

Jessi Elana Aaron\* and Maria Fionda

# Getting closer: Codification of subjective semantic prosody in Spanish continuative aspect

Folklore says syntax has a mind of its own, and does not need semantics. I wish to add here: the same goes for semantics! I agree that semantics is not what Montague Grammar makes us believe it is [. . .]. But I do not agree that sentence structures cannot be motivated from semantics. Even if we grant that syntax has its own way, the fact that at the end of the day we want to communicate certain meanings will put pressure on language to put up with the requirements of semantics. In other words: both syntax and semantics will have certain intrinsic properties, and they will conspire to produce the systems that we call natural languages. (Kracht 2007: 51)

**Abstract:** The syntactic correlates of the diachronic process of subjectification within grammatical constructions, unlike that of discourse markers and connectives, do not include a cancellation of syntax. This can make the identification of subjectification within some grammaticalization processes difficult to identify. Pairs of purportedly synonymous constructions, such as continuative aspectual markers in Spanish, offer an ideal site to explore how certain linguistic contexts, through frequency, can come to be associated with more or less subjectivity. Six forms are included in this study: the phasal adverbs *aún* ‘still’ and *todavía* ‘still’ and the “phasal periphrastic” (Laca 2000) constructions including (semi-)auxiliaries: *seguir* ‘follow’ + *Vndo* and *continuar* ‘continue’ for positive polarity, and the corresponding *seguir sin* ‘follow without’ + INF and *continuar sin* ‘continue without’ + INF for negative polarity. In a variationist study of 481 occurrences of these forms from 1760–1860 in *Corpus Diacrónico del Español* (CORDE) and 2762 occurrences from 1975–1980 from *Corpus de Referencia del Español Actual* (CREA), it is found that the difference between these “synonyms” is linked, on the one hand, to contextual elements indicative of subjectivity, and on the other, to register. Furthermore, it is suggested that variation due to differing levels of subjectivity and register variation may share some characteristic patterns in the distribution of grammatical features.

**Keywords:** Spanish, aspect, continuative, subjectivity, variation

---

**\*Corresponding author: Jessi Elana Aaron:** Spanish and Portuguese Studies, University of Florida, P.O. Box 117405, Gainesville, FL 32611-7405, USA. E-mail: jeaaron@ufl.edu

**Maria Fionda:** Modern Languages, University of Mississippi, P.O. Box 1848, University, MS 38677-1848, USA. E-mail: mifionda@olemiss.edu

# 1 Introduction

## 1.1 Subjectivity in language

Despite the general focus in linguistics on the “referential” functions of language at the expense of the “emotive” (Jakobson 1960: 353–354), the pervasive nature of subjectivity in language (Benveniste 1966) has proven to be a complex and fascinating topic for linguists since the 1980s (cf. Stein and Wright 1995). Company Company defines subjectification as “a movement from the external, objective, referential domain to the internal, personal or interpersonal domain” (2008: 204). It is a dynamic, diachronic process in which speakers’ points of view and attitudes about events and social actors (possibly including the interlocutor) are infused into the grammar of a language and become conventionalized.

The syntactic correlates of the diachronic process of subjectification in the development of discourse markers and connectives – the jumping-off point for many of the studies on this topic – are becoming ever clearer, and include syntactic cancellation alongside increased internal fixedness (Traugott 2003: 636–642). For example, Spanish *a pesar de* ‘in spite of’ lost internal structure as it gained external flexibility in its development from a nominal construction describing the sorrow of a human referent (literally ‘to the sorrow of’) into the concessive marker it is today (Torres Cacoullous 2006).

Discourse markers and connectives, however, are not the only elements of language considered to display the results of a process of subjectification. The development of *go*-futures like those found in English, French and Spanish involves the bleaching of motion meaning and the conventionalization of speaker prediction as the most salient aspect of meaning (Langacker 1990: 23; Traugott 1995: 36, 50; Aaron 2010b). Similarly, the pronominally marked Mexican Spanish *salirse* ‘go out’ + REFL) is said to indicate counter-expectation, in contrast to the unmarked *salir* ‘go out’ (Aaron 2004; Aaron and Torres Cacoullous 2005). In these cases, since syntactic cancellation is not relevant, the identification of subjectification in progress must depend upon the discovery of shifts in patterns of co-occurrence towards a more internalized or intimate contextual landscape. This is brought about through the contextual absorption of the linguistic context (Bybee et al. 1994: 296; Bybee 2010: 176–177), which differs from the absorption of pragmatic context that occurs with pragmatic inference (Hopper and Traugott 1993: 63–93). Linguistic contextual absorption refers to a situation in which, as a “grammatical morpheme becomes more and more bleached of its meaning, it may well be only conventionally a part of the construction and in fact may derive its meaning from the overall construction rather than making a contribution itself”

(Bybee 2010: 176). A full-fledged example of such absorption can be seen in French negator *pas*, whose original lexical meaning was ‘step’; this meaning was bleached in contexts in which *pas* was used as an intensifier alongside *ne* in *ne . . . pas* constructions. In present-day spoken French, *ne* is often left out entirely, as *pas* has absorbed the negation meaning originally contained in *ne*, as in *je sais pas* (cf. *je ne sais pas*) ‘I don’t know.’

Subsumed under such absorption of the linguistic context would be the absorption of the semantic context. Such absorption of the semantic context, or semantic prosody (Sinclair 1991), surrounding a construction can also be found at the lexical level (Traugott and Dasher 2002), since “words have a tendency to take on the meanings of their habitual collocates” (Orpin 2005: 39). Empirical evidence for this has been found in recent studies of socially loaded terms, such as the negative meanings attached to *elderly* (Mautner 2007), and the pejorative social meanings attached the modern use of *lame* (Aaron 2010a). Pairs of purportedly synonymous constructions, such as continuative aspectual markers in Spanish, offer an ideal site to explore how certain linguistic contexts, through frequency, can come to be associated with more or less subjectivity. In the only other quantitative analysis of the structural correlates of subjectivity of which we are aware, Aaron and Torres Cacoulios (2005) found that factors such as the speaker’s relationship to the referent correlated with the development of a semantics of subjectivity with Spanish *salirse*.

This paper will offer a variationist account of the relationship between syntax, semantics and subjective contextual landscape as they have been manifested diachronically. Since “the recurrent patterns that constitute grammatical structure are reflected in frequencies of (co-)occurrence” (Aaron and Torres Cacoulios 2005: 906), corpus-based studies offer a chance to uncover “structural patterns of subjectivity” (Scheibman 2002) even where no overt subjective meaning is found. This means that a form’s association with more subjective contexts may not be immediately recognizable to speakers through a contemplation of the form’s semantics, but may only be traceable through quantitative analysis. Subjectivity is one of many linguistic traits that may not be immediately noted by speakers. For example, negative polarity was found to have a strong (nearly definitive) association with the morphological future in a quantitative study of Canadian French, though neither speakers nor previous scholars had picked up on this fact (Poplack and Turpin 1999).

## 1.2 Spanish continuative aspect

Speakers of Modern Spanish are faced with a wide array of choices when they wish to express progressive continuative aspect. Though the Present Progressive

*estar* ‘be’ + *Vndo* is the most obvious choice, adverbs *aún* ‘still’ and *todavía* ‘still’ are also options, as shown in (1) and (2).<sup>1</sup> These adverbs, along with Spanish *ya* ‘already’, like French *toujours*, *encore*, *déjà* and *enfin*, form a small class called *phasal adverbs*, since they describe the transition or lack of transition between two phases (Löbner 1989; Mosegaard Hansen 2002: 146–147). These adverbs have not been paid great attention in Spanish linguistics *per se*, but such words have increasingly been the focus of comparative literature (e.g., König 1991; Garrido 1992; van der Auwera 1998), as well as the subject of language-specific expositions on French (e.g., Franckel 1989; Victorri and Fuchs 1992; Gosselin 1996; Muller 1999). Semantically rich and difficult to define, phasal adverbs are characterized by a polysemy, as is the “*règle dans les langues naturelles* [rule in natural languages]” (Gross 1998: 103), that has captured the attention of many scholars.

- (1) *va en derechura a la cama de su madre, ve que todavía respira, y llena de temor, de pena y sobresalto, espera que vuelva Eduardo con el médico*  
 ‘she goes directly to her mother’s bed, sees that she is *still* breathing, and filled with fear, with sorrow and alarm, she hopes that Eduardo comes back with the doctor’  
 (CORDE)
- (2) *los salarios de las pocas industrias que **aún** trabajan se hayan estancado*  
 ‘the salaries of the few industries that are **still** working have stagnated’  
 (CREA)

At the same time, “phasal periphrastic” (Laca 2000) constructions are also available, including (semi-)auxiliaries: *seguir* ‘follow’ + *Vndo* and *continuar* ‘continue’ for positive polarity, as in (3) and (4), and the corresponding *seguir sin* ‘follow without’ + INF and *continuar sin* ‘continue without’ + INF for negative polarity. It is important to note that, while negative polarity is not impossible with the + *Vndo* construction, only one occurrence – shown in (5) – was found in these data. Overwhelmingly more common is the infinitival construction shown in (6). In the sta-

---

1 As Ozete (1983) points out, the most appropriate grammatical category for the Spanish *-ndo* form is contested: some scholars characterize it as a gerund (e.g., Bello and Cuervo 1973: 161), while others argue that it is a present participle (e.g., Caro 1976 [1870]). Given this paper’s focus on function in context, this battle is of little interest here. In order to avoid this issue, we refer to these forms as *-ndo* where possible. However, in the interest of clarity and crosslinguistic generalization, we will sometimes refer to these forms as gerunds, which is the term we find preferable, albeit imperfect. For more on the origins and functions of these forms, see Azofra Sierra (2008).

tistical analyses presented in the results section, these V *sin* + INF variants are included along with their positive polarity counterparts.

- (3) *El piano sigue sonando.*  
 ‘the piano keeps on playing’  
 (CREA)
- (4) *Emanu y Foder continúan tocando.*  
 ‘Emanu and Foder continue playing’  
 (CREA)
- (5) *mi consideración de “paria oficial” no fue y **no** sigue siendo mera presunción mía.*  
 ‘my evaluation as an “official pariah” is not and continues not being merely my presumption.’  
 (CREA)
- (6) *la actual prensa española sigue sin ser “totalmente libre” por dos razones*  
 ‘today’s Spanish press continues not being “fully free” for two reasons’  
 (CREA)

The progressive aspectual nature of such expressions is subsumed under the more general aspectual category of imperfective. Comrie’s (1976: 25) subdivision of the aspectual domain can be seen in Figure 1. Comrie defines the progressive as “the combination of continuous meaning and nonstativity” (1976: 52), though in combination with stative verbs, it instead designates a “temporary (contingent) state” (1976: 38).

In effect, this study falls within the vast body of work on Romance verbal periphrases, and within the study of the development of aspectual meanings

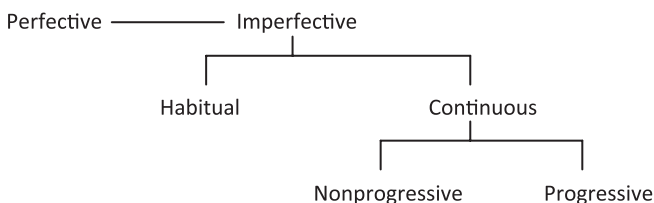


Fig. 1: Subcategories of imperfective aspect.

