

# At the boundaries of linguistic convergence: variation in presentational *haber / haver-hi*. A sociolinguistic comparative analysis of Spanish and Catalan grammars<sup>1</sup>

José Luis Blas Arroyo

Universitat Jaume I

In this paper, I focus on an eventual convergence outcome (the pluralization of presentational *haber/haver-hi*) in the grammar of two Romance languages, Spanish and Catalan, which have been living side-by-side for centuries in Eastern Spain. Taking into account the sociolinguistic comparative method and on the basis of several representative corpora of the two languages in contact, the data from this research offer evidence that points to a notable congruence between the underlying grammars of both languages, which would, at least partially, account for a similar diffusion of these vernacular pluralizations. Moreover, some of the few cases of disagreement found can be explained on the basis of both internal (such as the existence of points of structural conflict in some verbal paradigms) and external factors (such as hypercorrection), which certain social groups, particularly sensitive to normative pressure, are more receptive to.

**Keywords:** Presentational *haber / haver-hi*, convergence, comparative sociolinguistics, languages in contact, Catalan/Spanish

## 1. Introduction

The normative grammars of both Spanish and Catalan consider (1) and (2) below to be impersonal sentences. In both cases, the argument dependent on the verb *haber* is interpreted as the direct object, since it could be replaced by an accusative pronoun. Yet, it is known that in a number of speech communities, fragments like (3) and (4) are as, or more, frequently used by speakers. In these examples, the verbs *haber* and *haver-hi*

---

<sup>1</sup> This study is part of the project “Structural, social and idiolectal dimensions of linguistic change: new contributions from historical sociolinguistics to the study of Spanish” (2018-2021), funded by the Universitat Jaume I and the Spanish Ministry of Economy and Competitiveness (Ref. Ref. FFI2017-86194-P). I would like to thank Carme Barberà, Susana Martínez, Maria Chiara Marullo and Jordi Ayza for their collaboration in various tasks related to the project. I also thank two anonymous reviewers for their valuable comments, which have helped to improve the article. Any remaining errors are my sole responsibility.

agree with the adjacent noun in number and person, which act as though they were the subject:

- (1) ... gustara más que el que hago, pues probablemente *habría* posibilidades de que pudiera cambiar. (MCSCS, 417)  
“... liked more than what I do, then there would probably be the possibility to change.”
- (2) ... i ni *ha* coses, no, en el- la vida la fem conforme anem // a anem avançant. (MCSCS, 398)  
“... and there are things, no, in the, in life that we do as we go // as we go along.”
- (3) ... era el año cincuenta y tres/ pocos teléfonos *habrían* en Almazora pues. (MCSCS, 417)  
“... it was the year nineteen fifty-three/ there couldn't have been many telephones in Almazora then.”
- (4) ... a pesar d'això n'hi *han* coses d'atres persones que fas que sempre et venen al cap de dir... (MCSCS, 398)  
“... in spite of that there are other people's things that you do that always come into your mind to say...”

The linguistic alternation between the agreed and non-agreed variants of these verbs is one of the most conspicuous examples of syntactic variation in the two languages. In the case of Spanish, sociolinguistics has reported on the scope reached by pluralizations in different Hispanic regions, together with some structural and non-structural conditioning factors (Bentivoglio and Sedano, 1989; Montes de Oca, 1994; Díaz-Campos, 2003; D'Aquino, 2004; Hernández Díaz, 2006; Castillo Trelles, 2007; Freites, 2008; Rivas and Brown, 2012; Claes, 2014; Lastra and Martín Butragueño, 2016; Pato, 2016; Bouzouita and Pato, 2016, among others). A review of this literature, however, reveals that, albeit with some exceptions, most of this research has been conducted in America, with far fewer studies carried out on European Spanish.

There may be a number of reasons to account for this imbalance, such as the late arrival of sociolinguistics in Spain, but probably there is a far more prosaic one among them that cannot be ruled out: within the European Spanish speech communities as a whole, there is considerably less variation. Indeed, from the material available in some recent Spanish sociolinguistic corpora (see Section 4.2), it is easy to see how the pluralized variants hardly ever appear in some dialects, in clear contrast to those of America. Nevertheless, this panorama offers a couple of notable exceptions: the Canary Islands and several regions in the Catalan-speaking territories on the peninsula (essentially Catalonia and the Valencian Region) (Samper and Hernández, 2012; Blas Arroyo, 2016; Pato, 2016; Bouzouita and Pato, 2016; Claes, in press).

With regard to the Catalan-speaking communities, which are the focus of interest in this study, some recent research has drawn attention to the widespread diffusion that this vernacular phenomenon achieves in them (Blas Arroyo, 1996, 2016; Gómez Molina, 2013). The most strikingly singular aspect of these results lies in the fact that they are also widely produced in the other commonly used language in the society, Catalan, with which Spanish has coexisted for centuries (Fabra, 1956; IEC, 1992; Sanchis Guarner, 1993; Solà, 1994; Rigau, 1993; Badia i Margarit, 1994; Casals and Cervià, 1999; Ramos Alfajarín, 2001; etc.).

Is it therefore possible that the two languages could have provided each other with feedback on the social diffusion of this vernacular phenomenon? If this were the case, how can such an influence be conceptualized? And how can it be accounted for?

Regarding this last question, Poplack *et al.* (2012: 205) remind us that, in order to confirm whether a particular linguistic change phenomenon has its origin in the contact between languages, it is not sufficient to simply verify the existence of vernacular uses that are more or less extended throughout the speech community. On the contrary, it is necessary to provide empirical evidence of certain requirements, such as the following:

- (i) it must first be clear that the feature in question is in fact an innovation in the host language, (ii) that it was not present in the pre-contact variety, (iii) that it is not present in a contemporaneous non-contact variety, (iv) that it behaves in the same way as its putatively borrowed counterpart in the source language, and (v) that it differs in non-trivial ways from superficially similar constructions in the host language, if any.

Within this same heuristic framework, in this study we perform a comparative analysis of the patterns of variation that Spanish and Catalan present in their respective speech communities. To do so, we analyse the results of two independent variationist studies based on two corpora that are sufficiently comprehensive and representative of each community. In particular, our aims in these pages consist in:

- a) observing the scope of pluralizations in Spanish and comparing them with that observed in other Hispanic speech communities, as well as in Catalan;
- b) exploring, also from a contrastive perspective, the envelope of variation in which this variation is used, with the aim of determining which points in the system, together with those on the socio-stylistic axis, favour or, conversely, hinder agreed variants in both languages;
- c) determining whether any tendencies of linguistic change can be observed and, if so, in which direction they run in the two languages.
- d) To do so, we employ a number of different testing media, which will be detailed later on (see Section 3.1).

