphrases in Spanish between the 16th and 18th centuries: A diachronic variationist approach. *Diachronica* 34: 1-39.  
<https://doi.org/10.1075/dia.34.1.01bla>
- Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1): 1-48.  
<https://doi.org/10.18637/jss.v067.i01>
- Biber, Douglas. 2012. Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory* 8(1): 9-37.  
<https://doi.org/10.1515/cllt-2012-0002>
- Biber, Douglas & Susan Conrad. 2019. *Register, genre, and style*. Cambridge: Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511814358>
- Biber, Douglas & Bethany Gray. 2013. Being specific about historical change: The influence of sub-register. *Journal of English Linguistics* 41(2): 104-134.  
<https://doi.org/10.1177/0075424212472509>
- Biber, Douglas, Jesse Egbert, Bethany Gray, Rahel Oppliger, & Benedikt Szmrecsanyi. 2016. Variationist versus text-linguistic approaches to grammatical change in English: Nominal modifiers of head nouns. In M. Kytö & P. Pahta (eds.), *The Cambridge handbook of English historical linguistics*, 351-375. Cambridge: Cambridge University Press.  
<https://doi.org/10.1017/CBO9781139600231.022>
- Bosque, Ignacio. 1990. Sobre el aspecto en los adjetivos y en los participios. In I. Bosque (ed.), *Tiempo y aspecto en español*, 177-211. Madrid: Cátedra.
- Bosque, Ignacio. 2014. On resultative past participles in Spanish. *Catalan Journal of Linguistics* 13: 41-77.  
<https://doi.org/10.5565/rev/catjl.155>
- Boyd-Bowman, Peter. 1971. *Léxico hispanoamericano del siglo XVI*. London: Tamesis.

Casalicchio, Jan. 2019. Gerunds become prepositional infinitives in Romance Small Clauses: The effects of later Merge to the syntactic spine. *Probus* 31(1): 75-117.

<https://doi.org/10.1515/probus-2019-0003>

Dorgeloh, Heidrun & Wanner, Anja. 2010. Introduction. In H. Dorgeloh & A. Wanner (eds.), *Syntactic Variation and Genre*, 1-26. Berlin, New York: De Gruyter Mouton.

<https://doi.org/10.1515/9783110226485.0.1>

Goulart, Larissa, Bethany Gray, Shelley Staples, Amanda Black, Aisha Shelton, Douglas Biber, Jesse Egbert, & Stacey Wizner. 2020. Linguistic perspectives on register. *Annual Review of Linguistics* 6: 435-455.

<https://doi.org/10.1146/annurev-linguistics-011718-012644>

Hansen, Maj-Britt Mosegaard & Jacqueline Visconti (eds.). 2014. *The diachrony of negation*. Amsterdam: John Benjamins Publishing Company.

<https://doi.org/10.1075/slcs.160>

Horn, Laurence R. 1989. *A Natural History of Negation*. Stanford, CA: CSLI Publications.

Jenset, Gard B. & Barbara McGillivray. 2017. *Quantitative historical linguistics: A corpus framework*. Oxford: Oxford University Press.

<https://doi.org/10.1093/oso/9780198718178.001.0001>

Kabatek, Johannes. 2005. Tradiciones discursivas y cambio lingüístico. *Lexis* 29(2): 151-177.

Kabatek, Johannes (ed.). 2008. *Sintaxis histórica del español y cambio lingüístico: Nuevas perspectivas desde las Tradiciones Discursivas*. Frankfurt/Madrid: Vervuert/Iberoamericana.

<https://doi.org/10.31819/9783865278623>

Koch, Peter. 1997. Diskurstraditionen: zu ihrem sprachtheoretischen Status und ihrer Dynamik. In B. Frank, T. Haye & D. Tophinke (eds.), *Gattungen mittelalterlicher Schriftlichkeit*, 43-79. Tübingen: Narr.

Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36: 15–43.

Kytö, Merja. 2019. Register in historical linguistics. *Register Studies* 1(1): 136-167.

<https://doi.org/10.1075/rs.18011.kyt>

Oesterreicher, Wulf. 1996. Lo hablado en lo escrito: Reflexiones metodológicas y aproximación a una tipología. In T. Kotsch, W. Oesterreicher & K. Zimmermann (eds.), *El español hablado y la cultura oral en España e Iberoamérica*, 317-340. Madrid and Frankfurt: Iberoamericana/Vervuert.

Oesterreicher, Wulf. 1997. Zur Fundierung von Diskurstraditionen. In B. Frank, T. Haye & D. Tophinke (eds.), *Gattungen mittelalterlicher Schriftlichkeit*, 19-42. Tübingen: Narr.

Oesterreicher, Wulf. 2004. Textos entre inmediatez y distancia comunicativas: El problema de lo hablado escrito en el Siglo de Oro. In R. Cano Aguilar (ed.), *Historia de la lengua española*, 729-769. Barcelona: Ariel.

Pountain, Christopher J. 1993. Spanish Adjective Negation. In D. Mackenzie & I. Michael (eds.), *Hispanic Linguistic Studies in Honour of F.W. Hodcroft*, 107-126. Llangrannog: Dolphin.

Real Academia Española (RAE). 2009. *Nueva gramática de la lengua española*. Madrid: Espasa Libros.

Rosemeyer, Malte. 2015. Entrenchment and discourse traditions in Spanish auxiliary selection. In R. Kailuweit & M. Rosemeyer (eds.), *Auxiliary Selection Revisited: Gradience and Gradualness*, 301-332. Berlin, Boston: De Gruyter.  
<https://doi.org/10.1515/9783110348866-011>

Rosemeyer, Malte. 2017. La historia de las perífrasis deber/deber de+infinitivo: variación, norma y géneros textuales. In M. Garachana (ed.), *La gramática en la diacronía. La evolución de las perífrasis modales en español*, 147-195. Madrid/Frankfurt am Main: Iberoamericana Verveurt.  
<https://doi.org/10.31819/9783954877362-005>

Rosemeyer, Malte & Andrés Enrique-Arias. 2016. A match made in heaven: Using parallel corpora and multinomial logistic regression to analyze the expression of possession in Old Spanish. *Language Variation and Change* 28(3): 307-334.  
<https://doi.org/10.1017/S0954394516000120>

Rosemeyer, Malte & Mar Garachana. 2019. De la consecución a la contraexpectación: la construccionalización de lograr/conseguir+infinitivo. *Studies in Hispanic and Lusophone Linguistics* 12(2): 383-422.  
<https://doi.org/10.1515/shll-2019-2018>

Sánchez Marco, Cristina. 2012. Tracing the development of Spanish participial constructions: An empirical study of language change. Ph.D. thesis, Universitat Pompeu Fabra.

Schulte, Kim. 2007. *Prepositional infinitives in Romance: A usage-based approach to syntactic change*. Bern, Switzerland: Peter Lang.

Short, Mick. 1996. *Exploring the Language of Poems, Plays and Prose*. London: Longman.  
<https://doi.org/10.4324/9781315842080>

Van der Auwera, Johan. 2010. On the diachrony of negation. In L. Horn (ed.), *The expression of negation*, 73-110. Berlin/New York: De Gruyter Mouton  
<https://doi.org/10.1515/9783110219302.73>

Whitt, Richard J (ed.). 2018. *Diachronic Corpora, Genre, and Language Change*. Amsterdam: John Benjamins Publishing Company.  
<https://doi.org/10.1075/scl.85>

Yáñez-Bouza, Nuria. 2014. *Grammar, rhetoric and usage in English: Preposition placement 1500–1900*. Cambridge: Cambridge University Press.  
<https://doi.org/10.1017/cbo9780511732522>