(e.g., Comrie 1990; Bybee et al. 1994), including progressive.<sup>2</sup> Romance verbal periphrases are used to express an array of temporal and aspectual meanings. The verbs meaning *go*, for example, participate in constructions expressing perfective (in Catalan, *anar* + INF), future (in Portuguese, *ir* + INF, in Spanish, *ir a* + INF, and in French, *aller* + INF), and progressive (in Spanish and Portuguese, *ir* + *Vndo*). The details of the entire body of work on this complex area of Romance grammar, however, certainly cannot be given their due here. For now, it is enough to keep in mind that Romance languages draw heavily on periphrastic verbal constructions in the development of tense and aspect marking. In Spanish, continuative meaning is often expressed through such constructions.

The majority of the work on verbal periphrasis with + *Vndo* in Spanish has focused on progressive expression. The constructions *estar* + *Vndo* and *andar* + *Vndo* have been the focus of several variationist studies by Torres Cacoullos (1999, 2000, 2001), and the former also has been examined quantitatively by Cortés-Torres (2005) in Puerto Rican Spanish. These + *Vndo* constructions, like those that are the focus of this paper, exemplify what appear to be cross-linguistically relevant paths of development of continuative markers, which tend to originate in locative expressions (Bybee et al. 1994). Other languages that have had similar developments include Albanian (Alimhilli 2001: 385), Chinese Xiang dialects (Wu 2005), English, Italian, Lakota (Pustet 2008: 332), Papiamentu (as a borrowing from Spanish with frequency influence from English [Sanchez 2002]), and Tohono O'odham (Fitzgerald 2004).<sup>3</sup>

These studies, however, have left some important questions unanswered. First, many studies on Spanish + *Vndo* constructions have given little to no attention to the less-frequent *seguir* and *continuar*, focusing instead on the more general progressives with *estar* 'be' and *andar* 'go around' (e.g., Torres Cacoullos 2000, 2001; Sanchez 2002). Oddly, García Fernández and Martínez-Atienza (2003), in the only study of which we are aware that focuses solely on continua-

<sup>2</sup> On Romance verbal periphrases, see, for example: Gougenheim (1929); Lyer (1934); Fente Gómez et al. (1972); Rojo (1974); Dias da Costa (1976); Dominicy (1977); Böckle (1980); Fernández de Castro (1990); Brianti (1992); García González (1992); Amenta (1994–1995); Squartini (1998).

<sup>3</sup> It is interesting to note that while Spanish speakers have developed several aspectual *-ndo* constructions, speakers of hexagonal French do not widely use gerund-based progressive constructions (Claveres 1997; Carvalho 2003: 100), reminding us that applying form-based grammatical categories crosslinguistically can lead to "idées fausses [false ideas]" (Meillet 1948 [1920]: 181). In the case of English, which diverges in this respect from other Germanic languages, Bertinetto (2000: 559) notes that sometimes the development of English Progressive construction with *-ing* has been attributed to prolonged contact with Romance (French in particular), while others have argued that this was an autonomous development (e.g., Scheffer 1975).

tive aspect in Spanish, fail to mention periphrasis with *seguir* or *continuar* (as well as putting aside, perhaps more understandably, the adverbials *todavía* and *aún*). Second, when these constructions are examined, there is the problem of synonymy. While synonymy is an attractive and intuitive concept, discovering real-world cases has troubled linguists for centuries (e.g., Girard 1718).

Are the periphrastic constructions with *seguir* and *continuar* in fact interchangeable, as suggested by some Romance scholars (e.g., Gómez Torrego 1988; Olbertz 1998; Laca 2000)? Is the same true of the phasal adverbs *aún* and *todavía*? Or is it more the case that we find neutralization of certain meaning differences in context (Sankoff 1988)? If it is the latter, which factors constrain these contexts? Are the differences in syntactic structure concurrent with differences in contextual distribution? In a situation of presupposed synonymy, hypotheses regarding linguistic or other constraints on variation are difficult to formulate. The periphrastic constructions have similar lexical origins and structure. Though only *seguir* was originally a motion verb, defined as “Ir detrás de alguno [to go behind someone]” (DRAE 1739), both *continuar* and *seguir* were considered to have the same meaning in some contexts by the mid-18th century. In 1729, the Real Academia Española defined *continuar* in the following way: “Proseguir, perseverar, no cessar en la obra que se ha empezado [To proceed, persevere, not cease in the work that has begun].” Similarly, a 1739 definition of *seguir* includes: “Significa tambien proseguir, ò continuar en lo empezado [It also means to proceed, or to continue in what was begun].” It is this sense of the word, in both cases, that is the most likely to have grammaticalized into the continuative aspectual + *Vndo* constructions used in modern Spanish.<sup>4</sup> This parallelism makes grammaticalization-based hypotheses less helpful, since both *seguir* and *continuar* would lead to the same predictions. At the same time, previous work on *estar* + *Vndo* and *andar* + *Vndo* has shown significant distributional differences between those constructions, particularly in register (Torres Cacoullos 1999).

In a proposal aimed at improving the synonym function of word processors, with the eventual goal of processing synonymy automatically, Gross (1998) created a usage-based method of analysis. In seeking synonyms, he notes, “l’élément à paraphraser ne peut pas être un mot ‘morphologique’ mais *un* des sens de ce mot. Or, comme la plupart d’entre eux sont polysémiques, il faut être en mesure de reconnaître dans une structure donnée le sens ou plutôt l’emploi du mot dont

---

4 Noetzel (2007) has argued that, based on syntactic and distributional evidence, the Latin origin of the Spanish Progressive is to be found in the Latin gerundive, and not the Latin Ablative gerund, as suggested by Lausberg (1966) and Penny (1991).

on veut rendre compte [the element to be paraphrased cannot be a ‘morphological’ word, but *one* of the senses of the word. Thus, since the majority of them are poly-semantic, one must take care to recognize in a given structure the ‘sense’ or perhaps the *use* of the word one wishes to describe]” (1997: 104 [emphasis in original]). It is such a usage-based analysis, framed within the more general usage-based theoretical approach to linguistic analysis (e.g., Barlow and Kemmer 2000), that will be employed here.

The application of a usage-based framework to semantics is illustrated prominently in Traugott and Dasher’s (2002) work on regularity in semantic change. Corpus-based approaches to apparent synonymy or homonymy at the lexical and morphosyntactic levels have been successful in teasing apart distinctions or making connections between constructions that are not apparent to the naked eye. For example, Torres Cacoullos (2001) found both linguistic and social conditioning in a study of variation between aspectual expressions involving Spanish *estar* ‘be located’ and *andar* ‘go around’ + *Vndo*, both of which express progressive meaning, in Mexican oral corpora. In a lexical analysis, Mautner (2007) showed how data mining can offer a multi-dimensional portrayal of the representation of a social group in a corpus-based study of *elderly*. Similarly, Aaron (2010a) used a corpus-based analysis to demonstrate how the social meaning of *lame* ‘uninteresting, socially inept’ developed out of elements frequently co-occurring with earlier meanings of *lame* as physical impairment, reflecting social attitudes towards disability.

Though we were not expecting to uncover social meanings like those found in the studies mentioned here, Gross’s suggestion stands for all constructions, not just those that happen to be used in contexts in which their social meaning becomes salient. Thus, we have chosen a usage-based approach because we believe it is likely that semantics and pragmatics are not strictly divided, but rather that the latter can become infused into the former in usage events that occur with relatively high frequency.

## 2 Data and methods

### 2.1 Corpus and variants

The data for this study were extracted from two searchable online corpora available from the Real Academia Española, one historical, one modern. All occurrences of *seguir/continuar* + *Vndo*, *seguir/continuar* *sin* + INF were extracted, as well as all occurrences of *aún* and *todavía* in co-occurrence with a finite verb in



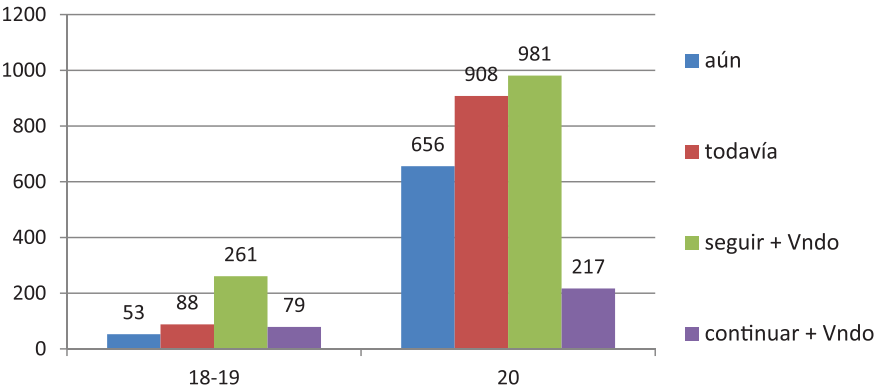


Fig. 2: Raw frequencies by data set, number of occurrences.

the Present tense.<sup>5</sup> The diachronic corpus, *Corpus Diacrónico del Español* (CORDE), consisted of 250 million words, and the 20th-century corpus, *Corpus de Referencia del Español Actual* (CREA), had 160+ million words. For the historical sample, the 100-year period from 1760–1860 was chosen, since the *seguir + Vndo* construction did not reach appreciable frequency before this period. This period yielded a total of 481 occurrences, after exclusions. The 20th-century period, a bit over a 100 years later, included a 5-year span, from 1975–1980, which yielded 2,762 occurrences after exclusions.<sup>6</sup>

The raw and relative frequencies of each variant by data set can be seen in Figure 2. In both data sets, the most frequent variant is *seguir* in construction (54% and 36%, respectively), followed by *todavía* (18% and 33%, respectively). *Aún* trails behind *todavía* in both time periods, at 11% and 24%, respectively. Constructions with *continuar*, third-most frequent in the 18th–19th-century data (16%), are much reduced in the 20th century, at only 8%.

5 Only orthographically accented occurrences of *aún* ‘still’ were extracted, which contrasts with unaccented *aun* ‘even’ according to normative standards. It is likely that some occurrences of *aun* could have been included; nonetheless, their exclusion should not affect our findings. The semantic link between these forms deserves more comment than can be given here, as does the possible polysemy of *todavía*, in ways similar to the contentious semantics of French *encore* (Victorri and Fuchs 1992; Mosegaard Hansen 2002).

6 A larger stretch of time was originally planned. The variants were found to be so frequent, however, that only a short time span, but with exhaustive extraction of all occurrences (offering the possibility of relative frequency calculation), was deemed most appropriate for our purposes.

## 2.2 Exclusions

Several contexts were excluded from this study in order to avoid the intrusion of factors not accounted for. These contexts, too infrequent to be studied quantitatively in these data, included: Verb Phrases with more than one variant, as in (7) and (8); the second verb in a parallel structure, as in (9); those that had no finite verb, as in (10); those presenting non-Present morphology, as in (11); and subjunctive mood, as in (12).<sup>7</sup> Also excluded were any occurrences in which the semantics of the form was clearly not progressive.<sup>8</sup>

- (7) *Aún* todavía continúan dándole fuego  
 ‘Even still they keep on feeding the fire’  
 (CORDE)
- (8) *La señora sigue confiando aún.*  
 ‘The lady **still** keeps trusting’  
 (CREA)
- (9) *y otros monarcas posteriores, que continúan reformando y adicionándolo hasta Egica inclusive.*  
 ‘and other earlier monarchs, who continue reforming and adding onto it even up to Egica’  
 (CORDE)
- (10) *Todo mi delito es haber querido y querer todavía un hombre que huyó.*  
 ‘my only crime is to have loved and to *still* love a man who fled’  
 (CREA)
- (11) *Los pamploneses ya machuchos y de buena memoria todavía **se estarán preguntando** por el significado*  
 ‘the guys from Pamplona, already all grown up and of good memory, **still must be asking themselves** about the meaning’  
 (CREA)

<sup>7</sup> Máslov (1984: 25, cited in Górbova 2000: 26) ascribes a pragmatic emphasis or “special emotionality” to the dual use of periphrastic progressive constructions and adverbs such as *siempre* ‘always’, as in (5), creating a kind of hyperbole.