## 2. The singularity of haber /haver-hi in syntax

### 2.1 The linguistic variable in Spanish

As Hernández Díaz (2006: 1147-1148) points out, recurring pluralizations prove that traditional interpretations of the argument as the direct object do not fit in with the syntactic-semantic characteristics of this verb in presentational contexts. On the contrary, *haber* presents a dissociation between the form and the semantics that is similar to that at play in the unaccusative verbs (*exist*, *float*, *occur*, etc.). In fact, all these verbs take a single participant, which is never an agent, but rather an experiencer or theme. In practice, *haber*'s pluralization comes as the solution to an obvious anomaly in Spanish syntax, in which the only argument dependent upon a verb (as in intransitive sentences) is usually coded as the subject (Blas Arroyo, 1996; Hernández Díaz, 2006: 1149).

All this is the result of the evolution of *haber* from Latin to Romance, as the verb lost its original possessive values in favour of *tener* in mediaeval Spanish (Garachana, 1997). At the same time, presentational *haber* participates in two important syntactic changes in history. As pointed out by Hernández Díaz (2006: 1124), the first of them concerns the introduction of a locative suffix (*ibi* → *y*) in the inflection of the present indicative (*hay*), a fact that could account for the delay in the spread of a second change, the “generalization of agreement”, to characterize the cases of direct (*habían flores*) and

indirect pluralization (*puede que hayan flores*).<sup>2</sup> Indeed, the agreement forms with the present indicative (*hayn, haen, etc.*) are still very marginal in the Spanish-speaking world, unlike what happens in Catalan (see Section 2.2).

The literature on this phenomenon was scarce until recent times, which suggests that the agreement must have been a stigmatized phenomenon, if only in the written language. Nonetheless, some scholars have proved the existence of examples of pluralization back in the 16th and 17th centuries, and on a more recurring basis as of the 18th century in Latin American countries like Argentina (Fontanella, 1992: 70). Quintanilla (2009) also found cases of plural agreement in journalistic texts in El Salvador from 1821 onwards, which gives the idea that, at that time, agreement must have been widespread, judging by the recriminations by grammarians such as Bello (1981 [1847]: 467), who drew attention to the “almost universal bad habit” of pluralization in Chile.

## 2.2 The linguistic variable in Catalan

As in Spanish, the normative grammar of Catalan tends to consider sentences formed with *haver-hi* as impersonal (Fabra, 1956; Moll, 1968; IEC, 1992; Sanchis Guarner, 1993; Badia i Margarit, 1994). Rigau (1993) endorses the impersonal nature of these constructions as a consequence of the grammaticalization of the locative pronoun *hi*, which, attached to the different forms of *haver*, turns it into a transitive verb, accompanied by a direct object or an internal argument of the predication. This is regardless of the fact that, in some dialects, this argument “may look like a subject, since it agrees with the verb” (Rigau, 1993: 39), as occurs in Central Catalan or in Valencian. In these cases, the agreement only affects the number, but not the grammatical person, which leads to such impersonal sentences as examples (2) and (4) cited earlier.

Being aware of how widespread the phenomenon of agreement actually is in these varieties, even the most conspicuous representatives of linguistic prescriptivism have displayed some degree of flexibility when it comes to evaluating these usages, by accepting them in certain registers. Thus, in his reprint of the *Gramàtica de la llengua catalana* (1956), Pompeu Fabra was far more benevolent towards agreement than he had been several decades before, when he had condemned their use outright. In this regard, he acknowledged that:

... these constructions, commonly used in the spoken language, have always been considered incorrect, *but perhaps one day they should be admitted into the written language, in accordance, of course, with the traditional constructions* (our italics).

Badia i Margarit (1994: 300-302), who was nearly always against any hint of pluralization, later appeared to be more tolerant regarding its use, although he restricted it to informal language. In the same line, a Valencian grammar published recently by the Universitat de València (2011) (*Gramàtica Zero*) remarks that, while the singular is the only form that should be used in formal and written language, in colloquial contexts “many speech varieties in Valencia and Catalonia make the singular or plural form of the verb *haver-hi* agree with the thing that there is”. More indulgent still, the grammar of the *Acadèmia Valenciana de la Llengua* (2006: 305) has even considered the agreed

---

<sup>2</sup> Hernández Díaz (2006: 1150) uses the concept “generation of indirect concordance” (*generación de concordancia indirecta*) to refer to those pluralizations that occur in the auxiliary verbs of periphrases that can select *haber* as the main verb.

usages as acceptable (“Both the lack of agreement and agreement are acceptable”), although immediately afterwards it goes on to warn that “in formal registers the lack of agreement is considered more *appropriate* [but therefore not necessarily the only correct form]” (both the italics and the comment in square brackets are our own).

Nevertheless, some authors have gone beyond these limits and have clearly defended the grammaticality – in the normative sense – of these pluralized variants. Probably the best known was Joan Solà (1994: 21-26), who noted that the opposition to these variants among Catalan grammarians has its origin in a mistaken normative criterion that has continued since the late 19th century, when the Spanish grammars published in Catalonia condemned the agreement that “Catalans made when they spoke Spanish” (Solà, 1994: 23). For him, the pluralized forms not only vary considerably from one dialect to another – which therefore rules out the idea of defending some as being more grammatical than others – but also display an obvious structural justification and, accordingly, he considers it preferable “to inflect the verb in the plural in constructions such as *Hi han quatre hom*”. At the same time, Solà believes that the influence of the mass media, and more especially that of television, is destined to become a powerful mechanism for diffusing the phenomenon, even in dialectal areas where it is traditionally far less characteristic. Indeed, the phenomenon of plural agreement displays a notable degree of disparity from one variety to another. Thus, pluralizations are not common in dialects such as the Catalan of the north-western territories, Rousillon, Alguer and the Balearic Islands. Conversely, these variants are notably more widespread both in Central Catalan and in Valencian.

In the development of *haver* from its original possessive values to those of a presentational nature that we are addressing here, *haver-hi* is closely tied to other verbs like *ésser* and, albeit to a lesser extent, *estar* (above all in Valencian), which are verbs that *haver-hi* has gradually replaced in different areas over the centuries. As stated by Ramos Alfajarín (2001), this evolution has passed through three periods. The first, which took place around the 16th-17th centuries, involved the affixing of the locative clitic pronoun *hi* (*haver* > *haver-hi*). Unlike Spanish, where the pronoun only appears grammaticalized in the third person singular of the present indicative (*ha + y*), in Catalan it is incorporated into the verb *haver* in all tenses, thus forming one single lexical element. The second period (18th-19th centuries) witnessed the generalization of the presentational usages of *haver-hi* with definite NPs (*hi havia la pubilla Cortines*) that were traditionally reserved for other “event” verbs (Solà, 1994: 22), such as *ésser* and *estar*, in an evolutionary process that again differs from that followed by Spanish.

Lastly, as of the 19th century and throughout the whole of the 20th, agreement also became a tendency within the spoken language. According to Ramos Alfajarín (2001: 138), one argument revealing the scarce occurrence of this phenomenon in earlier periods stems from the absence of any sort of reference to it in the grammars of the 18th century. In contrast, in the second half of the 19th century, the first fluctuations in its use began to appear, both in journalistic texts and in dramatic and satirical works that attempted to imitate popular speech.

### 3. Methodological issues

#### 3.1 Corpus and data processing

The material in Spanish used in this study comes from the *Macrocorpus sociolingüístico de Castellón y sus comarcas* (MCSCS – The sociolinguistic corpus of

Castellón and its province), a collection of interviews carried out by members of the Sociolinguistics Laboratory of the Universitat Jaume I between 1998 and 2006. It contains 210 sociolinguistic interviews, each lasting approximately 45-50 minutes. The method employed combined the anthropological method of life stories together with other thematic modules aimed at obtaining expository and argumentative texts.

Within the corpus, only 165 of the interviews contained examples related with the variable under study here, and thus the analysis is limited to those cases. The final sample was made up of 189 informants, who were initially stratified by gender, age and level of education, as well as by other social factors in a second post-stratification phase (place of residence and the dominant language).

The Catalan corpus consists of five textual databases that represent the two varieties of Catalan in which agreement is more prominent: Central Catalan and Valencian. Their distribution is as follows.

- For the Valencian variety, we have a collection of 20 interviews conducted using the same procedures as those outlined above by members of the Sociolinguistics Laboratory of the Universitat Jaume I.
- The other four subcorpora are made up of materials published electronically by the University of Barcelona to compile the *Corpus del català contemporani* (Corpus of contemporary Catalan). These are as follows:
  - a) *Corpus oral de varietats socials* (Cat-COS – Oral corpus of social varieties): which is a set of 25 semi-structured interviews conducted by means of life stories in the metropolitan area of Barcelona.
  - b) *Corpus oral de conversa col.loquial* (Cat-COC – Oral corpus of colloquial conversation): 10 conversations, each lasting on average around 30 minutes.
  - c) *Corpus oral de registres* (Cat-COR – Oral corpus of registers): consisting of a selection of 119 prototypical speech events from the Catalan speaking community, such as conferences, social and family gatherings, meetings, sermons, recipes, etc.
  - d) *Corpus audiovisual plurilingüe* (Cat-CAP – Multilingual audiovisual corpus): which is a selection of texts produced in Catalan by 12 informants, all of whom were female university students between 18 and 30 years of age.

After locating all the samples of the variable in both languages by means of a concordance program (*WordSmith* 6.0), they were submitted to a comparative sociolinguistic analysis (Poplack and Tagliamonte, 2001), consisting in performing two independent multivariate analyses with the same factor groups.

Taking into account the aims of this study, a number of means of evidence will be considered. First, we will assess the overall proportions reached by pluralization in Spanish and Catalan. For Spanish they will also be compared with those reached in other parts of the Spanish-speaking world, both in Latin America and in Spain.

Second, by means of a logistic regression analysis, and taking the pluralized form as the *application value*, we will perform a thorough analysis of the envelope of variation in both languages. In particular, we are interested in four statistical indices:

- a) The factor groups that are selected and not selected as being significant in each language
- b) The magnitude of those selected as significant, measured through their corresponding *ranges*

- c) The factor weights of each individual factor within those selected as significant; and
- d) The direction of effects, that is to say, the order of the factors on the axis that runs from the contexts that most favour the variant to the least favouring ones.

Our hypothesis is that the more indices are matched, the greater the congruence between the respective grammars is and, therefore, the higher the chances will be that both Spanish and Catalan have reached a more or less advanced level of convergence in this area of grammar as a consequence of the long-term contact. In contrast, finding openly opposing results would provide us with some clues about the diverging developments in the diffusion of the phenomenon.

### 3.2 Coding the linguistic variable

The factor groups examined are divided into three groups: linguistic, stylistic and social. In the following, we offer a schematic outline of them, while those factors selected as statistically significant will be analysed in more detail in a later subsection (see Section 4).

As far as the linguistic factors are concerned, they can be grouped in three domains:

- Syntactic: 1) verb category (simple/compound/periphrastic verb), 2) tense and mode (NB. only for the simple forms), 3) degree of determinacy of the argument (non-determinate/indefinite/definite), 4) category of the object (NP/PersP), 5) type of clause (subordinate/coordinate/others), 6) degree of adjacency between the verb and the argument (adjacent/non-adjacent), 7) position of the argument (pre-positioned/post-positioned), and 8) sentence modality (affirmative/negative/others).
- Pragma-semantic: 1) degree of animacy of the argument (animate/inanimate), 2) contextual modalization (intensified/neutral settings), 3) degree of semantic (im)personality of the sentence (active/passive and impersonal).
- Discursive: influence of the preceding context (*lexical priming*) (plural/singular/others).

The *stylistic* factors considered are:

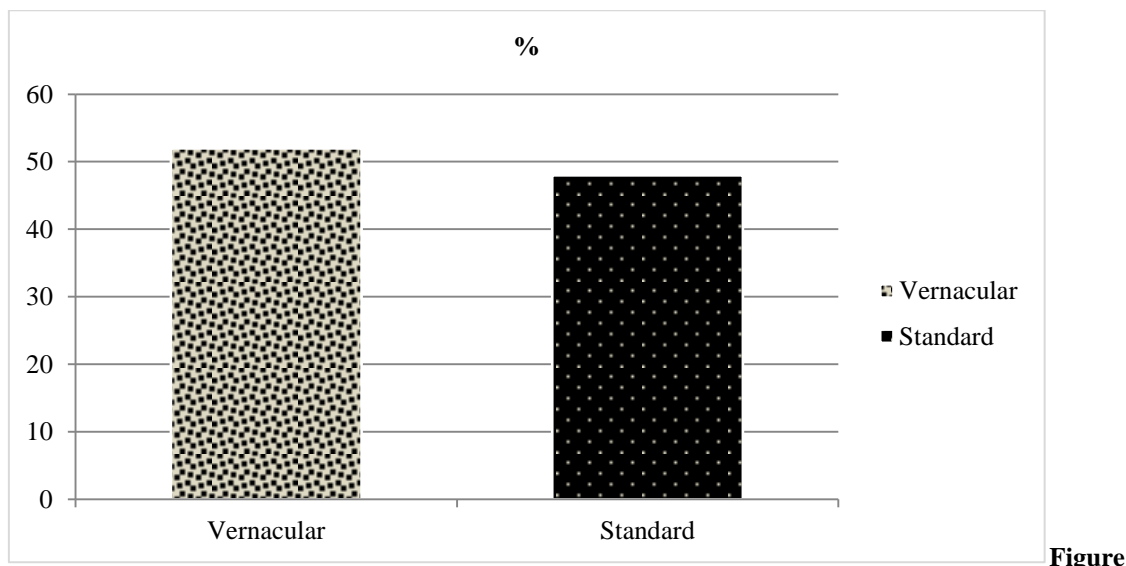
- Verbal interaction phase (initial/intermediate/final sequences)
- Tenor (+spontaneous/-spontaneous conversations)