<sup>8</sup> This was most common in lexicalized expressions like *todavía más* ‘even more.’ For example, in *Esta necesidad es todavía más urgente respecto de la amortización civil* ‘This need is **even more** urgent with respect to civil depreciation’ (CORDE), *todavía* does not express continuative aspect.

- (12) *sin que se **conozcan aún** las causas verdaderas*  
 ‘without the true causes **being known still**’  
 (CREA)

## 2.3 Coding and hypotheses

The remaining extracted occurrences were saved in an Excel spreadsheet, where they were then coded for seven linguistic factors and one stylistic factor. The linguistic factors included Verb class/verb type, Grammatical person, Clause type, Sentence type/polarity, Presence of a clitic, Presence of a locative, and Temporal adverbial presence and type. The breakdown of the original coding for each linguistic factor with examples can be seen below, as well as any hypotheses we had regarding these factors. All of these factors were included in all of the statistical runs; however, only the statistically significant factors in each analysis will be discussed in the results section.

### 2.3.1 Verb class

The verb class categories, initially guided by very broad notions of Aktionsart, such as stativity and dynamicity, were allowed to be molded by the distribution of the data. This meant that when a large group of items seemed to share particular semantic characteristics that were not shared by other items, a “category” emerged. Given the semantic compatibility of continuity with stativity, the categories that represented specific types of stativity, or even specific lexical items with lexical stativity, proliferated, while categories representing more dynamic types of action were fewer in number. The coding for verb class included the following categories: accomplishment/achievement (e.g., *lograr* ‘achieve’), activity (e.g., *escribir* ‘write’), durative (e.g., *durar* ‘last’), *estar* ‘be’, *gustar* ‘please’-type verbs,<sup>9</sup> *haber* ‘exist’, intransitive motion (e.g., *salir* ‘go out’), modal (e.g., *deber* ‘should’), other stative (e.g., *quedar* ‘stay’), perception (e.g., *ver* ‘see’), possessive (e.g., *tener* ‘have’), psychological (e.g., *desear* ‘wish’), verbs of saying (e.g., *hablar* ‘speak’), *ser* ‘be’, verbs of transfer (e.g., *dar* ‘give’), transitive displacement (e.g.,

<sup>9</sup> Verbs like *gustar* ‘please’ and *encantar* ‘enchant’ generally take a sentient being as an indirect object and often have an inanimate grammatical subject, while their English glosses would place the sentient being in the subject position. For example, *Le gusta el libro* would translate literally as ‘The book pleases her’, but it would generally be glossed as ‘She likes the book’.

*poner* ‘put’), and other. The verbs coded separately from any semantic category were chosen based on their extremely high frequency in these data, and were not chosen *a priori*. With regards to clearly polysemous verbs, e.g., *contar* ‘tell’ or ‘count’ or *tomar* ‘take’ or ‘drink’, each example was categorized according to the meaning of the verb in its context. In most cases, however, polysemies involved metaphorical extensions that could be applied to several verbs in the same verb class, and there was no need to divide these into separate categories. See Table A in the Appendix for details regarding the frequency of each category in each data set.

While we did not have hypotheses about each and every category, due to the purported synonymy of the variants, quantitative investigation often brings unexpected results; therefore, we aimed for a high level of detail. Nonetheless, we hypothesized that intransitive motion verbs and transitive displacement verbs would favor the use of the periphrastic constructions, due to their motion-based lexical origins. Similarly, statives, which include the general stative category as well as duratives, possessives, *estar* ‘be’, *haber* ‘exist’ and *ser* ‘be’, would disfavor these constructions, since their meaning is generally incompatible with the dynamicity of motion. Furthermore, we were interested in duratives and possessives in particular, since these contribute semantic content similar to the four variants of interest, so much so that there is a type of redundancy in these contexts, particularly in the former, as can be seen in (13).

- (13) *ha producido un grave fenómeno que aún perdura*  
 ‘it has produced a serious phenomenon that **still** goes on’  
 (CREA)

### 2.3.2 Grammatical person

This factor group encoded grammatical person, animacy, and certain special characteristics of the subject. This included: first-person singular; second-person singular informal; second-person singular formal; third-person singular animate; third-person singular inanimate; third-person singular with questionable animacy/subjecthood (body part, fetus, animal); third-person singular impersonal *se*; first-person plural; second-person plural; third-person plural animate; third-person singular inanimate; and official body. Again, a high level of detail was sought. These categories became more elaborate through our preliminary examination of the data. For example, “official body,” exemplified in (14), was added when we noticed that these were highly frequent in our data and could constitute a problem, since they were often third-person inanimate subjects (e.g.,

Bank of Spain) performing acts often carried out by animate subjects (e.g., regulate).

- (14) *del Banco de España que regula **aún** otros varios tipos de interés*  
 ‘of the *Bank of Spain* that **still** regulates various other types of interest’  
 (CREA)

### 2.3.3 Clause type

Clauses were coded as either main, as in (15), or subordinate, as in (16). It was hypothesized that the more marked subordinate context might correspond to a more formal register and therefore indicate an index of formality in its association with the forms under study.

- (15) *Ya ves, esto todavía me hace rabiar.*  
 ‘see, this *still* makes me enraged’  
 (CREA)
- (16) *Emanu, que sigue haciendo punto, intenta ir más de prisa*  
 ‘Emanu, who keeps knitting, tries to go more quickly’  
 (CREA)

### 2.3.4 Sentence type/polarity

This factor group included affirmative declarative (17), affirmative interrogative (18), negative declarative (19), and negative interrogative (20) contexts.

- (17) *Estoy un poco débil todavía.*  
 ‘I am *still* a little weak’  
 (CREA)
- (18) *¿Estáis aquí todavía?*  
 ‘Are you *still* here?’  
 (CREA)
- (19) *Enrique todavía no está preparado para esto.*  
 ‘Enrique is *still* not prepared for this’  
 (CREA)

- (20) *¿No te tientan todavía?*  
 ‘Don’t they *still* tempt you?’  
 (CREA)

### 2.3.5 Presence of a clitic

The presence or absence of a clitic was also coded, including both phonological cliticizing XPs (21) and head-adjoining clitics (22); an example of the absence of a clitic is seen in (23). It was hypothesized that the use of clitics with the analytic constructions under study would indicate higher levels of grammaticalization of these constructions as semi-auxiliaries.

- (21) *Puedo llegar a odiarte, creo que si sigues coaccionándome, controlándome, insultándome*  
 ‘I can come to hate you, I think that if you *keep coercing me*, controlling me, insulting me’  
 (CREA)
- (22) *a pesar de tanto como **le** sigo queriendo*  
 ‘in spite of so much, how I *keep on loving him*’  
 (CREA)
- (23) *porque los ojos siguen mirando lo que han leído*  
 ‘because the eyes *keep looking* at what they have read’  
 (CREA)

### 2.3.6 Presence of a locative

The data were coded for the presence or absence of a locative expression. It was hypothesized that the constructions that arose from verbs of motion, i.e., *seguir* and *continuar*, would show retention of older meanings through an association with locatives. It was further hypothesized that the least grammaticalized of these two adverbial constructions would be more strongly connected to its lexical origins and thus more strongly associated with locatives. Examples (24) and (25) show the presence and absence of a locative, respectively.

- (24) *Aquí en Andalucía todavía se va poco a los cementerios*  
 ‘here in Andalusia one *still* rarely goes *to the cemeteries*’  
 (CREA)

(25) *Nuestra vida sigue deslizándose como un manso río.*

‘our life *keeps slipping away* like a tame river’

(CREA)

It should be noted that locatives are generally semantically harmonic with verbs of motion, so certain types of verbs are more likely to co-occur with locative constructions. In fact, in an association that is in line with our hypothesis, locatives are significantly associated with intransitive motion verbs, which occurred with 11.5% ( $N = 32/278$ ) of the 20th-century locatives, compared to the rate of only 2% ( $N = 104/2,762$ ) intransitive motion verbs in the 20th-century data overall ( $p < .0000$ ). However, this measure cannot be said to correspond directly with verb class, as locatives nevertheless occurred across all coded verb classes.<sup>10</sup>

### 2.3.7 Temporal adverbial presence and type

The data were also coded for presence and type of temporal adverbial. It was hypothesized that the more grammaticalized variants would be less likely to co-occur with temporal adverbials that would support a continuative reading, since this reading would be the default interpretation for the variant itself. (For a longer discussion of the role of temporal adverbial co-occurrence in determining default meaning, see Aaron 2010b.) The temporal adverbial types coded included *ahora* ‘now’, *hoy* ‘today’, and other temporal adverbials, as in (26)–(28), respectively.

(26) *Y aquí sigue todavía **ahora** la escultura*

‘and here *still* remains the sculpture **now**’

(CREA)

(27) *situación que inexplicablemente hoy todavía se prolonga*

‘situation which **today** is *still* inexplicably prolonged’

(CREA)

(28) *A estas horas de la mañana todavía no se siente el calor*

‘**at these hours of the morning** you *still* don’t feel the heat’

(CREA)

<sup>10</sup> No such association was found, however, with transitive displacement verbs, which made up 3.2% ( $N = 9/278$ ) of occurrences with locatives, compared to 2.5% ( $N = 69/2,762$ ) in the 20th-century data as a whole, which is not a significant difference ( $p = .46$ ).

The first two were singled out for coding both due to their particularly strong contribution to continuative meaning and to their relatively high frequency, which offered some potential for quantitative analysis.

The 20th-century data were further coded for genre, which included press and literature. As explained in Section 1, very few hypotheses can be made *a priori* regarding the constraints on variation between these four variants. While some studies on morphosyntactic variation, such as those on future expression in Romance or subject expression in pro-drop languages like Spanish, have the benefit of decades of descriptive, prescriptive, and theoretical work on which to base hypotheses, this is not the case when forms have been generally considered to be interchangeable and thus relatively ignored in the literature (see, for another example, García (2011) on Spanish *quizá* and *quizás* ‘perhaps’). Our case here is particularly problematic because both periphrastic constructions share similar lexical origins. This study, then, is based on a wider hypothesis, indeed, the hypothesis that forms the basis for the variationist paradigm: structured heterogeneity (Weinreich et al. 1968). In their foundational 1968 paper, Weinreich et al. state,

The key to a rational conception of language change – indeed, of language itself – is the possibility of describing orderly differentiation in a language serving a community. We will argue that nativelike command of heterogeneous structures is not a matter of multidialectalism or “mere” performance, but is part of unilingual linguistic competence.

This paper, then, is a mission of discovery and a search for the orderliness these scholars proposed. The factors chosen, in this light, were those we thought to be the mostly likely to constrain the variability of interest here because they are often found to constrain morphosyntactic variation, both in Spanish and cross-linguistically.

Once the data were coded for these factor groups, they were then subjected to multivariate multiple regression analyses using GoldVarb X. Various factors were collapsed in the analyses; for clarity, these will be discussed as they arise in the results. The results of these analyses are presented in the next section.

### 3 Results

In this section, we will present the results of six separate variable rule analyses using GoldVarb X. While it is clear that all four variants are in competition with each other in discourse, here we have chosen to focus in on differences between pairs of forms, and not on the choice of each form versus the three others, as



might be expected in a function-based analysis of variation. This choice was made for two reasons. First, any attempt to lump these forms into one “whole” would offer the false impression that this is a representation of the variation of the functional space of continuative aspect. In fact, there are many forms that can have this function; we focus on two pairs that are relatively frequent and that have been treated as synonymous in the literature. Second, the combination of all four variants into one analysis, in which the adverbials and the periphrastic constructions would be combined, would prevent us from: a) teasing out differences between syntactically similar “synonyms”; and b) discovering what role, if any, the nature of the form (adverbial or periphrastic construction) has in predicting its contexts of occurrence. Thus, in all of our analyses, each form is treated separately. In the end, with the various analyses, we hope to have offered a multifaceted and coherent explanation of how formally diverse and apparently synonymous forms can compete in the expression of aspect.