Lastly, the *social* factors taken into account are:

- Gender (males/females)
- Age (-25 / 26-44 / 45-64 / 65-)
- Origin (Capital/Province)
- Level of education (Primary/Secondary/University)
- Dominant language (Spanish/Catalan/Balanced bilinguals)

## 4. Results and analysis

Figure 1 shows the overall levels of pluralization reached in the Spanish used in Castellón, compared with the normative variant.



1. General distribution of the variant in the MCSCS.

As can be seen, although the agreement phenomenon fails to reach the degrees of universality it was once presumed to possess,<sup>3</sup> it must be admitted that it does attain some far from negligible proportions that even exceed those of the standard variant.

### 4.1 Pluralization in Spanish-speaking America

On comparing these indices with those of some recently analysed in Latin American communities, it is observed that pluralizations in Castellón do not rank far behind the values found in America, at least in terms of urban speech types (see Table 1). On the whole, rural regions usually lead the ranking of vernacular variants, with percentages that exceed those found in the capital of their countries. Nevertheless, in the cities, and with the odd exception (the extreme cases of San Salvador and Mexico DF), the phenomenon is also clearly present, although, as in the case of Castellón, it does not reach the massive proportions once imagined.

**Table 1.** Comparison of the pluralized variants between Castellon and several Latin American speech communities (Sources: Castellon (this study), Caracas1 (Díaz Campos, 2003); Caracas2 (Aquino, 2004); Mérida (Domínguez *et al.*, 1998); Táchira Region (Freites, 2008); Mérida (Mx) (Castillo Trelles, 2007); Mexico DF (Lastra and Martín Butragueño, 2016); San Salvador (Quintanilla-Aguilar, 2009; Caguas (PR) (Brown and Rivas, 2012); Havana, Santo Domingo and Puerto Rico (Claes, 2014).

	%
<b>Castellón</b>	52
<b>Caracas1</b>	54

<sup>3</sup> As highlighted by Samper and Hernández-Cabrera (2012: 750) for the case of the Canary Islands: “The impressionist appreciations, even those of highly renowned linguists, may have highlighted the more striking forms and granted them a frequency that is higher than the one they really possess” (our translation).



<b>Caracas2</b>	63
<b>Mérida (Ven.)</b>	62
<b>Táchira (Ven.)</b>	82
<b>Mérida (Mx.)</b>	53
<b>México DF</b>	8
<b>San Salvador</b>	80
<b>Caguas (P.R.)</b>	56
<b>San Juan (P.R.)</b>	42
<b>Havana</b>	44
<b>Santo Domingo</b>	47

Likewise, in both the Castellón and the Latin American speech communities, significant sociolectal differences can be noted with regard to the diffusion of pluralizations, as can be seen in Table 2, where overall data on the highest sociolect in Castellón are compared with those obtained in other studies in different cities in Latin America (De Mello, 1991). As can be observed, the figures for Castellón are even higher than quite a few of these communities (in fact they are only surpassed by two cities, Lima and La Paz).

**Table 2.** Distribution of the pluralized variants in the high sociolects of nine Latin American cities (Source: DeMello, 1991: 449) and Castellón (this study).

<b>CITIES</b>	<b>%</b>
<b>Buenos Aires</b>	4
<b>Mexico DF</b>	8
<b>Bogotá</b>	16
<b>Havana</b>	27
<b>San Juan</b>	31
<b>Caracas</b>	36
<b>Santiago de Chile</b>	39
<b>Lima</b>	40
<b>La Paz</b>	60
<b>Mean (Latin America)</b>	26
<b>Castellón</b>	34

## 4.2 Pluralization in Spain

If the indices of pluralization in the region of Castellón are not far behind what happens on the other side of the Atlantic (at least as far as urban speech communities are concerned), what happens on this side of the ocean?

In an insight into the studies on grammatical variation on the PRESEEA project in Spain, Samper (2012: 48) states that:

all the teams in the project are studying two phenomena that need to be re-analysed in detail, with new theoretical and methodological formulations, in order to gain a more comprehensive understanding of them, namely, the expression of the subject pronoun and the use of personal *haber*, which is *hypothetically gaining ground in many Spanish speech communities* (the translation and italics are ours).

Without rejecting the possibility of this being the case in the future, the truth is that a brief review of the corpora published until now from several cities does not allow us to foresee such an expansion, at least for the time being (see Table 3). If we bear in mind that, in the eighties, Quilis (1983: 94) claimed to have found only two examples of

pluralization in a Madrid corpus made up of over 2,000 tokens of the variable, things do not seem to have changed very much in recent times. In the Alcalá de Henares oral corpus we only found seven examples, which represents barely 1.9% of the 361 instances of presentational *haber*. The same thing happens in Granada, with only three cases found in a total of 196 (1.5%). The figures for Malaga rise to 5.6% ( $n=22$ ), although most of them ( $n=16$ ) are cases of pluralization with the first person plural (e.g. *de nosotros ahora* *habemos tres*), which do not strictly form a part of the same variable, since this case involves agreement of the person and not just of the number. We should also bear in mind the total absence of the vernacular variant in the cultured speech used in cities like Salamanca, Madrid or Seville.

There are, however, two notable exceptions to what we have outlined in the previous paragraph: the speech communities of the Canary Islands and the regions in the east of the peninsula, such as Catalonia or the Valencian Region, with significantly higher indices of agreement. In the first case, dialectology had already drawn attention to the widespread diffusion of the phenomenon (Catalán, 1989: 155, 199), the conclusions being more recently confirmed by Pérez-Martín (2007) in a sociolinguistic study of the island of El Hierro. Nevertheless, the data from this study must be taken with caution, due to the low number of examples found in the corpus, as well as the fact that the study was limited to cases of variation affecting only the imperfect indicative tense. Conversely, other studies have restricted the supposed almost universal extension of pluralizations, at least among the higher sociolects. Indeed, in the ‘educated speech’ (*norma culta*) corpus of Las Palmas de Gran Canaria Samper and Hernández-Cabrera (2012: 750) find only seven examples of agreement in 42 tokens of presentational *haber* (17%), these being similar figures to those detected more recently among speakers with a high level of education in the PRESEEA corpus (20.5%).

In recent years, several studies have empirically confirmed the widespread diffusion of agreement in some Catalan-speaking territories. Thus, one pioneering study, conducted in the early nineties, highlighted the notable rise of agreement in the Valencian capital (57%) (Blas Arroyo, 1996). These data were completed with some positive attitudes towards the vernacular variant. Two decades later, Gómez Molina (2013) has found slightly lower levels of pluralization (46%) in this same city, but which are nonetheless high in comparison to those observed in other regions (see also Bouzouita and Pato, 2016).

In sum, several speech communities in the Catalan-speaking territories are clearly leading the way in pluralization in European Spanish, way ahead of other dialectal varieties, with the still-to-be-confirmed exception of the speech communities of the Canary Islands (further information on the sociolinguistic implications of these differences and affinities is given in the Final section).

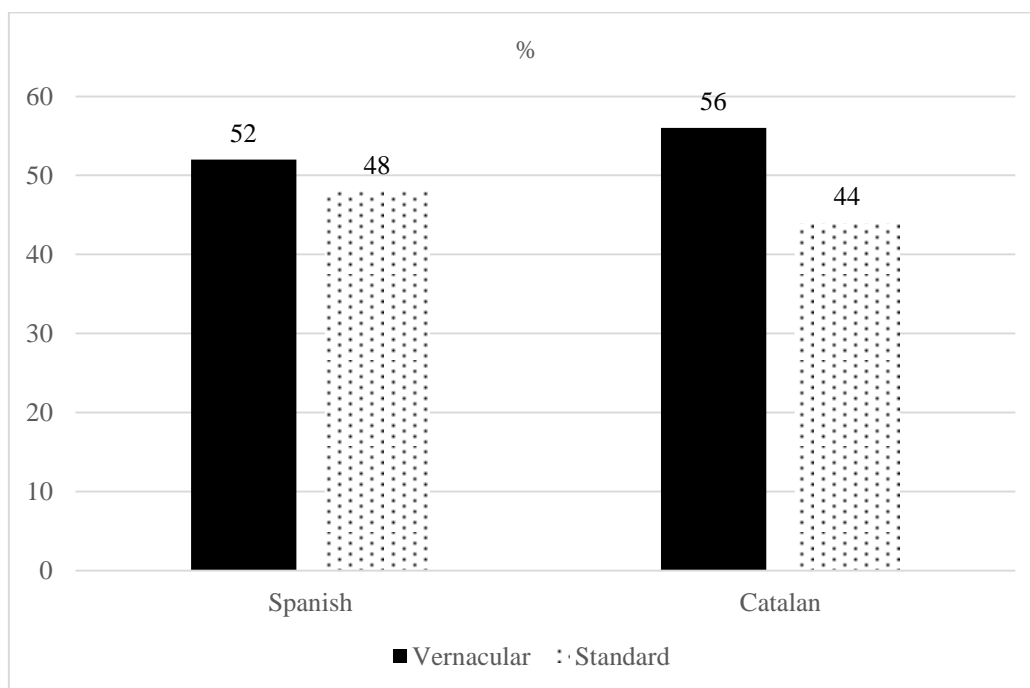
**Table 3.** Distribution of the pluralized variants in the different speech communities and sociolects of European Spanish (%) (Sources: El Hierro (Pérez-Martín, 2007); Las Palmas GC (high sociolect, PRESEEA; Samper and Hernández-Cabrera, 2012); Castellón (this study); Campanar-Valencia (Blas Arroyo, 1996); Valencia (Gómez Molina, 2013); Malaga (PRESEEA), Granada (PRESEEA), Alcalá (PRESEEA), Salamanca, Madrid (Quilis 1983), Las Palmas and Seville (*Norma Culta*)).

SPEECH COMMUNITIES	%
El Hierro	77
Valencia (1996)	57
Valencia (2013)	46
Castellón	52
Las Palmas de Gran Canaria (2012)	21

Las Palmas de Gran Canaria (norma culta)	17
Málaga	1,5
Granada	1,5
Alcalá de Henares	1
Madrid	0
Sevilla (norma culta)	0
Salamanca (norma culta)	0

### 4.3 Pluralization in Catalan

As can be seen in Figure 2, plural agreement with *haver-hi* occurs to a similar extent to that obtained in the Spanish of Castellón. Despite not reaching the massive indices that have sometimes been suspected, they do exceed those of the standard variant, with otherwise very similar figures (56%), although slightly higher than those of Spanish (52%). This coincidence is even greater when we compare the general data from the Spanish corpus and those from the subcorpus of the Catalan used in Castellón, as can be observed in Table 4 (52%).



Figure

2. Distribution of the variants in the two corpora (Spanish and Catalan).

All in all, the figures obtained in the different Catalan subcorpora are quite homogeneous in relation to each other, which lends support to the representativeness of the data. This gives us the idea that both in Valencian and in Central Catalan, the vernacular solutions present a diffusion that, although not overwhelming, are still particularly notable. The only exception to this panorama is that represented by the subcorpus *Corpus audiovisual* (Cat-CAP), with indices of pluralization that are slightly below the mean (43%). The reason for this may lie in the sample universe represented in this corpus, which consists exclusively of young female university students, a segment of the population generally more sensitive to normative pressure (see Section 4.4.2 below).

**Table 4.** Overall percentages of the realization of pluralized variants in the different Catalan subcorpora.

	%
<b>Catalan corpus (mean)</b>	56
--Cat-COC	60
--Cat-SOC	56
--Cat-COR	56
--Cat-CAP	43
--Cat-CS	51
<b>Spanish corpus (MCSCS)</b>	52

#### 4.4 Comparative analysis (so far...)

From what we have seen up until now, we can draw several conclusions. The phenomenon of pluralization with presentational *haber* is quite widespread in the Spanish spoken in the Castellón speech communities, even exceeding the standard realizations. As we have shown, this does not differ greatly from what has been observed in some Latin American speech communities, especially those of a more urban nature, but the difference is much greater in comparison to the other European speech varieties, with the exception of the Canary Islands, traditionally more linked to Latin American Spanish. Moreover, the analysis of the phenomenon in Catalan highlights a diffusion that is overall very similar, with proportions of pluralizations that follow close behind those observed in Spanish.