In Section 3.1, we will tackle the most obvious and also the easiest question: what are the distributional differences between the periphrastic constructions and the adverbs? In this case, we will rely on the most frequent variants in each category, i.e., *seguir* and *todavía*. In Section 3.2, we will examine the periphrastic constructions side by side, in order to determine if there are any significant differences in distribution. In Section 3.3, we will examine any significant differences in the patterns of distribution of the two adverbs under consideration. In both cases, we will see that these purportedly synonymous constructions, in context, are used somewhat differently. Each section will include both an analysis of 20th-century data alongside an analysis of 18th- and 19th-century data. In each case, we will argue that the difference between these “synonyms” is linked, on the one hand, to subjectivity, and on the other, to register. Finally, in section 3.4, we will look at the situation more globally in a search for a more general story of variation and change. We will suggest that variation due to differing levels of subjectivity and register variation may share some characteristic patterns in the distribution of grammatical features.

### 3.1 Most frequent variants: *seguir* + *Vndo* and *todavía*

Given the striking syntactic differences present between *seguir* + *Vndo* and *todavía*, it may seem strange to consider them as variants of the same variable at all. However, a function-based analysis, which does not separate *a priori* more grammatical expression from more lexical expression, would call for us to do just that. These differences provide several jumping-off points for the formation of hypotheses. The first two hypotheses are related to verb type. We understand

*seguir* as a motion verb, as its origins are found in Latin \**sequire*, from *sequor* ‘follow,’ as shown in (29).

(29) *Cēdāmus Phoebō et, monitī, meliōra **sequāmur***

‘Let us yield to Phoebus and, having been warned, **let us follow** better (paths).’

(Virgil)

While *seguir* (and *continuar*) have developed new meanings over time, it is the oldest and most concrete meanings that may offer the strongest cues regarding any patterns of occurrence that have become entrenched in usage over time. Thus, first, with the hypothesis that there will be persistence of earlier meanings, we can hypothesize that, as a motion-verb construction, *seguir* + *Vndo* will be disfavored in contexts incompatible with motion, such as stative and psychological contexts.

Second, we may posit that *seguir* + *Vndo* will be further disfavored with stative verbs because they are not frequently used as gerunds (Sanchez 2002). The next two hypotheses are related to syntactic differences between the two variants. The third hypothesis, then, is that the syntactically more marked construction (*seguir* + *Vndo*) will be disfavored in subordinate clauses, due to their own markedness.<sup>11</sup> Finally, fourth, based on Biber’s (1995) finding that more syntactically complex (or, here, marked) constructions are associated with higher registers, we hypothesize that the relatively more marked *seguir* + *Vndo* will show evidence of more formal or distant contexts. This hypothesis, in fact, would lead us to the opposite prediction, that is, that subordinate clauses would be favorable to *seguir* + *Vndo*. Given that these two equally plausible hypotheses point in different directions, the results will allow us to determine which factor is more likely to act as a stronger constraint, the avoidance of excessive markedness or the association of marked forms with marked contexts. In other words, do marked forms repel other marked forms, or does markedness tend to cluster in both form and context?

In Table 1, we see the results for the GoldVarb analysis of the historical data, which includes 349 occurrences, 75% (261) *seguir* + *Vndo* and 25% (88) *todavía*. In this data set, we can see that verb class is the factor group with the highest magnitude of effect on the choice to use *seguir* + *Vndo* instead of *todavía*, since it is the group with the highest Range (63). If we look in the probability column, we

<sup>11</sup> Here we use the term “marked” and “markedness” with the understanding advocated by Haspelmath (2006).

**Table 1:** Variable Rule Analyses, *seguir* + Vndo vs. *todavía* + P, 18th–19th centuries. Input probability: 0.841 (75%), N = 261/349

Factor Group	Rel. frequency	Probability	% of data
<b>Verb class</b>			
Other	93	.74	63
Psych, perception, modal	57	.24	14
Stative	36	.11	24
Range		.63	
<b>Temporal adverbial</b>			
Present	97	.89	8
Absent	73	.45	92
Range		.44	
<b>Clause type</b>			
Main	79	.56	67
Subordinate	66	.39	33
Range		.17	

$p < .05$ . Log likelihood = -130.445; Total Chi-square = 71.3171; Chi-square/cell = 0.7204  
Not significant: Grammatical person, Sentence type/polarity, Presence of a clitic, Presence of a locative.  
*Note:* As the input probability reflects the corrected mean, it differs from the general mean.

can see that this periphrastic construction is disfavored in contexts in which psychological, perception, modal and stative verbs are used, as in (30), as shown by the probability weights far below 0.50. All other verb classes are more favorable to *seguir* + Vndo than to *todavía*. The strong disfavoring with stative verbs was predicted in our hypotheses, and it is most likely related to the semantics of the progressive itself. Indeed, historically, like in English, the progressive construction may have been less felicitous with stative verbs than with other types of verbs. However, as we shall see in a moment, this general tendency no longer appears to be the case. This explanation does not extend to the psychological, perception, and modal verb types, nor can these results be explained by any other hypothesis we have made thus far. As is sometimes the case in variationist literature, particularly when addressing under-studied phenomena, discoveries that were not a part of the original hypotheses can be made. The emergence of surprising or unexpected results is one of the benefits of the quantification of natural data. This was the case, for example, in Poplack and Turpin’s (1999: 155) study of French future expression in Canada, in which negative polarity was found to be overwhelmingly associated with the Inflected Future, an effect that was generally “unknown, unacknowledged or merely noted in passing elsewhere” and was

therefore not featured as a strong hypothesis in their original plan. We will return later to the particular results presented here.

- (30) *Bien veo que seguís necesitando un consejero prudente*  
 ‘I see well that you continue needing a prudent adviser’  
 (CORDE)

The second-most influential factor group is that of the presence of temporal adverbials. As we can see in Table 1, when another temporal adverbial is present, writers of this period are much less likely to use the adverb *todavía*, instead opting for *seguir* + *Vndo*. This may be a reflection of the reluctance to use more than one temporal adverb in the same sentence, which, in fact, only occurred once in this data set, shown in (31).

- (31) ***Ahora** todavía es muchacho*  
 ‘**now** he is still a boy’  
 (CORDE) (only occurrence, N = 1/29)

The third significant factor group in the 18th- and 19th-century data was that of clause type. As predicted, writers preferred to use *todavía* in subordinate clauses. These results are in line with the third hypothesis, which stated that the syntactic markedness of subordinate clauses would lead writers to use *todavía* instead of *seguir* + *Vndo*. In turn, then, the fourth hypothesis is not borne out in these data; *seguir* shows no evidence of being linked to formal contexts or registers.

Now let’s compare the same two variants for the 20th century, shown in Table 2. As we can see, once again, verb class is the most influential factor group for this time period, with the highest range (67). Because more data are available for the century, we are able to see a more finely tuned portrait of what is going on with this factor group. For the most part, we see the same types of verbs disfavoring *seguir* + *Vndo* as they did in the century before: psychological verbs, modal verbs, and most stative verbs. However, we see that not all stative verbs are equal. While both *haber* and *estar* are strongly disfavorable to the *seguir* construction, other stative verbs are somewhat less so. In fact, the verb *ser* is more favorable to *seguir* + *Vndo* than to *todavía*, with a probability weight of 0.65. This suggests that the earlier tendency to avoid stative verbs with progressive tense in Spanish may be undergoing a process of enhanced lexical conditioning.

It is worth remembering, then, that stative verbs are not prohibited in Spanish progressive + *Vndo* constructions (Górbova 2000: 25,n12), as is sometimes suggested (e.g., Sanchez 2002). As Miguel Aparicio reminds us, “En general, casi

**Table 2:** Variable rule analyses, *seguir* + *Vndo* vs. *todavía* + *P*, 20th century. Input probability: 0.513 (52%), N = 981/1889

Factor Group	Rel. frequency	Probability	% of data
<b>Verb class</b>			
Other activity	77	.75	26
Other	71	.72	4
<i>Ser</i>	65	.65	21
Motion	61	.58	4
Psych, <i>gustar</i> -type	46	.48	12
Other stative	30	.29	19
<i>Estar</i>	16	.17	7
Modal	8	.08	5
<i>Haber</i>	7	.07	2
Range		.67	
<b>Grammatical person</b>			
Official body	92	.89	3
3 animate	65	.60	34
1s	44	.46	11
Inanimate, impers.	43	.41	48
2s+p	38	.39	3
Range		.50	
<b>Sentence type/polarity</b>			
Pos. declarative	56	.54	85
Neg. declarative	29	.28	13
Interrogative	22	.21	2
Range		.33	
<b>Clause type</b>			
Main or protasis	56	.54	67
Other subordinate	43	.41	33
Range		.13	
<b>Clitic</b>			
Absent	53	.52	79
Present	47	.43	21
Range		.9	

$p < .05$ . Log likelihood = -1002.389; Total Chi-square = 321.3232; Chi-square/cell = 0.9563  
 Not significant: Presence of a locative, Temporal adverbial presence and type.

todos los verbos estativos pueden entrar a formar parte de formas progresivas. La excepción es el propio *estar* [In general, almost all stative verbs can be a part of progressive forms. The exception is *estar* itself]" (1999: 3013). Here, there do not appear to be any absolute lexical restrictions, at least not with constructions other than *estar* + *Vndo*. This is revealed in (32)–(38); as the quantitative evidence

shows, such constructions are nonetheless strongly disfavored in most stative contexts. I provide several examples here to dispel any doubts about the viability of such a combination. Note that the English translations have been left purposefully awkward, in an attempt to demonstrate the typological differences between Spanish and English in this area of grammar.<sup>12</sup>

- (32) *Vicente Gaos, pues, no ha muerto; sigue estando conmigo.*  
 ‘Vicente Gaos, well, he hasn’t died; he keeps being with me’  
 (CREA)
- (33) *la bandera continúa estando ahí y continuará*  
 ‘the flag continues being there and it will continue’  
 (CREA)
- (34) *Su origen, posiblemente caucásico, sigue estando oscuro.*  
 ‘its origin, possibly Caucasian, keeps being obscure’  
 (CREA)
- (35) *pero sé que sigo teniendo miedo y todavía insinúo*  
 ‘but I know that I keep having fear [being afraid] and I still insinuate’  
 (CREA)
- (36) *mis propios actos siguen pareciéndome una incógnita.*  
 ‘my own acts keep seeming to me an unknown’  
 (CREA)
- (37) *Ahora: aquello fue aquello, y con esas cosas sigue habiendo un lío que a ver por dónde lo coges*  
 ‘Now: that was that, and with those things there keeps being a problem that let’s see where you grab it from [how you deal with it]’  
 (CREA)
- (38) *Tragedia parece que no hay, pero oscuridad sigue habiéndola.*  
 ‘It seems a tragedy that there is none, but there keeps being darkness’  
 (CREA)

---

<sup>12</sup> This tendency toward avoidance of using stative verbs in continuative/progressive aspect may have more to do with semantic similarity (and thus an apparent redundancy) than with any inherent semantic incompatibility. In fact, in the Salish language Quinault the same verbal morphology is used to mark both continuative and stative aspect (Rowicka 2006: 466). The most salient semantic difference between the two aspects may be that of the former as an internal process and the latter as an internal state (Lipsky 2003; Górbova 2000).