In view of the difficulties stemming from a direct causality between what happens in the Spanish speech used in the Catalan-speaking territories and what occurs thousands of kilometres away in the Canary Islands and Latin America: should we postulate a specific origin not so much for the phenomenon itself – the common structural conditionings of which are more than obvious (see Section 2.1 and 2.2) – but rather in relation to its particular diffusion in these peninsular varieties? And can these similarities be conceptualized as a manifestation of convergence in situations involving centuries of language contact, such as that which exists between Spanish and Catalan?

The similar proportions that this vernacular phenomenon shows in the two languages is a relevant piece of data to be taken into account, but in no way is it sufficient in its own right. Consequently, in the following, we conduct a comprehensive comparison of the envelope of variation in both languages, with the aim of determining the degree of affinity at deeper levels of the analysis, both on the linguistic and the socio-stylistic axes.

Starting from the idea that convergence is not a monolithic outcome – all or nothing – but instead a continuum in which the languages in contact are more open to solutions that are generally simpler and more regular than in their monolingual counterparts, our hypothesis is that the more similarities the respective underlying grammars offer, the higher the degree of convergence between the languages will be. And vice versa: the more differences we find, the greater the distance between the two languages will be.

#### 4.5 The envelope of variation

One element of proof to test the degree of congruence of the grammars in the variation phenomenon we are dealing with here is the number and the identity of the factor groups both selected and not selected as significant by the logistic regression program. A preliminary review of the data shows a notable degree of congruence between the two languages, as can be seen in Table 5.

**Table 5.** Factors selected and not selected as significant in explaining the variation in Spanish and Catalan.

FACTORS	Spanish	Catalan
Type of clause	–	–
Degree of determinacy	–	–
Degree of adjacency	–	–
Position of the argument	–	–
Sentence modality	–	–
Contextual modalization	–	–
Sentence (im)personality	–	–
Verbal interaction phase	–	–
Age	–	–
Origin	–	–
Lexical priming	+	+
Level of education	+	+
Tenor	+	+
Degree of animacy	+	+
Gender	+	+
Verb category/tense	+	–
Sentence modality	–	+
Dominant language	+	–

Hence, in 15 out of the 18 factor groups considered in the analysis, there is total coincidence, whereas only three show some divergence, either because they have been selected for Spanish, but not for Catalan, or the other way round. For expository purposes, in the next subsection we will address the factors in which there is coincidence, while in Section 4.5.2 we will analyse in detail those in which some divergence between the two languages is exhibited.

#### 4.5.1 Factors selected as significant in the two languages

Goldvarb selected five factor groups as significant in Spanish and Catalan, whereas the other twelve are rejected as being statistically non-relevant.

The first factor groups include two of a linguistic nature (*lexical priming* and degree of *animacy* of the argument) and three of a stylistic (*tenor* of the interactions) and social character (*level of education* and *gender*). Moreover, three of those factor groups run in the same explanatory direction (*lexical priming*, *level of education* and *tenor* of the interactions), whereas in the other two (*degree of animacy* and *sex*) the direction of effects is inverted. We will begin our analysis with the first groups.

**Table 6.** Factor groups selected as significant in the two languages (Goldvarb X).

	Spanish corpus	Catalan corpus
FACTORS	FW	FW
<b>Lexical priming</b>		
Yes (singular)	.33	.28
No	.47	.50
Yes (plural)	<b>.81</b>	<b>.88</b>
<i>Range</i>	48	60
<b>Level of education</b>		
Higher	.34	.30

Rest	<b>.55</b>	<b>.54</b>
Range	21	24
<b>Tenor (interviews)</b>		
-Spontaneity	.44	.48
+Spontaneity	<b>.56</b>	<b>.54</b>
Range	12	6
<b>Degree of animacy</b>		
Inanimate	.46	<b>.54</b>
Animate	<b>.60</b>	.39
Range	14	15
<b>Gender</b>		
Males	.44	<b>.60</b>
Females	<b>.54</b>	.45
Range	10	15

The most explanatory factor is the influence of *lexical priming* (Pickering and Ferreira, 2008). More specifically, agreement in the two languages is especially favoured when the same verb appears in plural in the previous context, with factor weights (FW) that are very high and similar to each other (Sp: .81; Cat: .88). This is the opposite to what happens when the linguistic variable is preceded by singular forms (Sp: .33; Cat: .28). Lastly, the other contexts fall in an intermediate position, close to neutrality (Sp: .47; Cat: .50).

The second factor group in which both languages agree is the level of education of the speakers. An initial frequency analysis confirms the existence of a notable degree of similarity in the majority sectors of both language communities, i.e. those made up of individuals with primary (Sp. 54%; Cat. 57%) and secondary education (Sp. 57%; Cat. 58%) only, with agreement proportions that are substantially higher than those found among the segments of the population with higher education (Sp. 34%; Cat. 40%). Thus, for the multivariate analysis we grouped the first two in one single group and compared it with that comprising individuals with a university education. The results again show a high degree of congruence between the two communities, with very similar probabilistic weightings and ranges. This would indicate that, despite the widespread social diffusion of the phenomenon, the normative pressures are perceived with greater intensity among the more cultured sectors of society.

The multivariate analysis also reveals that plural agreement is significantly more common in the interviews that allow for a higher degree of spontaneity between the interlocutors. In producing this factor we used three different parameters:

- a) The predominant topic in the interviews, a distinction being made between everyday topics and others of a more specialized nature (labour legislation, immigration, etc.);
- b) The degree of relationship existing between the participants prior to the interview (close/distant/non-existent relationship); and
- c) The intergenerational distance (same generation/difference of one or more generations).

After cross-checking all these factors, the interviews were classified into four groups, of which – in accordance with our current interests – we only code those lying at the extremes, these being the most and the least spontaneous conversations, respectively. The results show that, both in Spanish (.56) and in Catalan (.54) agreement was slightly favoured in the more spontaneous interviews than in those situated at the other end of this stylistic axis (Sp: .44; Cat: .48). Note again the similarity between the corresponding FW, as well as the fact that in both languages this factor group obtained

the smallest range of all. This should be interpreted as being a consequence of the fact that, despite being an element that is more characteristic of colloquial syntax, pluralizations also extend throughout other more formal registers in both languages.