The second-most influential factor group in this data set is one that did not appear as significant in the earlier data set: Grammatical person. This may be related to the elevated frequency of a certain kind of grammatical person, which may be due to the nature of the data in the 20th century. This “person” is what we have called “official body”. This category emerged during the coding of the data, and includes entities such as government bodies, countries, and institutions. It was decided that these entities required a separate category because they were neither singular animate third persons, in the normal sense (because they were not human beings or other animals *per se*), nor were they inanimate objects, since they were often portrayed as taking actions that only sentient beings can take (for example, deciding), as in (39). It is these entities, followed weakly by third-person animate subjects, that are most favorable to *seguir* + *Vndo* in the 20th century. All other persons are more favorable to *todavía*.

- (39) *salvo Irlanda, que sigue dependiendo cuantitativamente del sector primario*  
 ‘except for Ireland, which keeps depending quantitatively on the primary sector’  
 (CORDE)

The final three significant factor groups, we posit, are related to markedness. In all three cases, for sentence type/polarity, clause type, and clitic presence, *seguir* + *Vndo* is favored only in the simplest constructions: declaratives with positive polarity, main clauses, when there is no clitic present. It appears that the tendency for speakers/writers to favor *todavía* in more marked constructions carried over from the century before.

It is interesting to note that the temporal adverbial constraint that we saw in the earlier data set is not statistically significant in the 20th-century data set. Preliminary results showed, however, that *seguir* + *Vndo* was still relatively more likely to occur with a temporal adverb than was *todavía*, except for the case of *hoy* ‘today’, which is more likely to occur with *todavía*. Figure 3 shows these results in more detail.

In sum, then, we have found that the 19th-century and the 20th-century data sets tell more or less the same story. Some parts of it were predictable (and predicted), such as the tendency of *seguir* + *Vndo* to be less frequently used with stative verbs and in more marked constructions. Other parts, however, were not predicted. For example, our hypotheses did not explain why psychological and modal verbs would be less favorable to *seguir* + *Vndo*, or why official bodies would be more favorable. It is possible that these two factors are in fact a reflection of the same kind of constraint. In the first case, we see verb types that are very close to the speaker/writer/narrator and her or his desires, experiences or

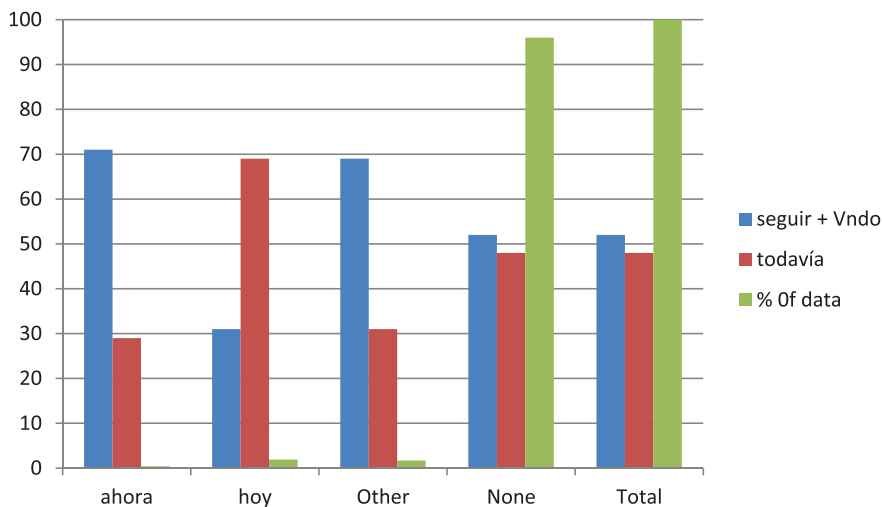


Fig. 3: Temporal adverbial co-occurrence with *seguir + Vndo* and *todavía*, percentages.

opinions, as illustrated in (40)–(42), from the 18th and 19th centuries, and in (43)–(45), from the 20th century.

- (40) *Y rehusé, Clemencia -prosiguió Pablo-, porque tú hacías un sacrificio grande en casarte conmigo, y yo uno cruel en negarme a ello, y quise que el sacrificio estuviese de mi parte y no de la tuya esto prueba que te amaba, y sigo amando sin esperanza, Clemencia; y el amor que vive sin alimento, esto es, sin esperanzas que lo sostengan, es de alta esfera, o inmortal como alma.*

‘And I refused, Clemencia – continued Pablo –, because you were making a big sacrifice in marrying me, and I a cruel one to refuse it, and I wanted the sacrifice to be mine and not yours, this proves that I loved you, and I *keep loving* without hope, Clemencia; and love that lives without nourishment, that is, without hopes to sustain it, is of a high sphere, or immortal, like a soul.’

(CORDE)

- (41) *Esperad, que no sabéis todavía el exceso de vuestras desgracias. ¿Cómo, quedan todavía rayos que disparar a mi rabiosa suerte? ¿Mi sufrimiento no agotó toda la saña de su furioso poder?*

‘Wait, for you do not yet **know** the excess of your disgraces. What, there are still lightning bolts to strike at my rabid luck? My suffering did not deplete all of the ire of its furious power?’

(CORDE)



- (42) *¡Oh terribles atractivos de aquellos dulces ojos, fraguas de ardientes rayos que llegan a inflamar mi memoria y los deseos que **debo** sofocar todavía!*  
 ‘O terrible attractions of those sweet eyes, forgeries of burning rays that come to inflame my memory and the desires that I **must** suffocate still!’  
 (CORDE)
- (43) *Y ayúdale tú, querida, con sencillez y cariño, si todavía **quieres** casarte con un arquitecto.*  
 ‘And help him yourself, my dear, simply and with affection, if you still want to marry an architect.’  
 (CREA)
- (44) *Estoy parado en una esquina de la plaza, contemplando la fachada de la iglesia. [...] Estas iglesias del Levante español no son muy notables, para nada figuran en los libros de arte; pero a mí me gusta visitarlas. Y, sin duda, no sólo a mí: junto al atrio, en estas horas tempranas de la mañana, veo estacionados varios coches de matrícula extranjera: turistas alemanes, suecos, belgas . . . A estas horas de la mañana todavía no se **siente** el calor, y el silencio es perfecto.*  
 ‘I am standing at the corner of the plaza, contemplating the church’s façade. [...] These churches from the Spanish Levante are not very noteworthy, they don’t figure at all in the art books; but me, I like to visit them. And, without a doubt, it’s not just me: next to the atrium, at these early hours of the morning you *still* don’t **feel** the heat, and the silence is perfect.’  
 (CREA)
- (45) *Volvió a analizar mi rostro y habló con tono envejecido: Usted no debe tener hijos todavía. **Aún** es joven.*  
 ‘He analyzed my face again and spoke with an aged tone: You shouldn’t have children yet. You’re **still** young.’  
 (CREA)

The older examples are particularly dramatic. In (40), the narrator is speaking to a woman for whom he has an undying love and for whom he has sacrificed his heart to see her free. In (41), the narrator is quoting his servant, who is telling him that things are worse than he thinks. The narrator’s son was sentenced to death, his wife killed, and his daughter raped. He is holding a sword in his hands, ready to avenge the violence done against her, when his servant informs him of the “excesses” of his misfortune: all of his earthly possessions are to be confiscated that very same day. The narrator in (42) is in the rapture of a love that must be suppressed. We can also see the link between psychological and modal verbs and more personal contexts in the 20th-century data. In (43), a mother offers advice to

her daughter about her potential future marriage. In (44), the narrator describes the pleasure of his early-morning stroll. In (45), the narrator receives advice on the very personal choice of whether it is time for him to have children.

On the other hand, official bodies favor *seguir* + *Vndo*. Here, we are dealing with bodies that are, by their very nature, relatively distant from the personal experience or desires of the average speaker/writer/narrator, as we see in the 20th-century examples (46)–(47).

- (46) *otro **sector de la dirección** no acepta todavía la validez de esa fórmula*  
 ‘another **sector of the administration** still does not accept the validity of this formula’  
 (CREA)

- (47) *Irán sigue siendo un país de poder fragmentado*  
 ‘**Iran** continues to be a country with fragmented power’  
 (CREA)

In both cases, then, we have *todavía* associated with more personal or subjective contexts, and *seguir* + *Vndo* associated with more official contexts (though we must note, however, that it is slightly disfavored (.41) with third-person inanimates and impersonal constructions, just as it is with first and second person). Note that the former also appears to be associated with more marked syntax, and the latter with less marked syntax. These results, in themselves, are counter-intuitive, since it does not necessarily make sense that more marked syntax would co-occur more frequently with subjective contexts. Also, as we will see in the next two sections, such a conclusion is not only counter-intuitive, but also incorrect, mainly because the picture of the interplay between form and function is incomplete. While we cannot present the entire picture here, quantitative evidence of the patterns of occurrence of these variants’ less-frequent counterparts, *continuar* + *Vndo* and *aún*, will help us to understand the relationship between subjectivity and syntax in this variable context.

### 3.2 Periphrastic constructions

If the significant differences that we found between the most frequent variants, *seguir* + *Vndo* and *todavía*, are due to different preferences based on syntactic characteristics, then we would expect that constructions that overlap in meaning and also have similar syntactic characteristics and origins would not show such differences. To see if this is indeed the case, we will now examine any variable

patterns in the occurrence of the two periphrastic constructions *seguir* + *Vndo* vs. *continuar* + *Vndo*. Not only do both *seguir* and *continuar* overlap semantically in that they are both used to create grammatical constructions with continuative aspect, but also in their everyday lexical usage they overlap, as shown in (48).

- (48) *las ferias que se inician en París, sigue en Elda y continúa en Italia.*  
 ‘the vacation that starts in Paris, goes on to Elda and continues in Italy.’  
 (CREA)

Given this situation, hypotheses are much more difficult to formulate. In their examination of the cross-linguistic characteristics of posture verbs, Schaefer and Egbokhare (2008) found that only the most broad-based hypotheses could be formulated:

Through a survey of geographically diverse languages, Newman (2002a) proposes that the human experience of posture underlies the cross-linguistic properties of these verbs. The classic dilemma for such a position concerns why some elements of human experience and not others find their way into languages, and why these elements get arranged the way they do. Two contrasting hypotheses arise: posture verb properties will differ substantially from language to language, *exhibiting extreme variability*, or they will differ little, *manifesting great similarity*. (Schaefer and Egbokhare 2008: 216 [emphasis ours])

In other words, we may hypothesize that these purportedly synonymous and syntactically similar constructions may be very different, or very much the same. Of course, a third possibility also arises: these constructions may be slightly different in their distribution and use, but not as different as constructions that are syntactically less similar (like adverb *todavía*). In this case, these differences might arise because of a difference in frequency of use, which might point to (or lead to) a difference in the extent of grammaticalization and/or subjectification of each form.

Table 3 shows the results of the GoldVarb analysis for the production of *seguir* + *Vndo* vs. *continuar* + *Vndo* in the 18th- and 19th-century data. Three factor groups are found to be significant during this time period: verb class, the presence of a temporal adverbial, and the presence of a clitic. We have also included in Table 3 the non-significant results for grammatical person, which will be relevant to our discussion of our results from the 20th-century data.