The magnitude of the other two factors selected as significant is also practically identical in Spanish and Catalan, although in this case the direction of effect is inverted. Indeed, whereas in Spanish the vernacular variant is especially favoured in arguments of an animated nature (.60), the opposite occurs in Catalan, where this role appears to be reserved for inanimate objects (.54). In this regard, the Spanish used in Castellón displays similar constraints to those found in other Spanish dialects (Caracas (Bentivoglio and Sedano, 1989), Mérida (Venezuela) (Domínguez *et al.*, 1998), San Cristóbal de los Andes (Freites, 2008), Mexico (Montes de Oca, 1994), Mérida (Mexico) (Castillo Trelles, 2007), El Hierro (Pérez-Martín, 2007), Puerto Rico (Rivas and Brown, 2012) and El Salvador (Quintanilla, 2009)).

The same Pan-Hispanic profile is displayed by the *gender* factor group. In this regard, the Castellón corpus shows a tendency that has been repeatedly observed in studies on Spanish, with females favouring the agreed forms slightly more (.54) than men (.44) (cf. Caracas (Bentivoglio and Sedano, 1989; Díaz Campos, 2003), Santiago de Chile (Díaz Campos, 1999-2000), Mérida (Mexico) (Castillo Trelles, 2007), El Salvador (Quintanilla, 2009) or San Juan de Puerto Rico (Claes, 2014), among others). In contrast, in the Catalan speech community that role seems to be played mainly by males (.60).

#### 4.5.2 Factors not selected as significant in the two languages

An element that is equally worth noting is the even longer list of factor groups rejected by the multivariate analysis (see Table 7).

**Table 7.** List of non-selected factor groups in the two languages (Goldvarb X) (NB. In the absence of probabilistic weightings, the percentages of these non-significant factors are given between square brackets).

<b>FACTORS</b>	<b>Spanish</b>	<b>Catalan</b>
<b>Syntactic category</b>	<i>n.s.</i>	<i>n.s.</i>
NP	[49]	[52]
PersP	[58]	[54]
<b>Degree of determinacy</b>	<i>n.s.</i>	<i>n.s.</i>
Determined/Definite	[52]	[59]
Determined/Indefinite	[56]	[51]
Non-determined	[46]	[50]
<b>Degree of adjacency</b>	<i>n.s.</i>	<i>n.s.</i>
Adjacent	[51]	[52]
Non-Adjacent	[52]	[52]
<b>Position</b>	<i>n.s.</i>	<i>n.s.</i>
Post-positioned	[50]	[52]
Pre-positioned	[58]	[50]
<b>Sentential syntax</b>	<i>n.s.</i>	<i>n.s.</i>
Relative subordinate clauses	[48]	[57]
Other subordinate clauses	[48]	[59]
Coordinated or main clauses	[51]	[51]
<b>Contextual modalization</b>	<i>n.s.</i>	<i>n.s.</i>
Intensified	[51]	[50]
Neutral	[52]	[54]
<b>Interaction phase</b>	<i>n.s.</i>	<i>n.s.</i>
Initial	[49]	[54]
Intermediate	[56]	[58]
Final	[48]	[48]

Age	<i>n.s.</i>	<i>n.s.</i>
-25	[48]	[45]
26-45	[52]	[58]
46-64	[50]	[57]
>65	[54]	[48]
Origin	<i>n.s.</i>	<i>n.s.</i>
Capital	[47]	[55]
Province	[54]	[50]

These include linguistic factors, such as the verbal category of the argument (NP/PersP), the degree of adjacency (Adjacent/Non-adjacent) and the position (pre-positioned/post-positioned) with respect to the verb. As can be observed in the table, the differences within these groups are minimal in the two languages. In the same vein, the syntactic type of sentences in which the presentational verb appears does not affect the degree of agreement in either of the two languages. And neither does agreement seem to be an iconic resource aimed at strengthening the illocutionary force of the utterances, as other syntactic phenomena have been interpreted as doing, such as the insertion of *de* in *dequeísmo* or in the modal periphrases with *deber* (Blas Arroyo, 2011).

Even more revealing is the fact that the degree of determinacy was not selected. As we saw in an earlier section, one of the main differences between the syntax of the presentational predicates with *haber* and *haber-hi* lies in the greater capacity of Catalan to combine definite arguments, especially with the article, which is a more restricted possibility in Spanish. Our data show that in this context agreement takes place more often in Catalan (59%) than in Spanish (52%), but the differences, as in the other settings, are small and not significant in any of the cases.

On the stylistic axis, the widespread diffusion of the vernacular variant, often below the level of awareness, would explain why we also fail to find any significant differences in relation to the interaction phase. If this were the case, we would expect to find more cases of agreement in the later stages of the semi-structured interviews, those in which the initial tension of the interview gives way to a more relaxed atmosphere between the participants. As we can see, this is not the case.

Finally, on a social level, neither age nor the origin of the speakers is found to be significant in either of the communities. The phenomenon does not appear to be especially associated to the younger segment in the Spanish used in Castellón (48%), and neither is this the case in the Catalan-speaking community (45%). Moreover, in both linguistic communities, this first generational cohorts last in terms of the number of agreed tokens. This fact may suggest the existence of a certain degree of stabilization of the vernacular phenomenon or even a slight regression, although the differences with respect to the other age groups are so small that they do not allow us to draw any more accurate conclusions. Likewise, the setting in which the informants live (capital/province) does not make it possible to establish any relevant differences in either of the two communities.

#### 4.5.3 Differentiating factors

As we pointed out at the beginning of this section, there are three factor groups which were selected by the statistical program as displaying differences in the two speech communities. Two of them are of a linguistic nature (verb category and sentence modality) and the other one is of an extralinguistic character (the dominant language).



**Table 8.** Factor groups selected as significant in one language but not in the other (Goldvarb X).

FACTORS	Spanish	Catalan
<b>Verb category</b>	<b>FW</b>	<b>FW</b>
Compound forms	.27	[64]
Simple forms	.50	[52]
Periphrastic forms	<b>.64</b>	[46]
<i>Range</i>	39	n.s.
<b>Sentence modality</b>		
Affirm. Declarative	[51]	<b>.52</b>
Neg. Declarative	[51]	.32
Non-declarative	[51]	--
<i>Range</i>	n.s.	20
<b>Dominant language</b>		
Valencian	.39	[56]
Both	.50	[49]
Spanish	<b>.62</b>	[37]
<i>Range</i>	23	n.s.