Since the two variants examined here share both syntax and general semantic origins, any significant differences can most likely be attributed to the conventionalization of differences in usage during the process of grammaticalization. The appearance of Motion verbs (semantically harmonic with the lexical origins of both constructions, but somewhat more so with *seguir*) as more favorable to

**Table 3:** GoldVarb Analyses, *seguir* + *Vndo* vs. *continuar* + *Vndo*, 18th–19th centuries. Input probability: 0.793 (76%), N = 250/329

Factor Group	Rel. frequency	Probability	% of data
<b>Verb class</b>			
Psychological	96	.88	8
Motion	88	.70	18
Other	77	.46	9
Activity	76	.45	36
Stative	73	.42	12
Change of state	56	.25	16
Range		.63	
<b>Temporal adverbial</b>			
Present	87	.77	10
Absent	75	.47	90
Range		.30	
<b>Clitic</b>			
Absent	80	.56	33
Present	69	.39	67
Range		.17	
<b>Grammatical person</b>			
1s+p	90	[n.s.]	12
3s animate	82	[n.s.]	23
2s	78	[n.s.]	3
3p animate	77	[n.s.]	15
3 inanimate/animal	70	[n.s.]	28
3 impersonal	68	[n.s.]	19
2p	67	[n.s.]	1

$p < .05$ . Log likelihood = -158.346; Total Chi-square = 119.1481; Chi-square/cell = 0.8511  
 Not significant and not shown: Clause type, Sentence type/polarity, Presence of a locative.

*seguir* + *Vndo* could point to this variant as a later development, and thus less generalized contextually. This interpretation is supported by our failed search for this construction in earlier texts. The most favorable verb class for *seguir* + *Vndo* is psychological verbs, as in (49). In this case, we have the most frequent variant associated also with what are likely to be more subjective contexts.

- (49) *Valentín. Vengo á felicitarte. ¿Tu adorada duquesa habrá colmado ya tus deseos? Eres el más feliz de los hombres, ¿eh?*  
*General. ¡Tengo destrozado el corazón! Mis esperanzas, mis ilusiones . . . , todo ha desaparecido!*  
*Valentín. (Riendo.) ¡Calla! . . . ¿Esas tenemos? . . . ¿Y tú sigues adorándola?*

*General. ¡La aborrezco y la desprecio!*

'Valentine. I've come to congratulate you. Your beloved duchess must have fulfilled your wishes? You're the happiest man alive, huh?

General. My heart is in shreds! My hopes, my dreams . . . , everything has disappeared!

Valentine. (Laughing.) Shut up! Do we have those? . . . And you still adore her?

General. I loathe and despise her!'

(CORDE)

The two other factor groups that had a significant effect in the 18th–19th-century data were the presence of a temporal adverbial and the presence of a clitic pronoun. The co-occurrence of a temporal adverbial, as in (50), is favorable to *seguir* + *Vndo*, but the presence of a clitic pronoun, also in (50), has the opposite effect. Both of these may point to the fact that the *seguir* construction was still in its early stages of development. On the one hand, this construction may still have depended on temporal adverbials (such as *hoy* 'today' in [50]), to assure a continuative reading. Also, it is possible that the lack of clitic co-occurrence with the *seguir* construction shows a lack of full grammaticalization of this construction.

(50) *Estaba yo en aquel tiempo (como sigo estándolo hoy) verdaderamente prendado de mi mujer*

'At that time I was (as I still am today) truly attached to my wife'

(CORDE)

Table 4 shows the results for the variable rule analyses for these same two variants the 20th century. Here, surprisingly, the factors that had a significant effect in the 19th-century formula were not shown to be significant in this analysis. Instead, grammatical person, not significant in the earlier data set, is the only factor group shown to be significant in this data set. First- and second-person singular are the only two contexts shown to be strongly favorable to *seguir* + *Vndo*. First-person plural, third-person singular, and animate third-person plural are neither strongly favorable nor strongly disfavorable to either variant. On the other hand, three grammatical persons are less favorable to *seguir* + *Vndo*: third-person inanimates, official bodies, and the third-person impersonal.

If we compare the results in Table 4 with the results for grammatical person shown in Table 3, we can see that these patterns were also in effect in earlier centuries, but not at significant levels. These particular results suggest that *seguir* + *Vndo* is favored over *continuar* + *Vndo* in personal contexts, referring to

**Table 4:** Variable rule analyses, *seguir* + Vndo vs. *continuar* + Vndo, 20th century. Input probability: 0.831 (82%), N = 981/1198

Factor Group	Rel. frequency	Probability	% of data
<b>Grammatical person</b>			
1s	97	.86	5
2s	92	.71	2
1p	85	.53	3
3s	84	.52	58
3p animate	80	.45	16
3p inanimate	73	.36	6
Official body	66	.29	7
3 impersonal	64	.26	3
Range		60	

$p < .05$ . Log likelihood = -538.298; Total Chi-square = 278.1855; Chi-square/cell = 0.9593  
 Not significant: Verb class/verb type, Clause type, Sentence type/polarity, Presence of a clitic, Presence of a locative, Temporal adverbial presence and type.

emotions and actions of those in the immediate situation, as shown in (51) and (52).

(51) *Usted me cae bien, joven -otra vez-. Por eso le sigo aguantando, a pesar de sus impertinencias.*

‘I like you, young man –again-. That’s why I continue to put up with you, despite your impertinences.’

(CREA)

(52) *En los días sucesivos continúo hablando de la playa*

‘in the next few days I keep talking about the beach’

(CREA)

Such systematic results may point to register differences that would not be evident in a purely semantic analysis. These differences, however, could nonetheless lead to subjectification, which could, in turn, bring about an eventual semantic transformation of one construction (*seguir* + Vndo), while the other (*continuar* + Vndo) would not go through such a process.

### 3.3 Adverbs

If we expected to find similar patterns of occurrence for the periphrastic constructions, this would also be the case for the adverbs *todavía* and *aún*. However,

**Table 5:** Variable Rule Analyses, *aún* + P vs. *todavía* + P, 18th–19th centuries. Input probability: 0.345 (38%), N = 53/141

Factor Group	Rel. frequency	Probability	% of data
<b>Verb class</b>			
Durative/possess.	72	.84	26
Other	33	.44	37
<b>Other stative</b>	<b>19</b>	<b>.29</b>	<b>38</b>
Range		55	
<b>Grammatical person</b>			
<b>3p animate</b>	<b>62</b>	<b>.80</b>	<b>9</b>
<b>Impersonal</b>	<b>59</b>	<b>.70</b>	<b>12</b>
1+2	43	.61	20
3p inanimate	43	.54	15
3s	23	.31	44
Range		49	

$p < .05$ . Log likelihood = -69.610; Total Chi-square = 90.8580; Chi-square/cell = 1.1357  
Not significant: Clause type, Sentence type/polarity, Presence of a clitic, Presence of a locative, Temporal adverbial presence and type.

just as in the previous case, significant differences in distribution are found both in the 18th–19th centuries, and in the 20th century. Again, we will see that these differences are not semantic, since they’re not necessarily a salient element of meaning to speakers, or even an element that can be brought about upon reflection by linguists interested in the matter. They reflect the subjectification and routinization in certain contexts of the most frequent form, *todavía*.

Table 5 shows the results for the 18th and 19th-century data set. Two factor groups are shown to be significant here: Verb class and Grammatical person. In the case of Verb class, we see that while *aún* is preferred in durative and possessive contexts, as in (53) and (54), it is less preferred in all other stative contexts, like those in (55) and (56).

- (53)

*Sólo mi corazón **aún permanece** cubierto de densas y espantosas tinieblas.*  
‘Only my heart **still remains** covered in dense and frightening shadows’  
(CORDE)
- (54)

*Esta provincia **aún conserva** cierto orgullo nacido de su antigua grandeza*  
‘This province **still conserves** a certain pride borne of its ancient grandeur’  
(CORDE)

- (55) *no desperdiciemos las pocas fuerzas que **aún** nos **quedan** . . .*

‘let’s not waste the bit of strength that we **still have left** [lit. are left to us]’  
(CORDE)

- (56) *Arraigó el tronco en su alma la compasión y la humanidad, y si todavía **quedan** en ella resabios de vanidad*

‘The trunk of compassion and humanity rooted itself in his soul, and if there **are still** traces of vanity **left** in it’  
(CORDE)

At first, the verb class results make little intuitive sense. Why would a form be significantly favored in a diverse subset of stative verbs (i.e., durative and possessive), yet much less favored with other, semantically similar verbs (i.e., other stative verbs)? The answer lies in a closer look at the data: in this case, we did not find semantic conditioning according to verb class, but rather lexical conditioning combined with semantic conditioning. The semantic contexts of durativity and possession are favorable to the use of *aún*, but two highly frequent stative verbs outside these specific contexts, namely, *ser* ‘be’ and *haber* ‘exist’ do not occur at all with *aún* in this data set. Of the 13 occurrences of *ser* and 4 of *haber*, all were with *todavía*, as shown in (57) and (58).

- (57) ***Sois** todavía hijo de familia*

‘You **are still** a son of [a well-bred] family’  
(CORDE)

- (58) *¿No hay remedio, no **hay** remedio todavía?*

‘There is no remedy, **there is still** no remedy?’  
(CORDE)

The other factor group shown to be statistically significant for the early data set was grammatical person. Two contexts are highly favorable to *aún*: third-person plural animate subjects (0.80) and impersonal subjects (0.70). Together, these two contexts account for only 17% of the data. First- and second-person (0.61), along with third-person inanimate plural subject contexts (0.54), are slightly favorable, and the only context disfavorable to *aún* is also the most frequent: third-person singular (0.31), which comprises 44% of the data and is used with *aún* only 23% of the time. The contexts most favorable to *aún*, in this case, are also those most likely to refer to a situation outside the immediate personal context of the writer/reader, such as the actions of an outside group (in the case of third-person plural animate) or people’s general behavior (in the case of impersonal).



**Table 6:** Variable Rule Analyses, *aún* + P v. *todavía* + P, 20th century. Input probability: 0.417 (42%), N = 656/1564

Factor Group	Rel. frequency	Probability	% of data
<b>Grammatical person</b>			
Official body	86	.88	2
1s+2s	46	.54	13
3s + inanimate	41	.49	71
Plural animate	36	.41	14
Range		47	
<b>Temporal adverbial</b>			
Other	50	.57	1
None	42	.51	97
<i>hoy, ahora</i>	16	.20	2
Range		37	
<b>Verb class</b>			
Other stative + dur.	49	.57	31
Psych, percep., modal	47	.57	26
<i>Estar</i>	36	.42	11
Activity	33	.41	11
<i>Ser, haber</i>	30	.37	16
Other	29	.36	2
Range		21	
<b>Register</b>			
Press	46	.54	40
Literature	39	.47	60
Range		7	
<b>Clitic</b>			
Absent	42	.52	77
Present	41	.45	23
Range		7	

$p < .05$ . Log likelihood = -1015.218; Total Chi-square = 241.5682; Chi-square/cell = 0.9744  
 Not significant: Clause type, Sentence type/polarity, Presence of a locative.

For the 18th and 19th-century data set, then, we have found, on the one hand, a tendency for *aún* to occur in durative and possessive contexts referring to situations outside the writer's personal realm, and on the other, lexical conditioning toward non-occurrence of *aún* with *ser* and *haber*.

The 20th-century results, shown in Table 6, tell a similar story. Both factors shown to be significant in the earlier data set are still significant in the 20th-century data set. This time, however, verb class is no longer the strongest factor group; grammatical person now has the strongest magnitude of effect, with a range of 47, while verb class is third, with a range of 21. Though the favorable

grammatical person contexts for *aún* were limited to the smallest groups in the earlier data set, here the favorable contextual environment is even smaller, with official bodies (0.88) as in (59) representing only 2% of the data. All other grammatical persons hover right around 0.5.

- (59) *Zatopek no entiende por qué **España aún** duda en participar en Moscú*  
 ‘Zatopek does not understand why **Spain still** has doubts about participating in Moscow’  
 (CREA)

The lexical conditioning for verbs found in the earlier data set is found again, and even more clearly this time, in the 20th century. This conditioning, however, is no longer categorical, as shown in (60). The three most highly frequent stative verbs, *ser* ‘be’ and *haber* ‘exist’ (0.37 combined), and *estar* ‘be’ (0.42), are disfavorable to *aún*, along with activity verbs (0.41).

- (60) *pues **aún hay** otro libro inédito y sin título*  
 ‘well **there is still** another unpublished and untitled book’  
 (CREA)

In this data set, other significant factor groups also emerged. Temporal adverbials (Range = 37), in fact, the lexical items *hoy* and *ahora*, as in (61) and (62), are highly unlikely to occur with *aún* (0.20). Here, again, then, we have a case of lexical conditioning.

- (61) *su nueva versión de eurocomunistas, que es **aún hoy** una pura palabra sin ningún ejemplo práctico*  
 ‘his new version of Eurocommunists, which **still today** is just a word without any practical example’  
 (CREA)
- (62) *de lo que se trata es de transformar lo que **hoy** es todavía para muchos una expresión esotérica*  
 ‘what it’s about is transforming what **today** is still for many an esoteric expression’  
 (CREA)

In (61), we can see how certain lexical combinations may have encouraged the pragmatic inferences that link the different meanings ascribed to the polysemous *aún/aun* and *todavía*; in (61), the collocation *aún hoy* ‘still today’ can also (but not

only) be taken to mean ‘even today’. In (62), however, this is not the case, as there is intervening material between *hoy* ‘today’ and *todavía* ‘still’. However, the same such reading is possible for *todavía* in combination with *hoy* ‘today’, as seen in (63). Such semantic extension of these adverbials appears to be part of their shared path of grammaticalization; as such, these cannot account for the rise in frequency of one adverbial form over another.

(63) *pero conserva todavía hoy la intensidad febril de sus ojos de hielo*  
‘but she retains still/even today the feverish intensity of her icy eyes’  
(CREA)

These three factor groups (grammatical person, temporal adverbials, and verb class) all point toward lexically constrained *aún* associated with distant, impersonal, general states (official bodies, not ‘today’ or ‘now’, and not activities). The final two factor groups to show a significant, though weak, effect were register and the presence of a clitic, both with a Range of 7. The less frequent form, *aún*, is slightly favored in the press and when no clitic is present.

3.4 General tendencies: Genre differences or typological shift

Figure 4 shows the relative frequencies of each variant by data set (already given in the Introduction). What is interesting is that, although *seguir* constructions

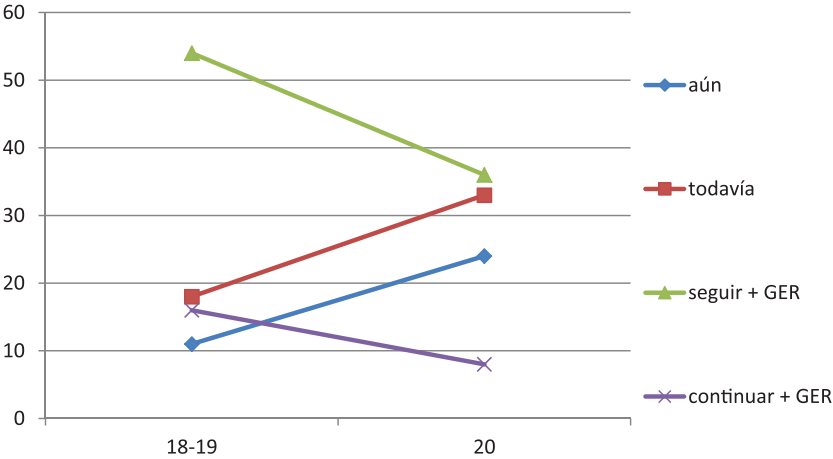


Fig. 4: Relative frequencies by data set, percentages.

and *todavía* are the most frequent in both data sets, this is not true to the same extent for each. In fact, *seguir* constructions lose considerable ground. But this “loss” is not to the syntactically similar *continuar* constructions, but rather to the phasal adverbs, both of which show a gain in relative frequency in the 20th-century data.

These results show a general increase in relative frequencies of continuative adverbials, and a concurrent general decrease in relative frequencies of continuative semi-auxiliary constructions. It appears that adverbial expression of progressive aspect in this domain is increasingly preferred over the periphrastic constructions. This is particularly fascinating given the pan-Hispanic increase in frequency other scholars have found for the *estar* + *Vndo* construction. We may have here, then, an example of what Sapir (1921) calls “drift”, i.e., the tendency for languages to shift typologically over time (cf. Greenberg 1960):

The linguistic drift has direction. In other words, only those individual variations embody it or carry it which move in a certain direction, just as only certain wave movements in the bay outline the tide. The drift of a language is constituted by the unconscious selection on the part of its speakers of those individual variations that are cumulative in some special direction. This direction may be inferred, in the main, from the past history of the language.

While such results are both exciting and suggestive, we must not forget the inescapable fact that the nature of the two data sets is different. This means that quantitative evidence about frequency alone is not enough to make strong claims about the nature of the grammar of a given language at a given moment in time (Rinke and Elsig 2010). The deeper, more detailed analyses provided through variable rule analyses are, in the end, more revealing. Whether the “change” in Figure 4 represents, in fact, a real change, or simply a genre shift in the data, must be left to further investigation.

## 4 Discussion and conclusions

The quantitative analysis of four (or six, if periphrastic negative polarity constructions are counted separately) variants of variable continuative aspect marking in Peninsular Spanish has revealed a complex, multi-layered picture of language variation and change, in which syntactic constraints, typological tendencies, and frequency-based diachronic processes of change present a nuanced system of variation that has, until now, been invisible to scholars who have only worked with qualitative analyses. In terms of syntax, we found that speakers prefer *todavía* over the more marked *seguir* + *Vndo* in more marked contexts, since it

is favored in subordinate clauses, with negative polarity and when a clitic is present. We also found that overall relative frequencies over the past 200 years indicate a relative rise in adverbial constructions at the expense of periphrastic constructions. This may point to a general change in Spanish typology, or may simply reflect genre differences between data sets.

Alongside syntactic and typological differences, we also found significant differences between purported synonyms that appear to be due, on the one hand, to being at different stages in the process of grammaticalization (in the case of the periphrastic constructions), and on the other, to the absorption of frequently occurring linguistic contexts in the most frequent constructions, leading to what appears to be an incipient subjectification. Evidence for the relative newness of *seguir* + *Vndo* as a grammaticalizing construction marking continuative aspect was found in the semantic persistence of older motion meanings through the 19th century. At the same time, we find that the less frequent constructions *aún* and *continuar* + *Vndo* are associated with less personal contexts. Since it is the more frequent forms that would be more likely to occur in casual speech (Biber 1995), it makes sense that these forms would be the ones that occur most often with more personal contexts. This accident of frequency has begun to make its way into the grammar, revealing a correlation between register variation and the probability of the subjectification of a given construction.

## Data sources

CORDE = *Corpus diacrónico del español*. Real Academia Española. <http://corpus.rae.es/cordenet.html>

CREA = *Corpus de referencia del español actual*. Real Academia Española. <http://corpus.rae.es/creanet.html>

Real Academia Española. 1729. DRAE = RAE A 1729. *Diccionario de la lengua castellana, en que se explica el verdadero sentido de las voces, su naturaleza y calidad, con las frases o modos de hablar, los proverbios o refranes, y otras cosas convenientes al uso de la lengua* [...]. *Compuesto por la Real Academia Española. Tomo segundo. Que contiene la letra C*. Madrid: Imprenta de Francisco del Hierro. Reproducido a partir del ejemplar de la Biblioteca de la Real Academia Española. <http://buscon.rae.es/ntlle/SrvltGUILoginNtlle>. (Accessed on 10 September 2010)

Real Academia Española. 1739. DRAE = RAE A 1739. *Diccionario de la lengua castellana, en que se explica el verdadero sentido de las voces, su naturaleza y calidad, con las frases o modos de hablar, los proverbios o refranes, y otras cosas convenientes al uso de la lengua* [...]. *Compuesto por la Real Academia*

*Española. Tomo sexto. Que contiene las letras S.T.V.X.Y.Z.* Madrid: Imprenta de la Real Academia Española, por los herederos de Francisco del Hierro. Reproducido a partir del ejemplar de la Biblioteca de la Real Academia Española. <http://buscon.rae.es/ntlle/SrvltGUILoginNtlle>. (Accessed on 10 September 2010)

## References

- Aaron, Jessi Elana. 2004. The gendered use of *salirse* in Mexican Spanish: *Si me salía yo con las amigas, se enojaba*. *Language in Society* 33. 585–607.
- Aaron, Jessi Elana. 2010a. An awkward companion: Disability and the semantic landscape of English *lame*. *Journal of English Linguistics* 38(1). 35–55.
- Aaron, Jessi Elana. 2010b. Pushing the envelope: Looking beyond the variable context. *Language Variation and Change* 22. 1–36.
- Aaron, Jessi Elana & Rena Torres Cacoullos. 2005. Quantitative measures of subjectification: A variationist study of Spanish *salir(se)*. *Cognitive Linguistics* 16(4). 607–633.
- Alimhilli, Gjilda. 2001. L'Aktionsart nella lingua albanese. L'azione verbale continuativa e stativa. *Studi Italiani di Linguistica Teorica e Applicata* 30(2). 371–395.
- Amenta, Luisa. 1994–1995. *La perifrasi aspettuale stare + gerundio*. Palermo: University of Palermo dissertation.
- van der Auwera, J. 1998. Phasal adverbials in the languages of Europe. In Johan van der Auwera & Dónall P. Ó Baoill (eds.), *Adverbial constructions in the languages of Europe*, 125–145. Berlin & New York: Mouton de Gruyter.
- Azofra Sierra, María Elena. 2008. Origen y valores de formas nominales en *-ndo* en español. *Verba* 35. 35–54.
- Barlow, Michael, & Suzanne Kemmer. 2000. *Usage-based models for language*. Stanford, CA: Center for the Study of Language and Information.
- Bello, Andrés & Rufino J. Cuervo. 1973. *Gramática de la lengua castellana*, with commentary by Niceto Alcalá-Zamora y Torre, 9th edn. Buenos Aires: Editorial Sopena Argentina.
- Benveniste, Émile. 1966. De la subjectivité dans le langage. In Émile Benveniste (ed.), *Problèmes de linguistique générale*, 258–266. Paris: Gallimard.
- Bertinetto, Pier Marco. 2000. The progressive in Romance, as compared with English. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, 559–604. Berlin & New York: Mouton de Gruyter.
- Biber, Douglas. 1995. *Dimensions of register variation*. Cambridge: Cambridge University Press.
- Böckle, Klaus. 1980. Zum Diasystem der portugiesischen Verbalperiphrasen mit dem "gerundialen Infinitif". *Zeitschrift für Romanische Philologie* 96. 333–354.
- Brianti, Giovanna. 1992. *Périphrases aspectuelles de l'italien: Le cas de andare, venire et stare + gérondif*. Bern: Lang.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan, Revere Perkins, & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.
- Caro, Miguel Antonio. 1976 [1870]. *Tratado del participio*. Bogotá: Instituto Caro y Cuervo.

- de Carvalho, Paulo. 2003. "Gérondif", "participe présent" et "adjectif déverbal" en morphosyntaxe comparative. *Langages* 37(149). 100–126.
- Claveres, Marie-Hélène. 1997. Requiem pour le présent progressif. *Les Langues Modernes* 9. 37–45.
- Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
- Comrie, Bernard. 1990. The typology of tense–aspect systems in European languages. *Lingua e Stile* 25.259–272.
- Cortés-Torres, Mayra. 2005. ¿Qué estás haciendo?: La variación de la perífrasis *estar + -ndo* en el español puertorriqueño. In David Eddington (ed.), *Selected proceedings of the 7th Hispanic Linguistics Symposium*, 42–55. Somerville, MA: Cascadilla Proceedings Project.
- Dias da Costa, Albano. 1976. Periphrastic verbal expressions in Portuguese. In Jurgen Schmidt-Radefeldt (ed.), *Readings in Portuguese linguistics*, 187–243. Amsterdam: North-Holland.
- Dominicy, Marc. 1977. Les périphrases verbales des langues romanes. *Revue Belge de Philologie et d'Histoire* 55(3). 914–934.
- Fente Gómez, Rafael, Jesús Fernández Alvarez & Lope G. Feijóo. 1972. *Perífrasis verbales*. Madrid: Sociedad General Española de Librería.
- Fernández de Castro, Félix. 1990. *Las perífrasis verbales en español: Comportamiento sintáctico e historia de su caracterización*. Oviedo: Departamento de Filología Española.
- Fitzgerald, Colleen M. 2004. Language change and motion verbs in Tohono O'odham. Paper presented at LASSO XXXIII, New Orleans, September 10–12. <http://www.uta.edu/faculty/cmfitz/2004lassohandout.pdf> (accessed 1 July 2013).
- García, Christina. 2011. Distinguishing two "synonyms": A variationist analysis of *quizá* and *quizás* in six Spanish dialects. In Jim Michnowicz & Robin Dodsworth (eds.), *Selected proceedings of the 5th Workshop on Spanish Sociolinguistics*, 103–112. Somerville, MA: Cascadilla Proceedings Project. <http://www.lingref.com/cpp/wss/5/paper2510.pdf> (accessed 1 July 2013).
- García González, Javier. 1992. *Perífrasis verbales*. Madrid: SGEL.
- García Fernández, Luis & María Martínez-Atienza. 2003. La expresión de los eventos inconclusos en español. *Revista Española de Lingüística* 33(1). 29–67.
- Garrido, Joaquín. 1992. Expectations in Spanish and German adverbs of change. *Folia Linguistica* 26(3/4). 357–402.
- Girard, M. l'Abbé. 1718. *La justesse de la langue françoise ou les différentes significations des mots qui passent pour synonymes*. Paris: Laurent d'Houry. <http://gallica.bnf.fr/ark:/12148/bpt6k50626b> (accessed 1 July 2013).
- Gómez Torrego, Leonardo. 1988. *Perífrasis verbales*. Madrid: Arco Libros.
- Górbova, Elena V. 2000. Acerca del problema del potencial semántico del progresivo en español. *Language Design: Journal of Theoretical and Experimental Linguistics* 3. 23–35.
- Gosselin, Laurent. 1996. *Sémantique de la temporalité en français: Un modèle calculatoire et cognitif de temps et de l'aspect*. Louvain-la-Neuve: Duculot. <http://lidifra.free.fr/?q=node/668> (accessed 1 July 2013).
- Gougenheim, Georges. 1929. *Étude sur les périphrases verbales de la langue française*. Paris: Nizet.
- Greenberg, Joseph H. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26. 178–194.
- Gross, Gaston. 1998. Pour une véritable fonction « synonymie » dans un traitement de texte. *Langages* 32(131). 103–114. <http://www.persee.fr> (accessed 1 July 2013).

- Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42. 25–70. doi: 10.1017/S0022226705003683.
- Hopper, Paul & Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge, UK: Cambridge University Press.
- Jakobson, Roman. 1960. Concluding statement: Linguistics and poetics. In Thomas Albert Sebeok (ed.), *Style in language*, 350–377. New York: John Wiley and Sons.
- König, Ekkehard. 1991. *The meaning of focus particles: A comparative perspective*. London: Routledge.
- Kracht, Marcus. 2007. The emergence of syntactic structure. *Linguistics and Philosophy* 30(1). 47–95. doi: 10.1007/s10988-006-9011-5.
- Laca, Brenda. 2000. Les périphrases de phase en catalan contemporain. In Bruno Staib & Wolf Dietrich (eds.), *Linguistica Romanica et Indiana*, 249–269. Tübingen: Narr.
- Langacker, Ronald W. 1990. Subjectification. *Cognitive Linguistics* 1. 5–38.
- Lausberg, Heinrich. 1966. *Lingüística románica*. Madrid: Gredos.
- Lipsky, Angela. 2003. Pour une description sémantique et morpho-syntaxique du participe français et allemand. *Langages* 37(149). 71–85.
- Löbner, Sebastian. 1989. German *schon – erst – noch*: An integrated analysis. *Linguistics and Philosophy* 12. 167–212. <http://docserv.uni-duesseldorf.de/servlets/DerivateServlet/Derivate-12756/S0089415.pdf> (accessed 1 July 2013).
- Lyer, Stanislav. 1934. *Syntaxe du gérondif et du participe présent dans les langues romanes*. Paris: Droz.
- Mautner, Gerlinde. 2007. Mining large corpora for social information: The case of *elderly*. *Language in Society* 36(1). 51–72.
- Meillet, Antoine. 1948 [1920]. *Linguistique historique et linguistique générale*. Paris: Champion.
- Miguel Aparicio, Elena de. 1999. El aspecto léxico. In *Gramática descriptiva de la lengua española*, 2977–3060. Madrid: Espasa.
- Mosegaard Hansen, Maj-Britt. 2002. La polysémie de l’adverbe *encore*. *Travaux de Linguistique* 44. 143–166.
- Muller, Claude. 1999. Encore et toujours les modifieurs aspectuels: De *encore* à *toujours*. In Marc Plénat, Michel Aurnague, Anne Condamines, Jean-Pierre Maurel, Christian Molinier & Claude Muller (eds.), *L’emprise du sens: Structures linguistiques et interprétations: mélanges de syntaxe et de sémantique offerts à André Borillo*, 217–237. Amsterdam: Rodopi.
- Newman, John. 2002. A cross-linguistic overview of the posture verbs ‘sit’, ‘stand’, and ‘lie’. In John Newman (ed.), *The linguistics of sitting, standing and lying* (Typological Studies in Language 51), 1–24. Amsterdam & Philadelphia: John Benjamins.
- Noetzel, Lisa M. 2007. The Spanish progressive construction and its Latin source. *Neophilologus* 91(1). 51–61. doi: 10.1007/s11061-006-9006-z.
- Olbertz, Hella. 1998. *Verbal periphrasis in a functional grammar of Spanish*. New York: Mouton de Gruyter.
- Orpin, Debbie. 2005. Corpus linguistics and critical discourse analysis: Examining the ideology of sleaze. *International Journal of Corpus Linguistics* 10(1). 37–61.
- Ozete, Oscar. 1983. On the so-called Spanish gerund/participle. *Hispania* 66(1). 75–83.
- Penny, Ralph. 1991. *A history of the Spanish language*. Cambridge: Cambridge University Press.
- Poplack, Shana & Danielle Turpin. 1999. Does the *Futur* have a future in (Canadian) French? *Probus* 11(1). 133–164.



- Pustet, Regina. 2008. The Lakota aspect/modality markers *-kinica* and *tkhá*. In Werner Abraham & Elisabeth Leiss (eds.), *Modality–aspect interfaces: Implications and typological solutions*, 331–355. Amsterdam & Philadelphia: John Benjamins.
- Rinke, Ester & Martin Elsig. 2010. Quantitative evidence and diachronic syntax. *Lingua* 120(11). 2557–2568. doi: 10.1016/j.lingua.2010.06.012
- Rojo, Guillermo. 1974. *Perífrasis verbales en el gallego actual: Verba*: Annex 2. Santiago de Compostela: Universidad de Santiago de Compostela.
- Rowicka, Graxyna J. 2006. Pronominal markers in Quinault (Salish). *International Journal of Applied Linguistics* 72(4). 451–477.
- Sanchez, Tara. 2002. The interacting influences of Spanish and English on the creole Papiamentu. In Daniel Ezra Johnson & Tara Sanchez (eds.), *Selected Papers from NNAV 30, Raleigh, NC* (University of Pennsylvania working papers in linguistics 8.3), 235–247. Philadelphia, PA: Penn Graduate Linguistics Society.
- Sankoff, David. 1988. Sociolinguistics and syntactic variation. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge Survey*, 140–161. Cambridge: Cambridge University Press.
- Sapir, Edward. 1921. *Language*. New York: Harcourt Brace and World. <http://www.bartleby.com/186/> (accessed 15 September 2013).
- Schaefer, Ronald P. & Francis O. Egbokhare. 2008. A preliminary assessment of Emai posture verb parameters. *Journal of African Languages and Linguistics* 29(2). 215–235. doi: 10.1515/JALL.2008.010.
- Scheibman, Joanne. 2002. *Point of view and grammar: Structural patterns of subjectivity in American English conversation*. Amsterdam & Philadelphia: John Benjamins.
- Scheffer, Johannes. 1975. *The progressive in English*. Amsterdam: North-Holland.
- Sinclair, John M. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Squartini, Mario. 1998. *Verbal periphrases in Romance: Aspect, actionality, and grammaticalization*. Berlin & New York: Mouton de Gruyter.
- Torres Cacoullos, Rena. 1999. Construction frequency and reductive change: Diachronic and register variation in Spanish clitic climbing. *Language Variation and Change* 11(2). 143–170. doi: 10.1017/S095439459911202X.
- Torres Cacoullos, Rena. 2000. *Grammaticalization, synchronic variation and language contact: A study of Spanish progressive -ndo constructions*. Amsterdam & Philadelphia: John Benjamins.
- Torres Cacoullos, Rena. 2001. From lexical to grammatical to social meaning. *Language in Society* 30. 443–478.
- Torres Cacoullos, Rena. 2006. Relative frequency in the grammaticization of collocations: Nominal to concessive *a pesar de*. In Timothy L. Face & Carol A. Klee (eds.), *Selected proceedings of the 8th Hispanic Linguistics Symposium*, 37–49. Somerville, MA: Cascadia Proceedings Project.
- Traugott, Elizabeth Closs. 1995. Subjectification in grammaticalization. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectivisation*, 31–54. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs. 2003. Constructions in grammaticalization. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 624–647. Oxford: Blackwell.
- Traugott, Elizabeth Closs & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.

- Victorri, Bernard & Catherine Fuchs. 1992. Construire un espace sémantique pour représenter la polysémie d'un marqueur grammatical: L'exemple de *encore*. *Lingvisticae Investigationes* 16(1). 125–153.
- Weinreich, Uriel, William Labov & Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In Winfred P. Lehmann and Yakov Malkiel (eds.), *Directions for historical linguistics*, 95–195. Austin: University of Texas Press. <http://www.utexas.edu/cola/centers/lrc/books/hist05.html> (accessed 1 July 2013).
- Wu, Yunji. 2005. *A synchronic and diachronic study of the grammar of the Chinese Xiang dialects*. Berlin & New York: Mouton de Gruyter.

## Appendix

**Table A:** Number of occurrences by verb class

Dataset	XVIII–XIX	XX
Accomp./achiev.	59	151
Activity	78	175
Durative	8	73
<i>Estar</i>	9	193
<i>Gustar</i> -type	3	18
<i>Haber</i>	5	55
Intransitive motion	33	104
Modal	7	193
Other stative	37	339
Perception	8	54
Possession	36	207
Psychological	39	324
Verbs of saying	44	173
<i>Ser</i>	35	497
Verbs of transfer	18	43
Transitive displacement	30	69
Other	32	94
Total	481	2,762