As far as the first factor group is concerned (verb category), in the Spanish corpus a significant difference can be observed between the cases in which *haber* is the only representative of the verb group and others in which it is accompanied by other verbs. In the latter, however, there is a very significant distance between the compound forms of *haber*, such as the examples in (5), and the cases in which it acts as the main verb in a periphrastic construction, as in (6).

- (5) *Han habido* varias personas, una de ellas mi madre. (MCSCS 212) “There have been several people, one of them being my mother.”
- (6) No:: pero:: + no sé, pienso que:: igual que estamos nosotros aquí, pueden, *pueden haber* más personas, no? (MCSCS 333)  
 “No, but, I don’t know, I think that, just as we are here, there may, there may be more people, right?”

The former are by far the contexts that least favour the vernacular variant (.27; 29%), while the latter favour it to a notable extent (.64; 62%). This second group of contexts seems to confirm the hypothesis defended by Hernández Díaz (2006: 1152), who, based on a cognitive argument, holds that pluralization occurs more often when the speaker “has less control over the forms”, which gives rise to an “indirect agreement” between the auxiliary verb and the argument of *haber*. However, this author suggests that this type of agreement also takes place in the compound forms of the verb, a fact that is not found in our corpus. In fact, quite the opposite occurs: the number of pluralizations drops sharply in this context, which perhaps could now be interpreted as a matter of syntactic phonetics. Indeed, the agreement breaks the synalepha between two forms that end and begin with the same unstressed *a* (*ha(n) habido*), and which speakers often pronounce run together as one sound in spontaneous conversation.

Between these two extremes, there are the simple forms of the conjugation, the overall contribution of which is apparently neutral (.50; 51%). A more detailed analysis, however, reveals the existence of notable differences from one paradigm to another. On the side that favours pluralizations, we find the imperfect subjunctive (.77), the simple conditional (.59) and the imperfect indicative (.53). Attention must be drawn to the fact that this latter paradigm, which has often been seen as more favourable for agreement (Bentivoglio and Sedano, 1989; Díaz-Campos, 1999-2000; Pérez-Martín, 2007;

Quintanilla, 2009), is not so in practice in the Castellón corpus (see also Bouzouita and Pato, 2016), despite the fact that its notable recurrence in discourse (54% of all the examples of the variable) may be misleading.

We also have other forms, such as the future simple (.35) and above all the past simple (.14), which disfavour the variant. The particular resistance of the past simple to agreement (*hubieron fiestas*) is one of the most unanimous findings in the sociolinguistic literature. Several hypotheses have been put forward to justify this fact, such as the scant frequency of this tense in discourse (Quintanilla 2009: 185) or the 'excess' of phonetic material in *hubieron* (and the corresponding higher social stigma resulting therefrom) (Hernández Díaz, 2006: 1151). Yet, both criteria are to be found in other forms, such as the imperfect subjunctive (*hubiera(n)*, *hubiese(n)*), which does not prevent it, curiously enough, from being the paradigm that most favours pluralization in the Castellón corpus (.77).

Hardly any of the foregoing occurs in Catalan. However, on this point the grammar of this language presents an important structural conflict with Spanish, which makes any comparison worthless. Indeed, whereas Spanish practically precludes the possibility of agreement in the present indicative,<sup>3</sup> the opposite occurs in Catalan. As a matter of fact, it is in this last tense that the linguistic variable appears on a massive scale in the Catalan corpus, since it alone accounts for 81% of all its occurrences ( $n=428$ ), followed far behind by the imperfect indicative ( $n=66$ ) and, with practically negligible figures, by the remaining paradigms. In practice, this significant imbalance rules out any chance of comparison with Spanish as regards the verbal category. In fact, all the simple forms account for 95.5% of the examples of the variable ( $n=504$ ) in Catalan, versus a scant 2% ( $n=11$ ) for the compound forms and 2.5% ( $n=13$ ) for the other periphrastic forms.

In sum, the difference between Spanish and Catalan on this point lies not so much in the different significance of the factors as in the distinct configuration of the grammar (blockage of the present indicative in Spanish) and the extraordinary sampling mismatch between different contexts in the two corpora. What would have happened in a broader corpus that contained a more comprehensive and representative selection of forms other than the present? Logically, it is not possible to put forward a definitive answer to this question with the information available, but some data from the present study do appear to be revealing. Thus, agreement with the imperfect indicative in Catalan – the second most frequent in discourse, although at a considerable distance behind the present – is even more common in this language (62%) than in Spanish (55%).

Sampling issues also underlie the different treatment offered by the sentence modality in each language. While no differences whatsoever are observed in the agreement indices between both affirmative and negative declarative sentences and non-declarative sentences in Spanish (51% in all three), the difference between the first two<sup>4</sup> initially appears as significant in the Catalan corpus. In this latter corpus, pluralizations are less frequent in the negative polarity (.30; 32%) than in the affirmative (.52; 54%).

In both languages, the overall figures for affirmative sentences are very similar, while the differences are concentrated in the negative ones. Why? Again the reason is to be found in a peculiarity of the sample of the Catalan corpus, this time deriving from the

---

<sup>3</sup> As stated above, the cases of vernacular forms like *hayn* or *haen* are still very scarce in the Spanish-speaking world. In his study, Claes (2014: 81) only detects 53 examples in a Caribbean corpus containing over five thousand instances of the variable.

<sup>4</sup> In the Catalan corpus we found only six instances of non-declarative sentences, and thus they were left out of the multivariate analysis.

interaction between this factor and the level of education. As the reader will recall, individuals with a higher education stood out above the rest in terms of a significantly lower presence of the vernacular variant. Cross tabulation with the sentence modality now reveals that of the 38 examples of negative sentences in the Catalan corpus, no fewer than 34 correspond to persons included in this high sociolect. And of these, at least seven involve the repetition of the same sentence (*no hi ha nazis vienesos*) by several actors in a play, whose rehearsals are included in the oral texts in one of the Catalan subcorpora. In sum, we find ourselves before a new factor group in which the disparity between the two languages is more a result of an imbalance in the sampling than due to structural differences.

More intriguing – and initially counterintuitive – is the apparent influence of the dominant language.

Although practically all the members of the Spanish corpus display some degree of bilingualism in the two languages, things are quite different in terms of their active usage. Although all the habitual Catalan-speakers are in turn, albeit with varying levels of proficiency and frequency of use, speakers of Spanish, the same cannot be said of the Spanish-speakers, some of whom hardly ever use Catalan in their everyday communication. Accordingly, in order to code this independent variable, the sample was divided into three groups: a) speakers with Catalan as their dominant language; b) speakers with Spanish as their dominant language; and c) speakers with a balanced proficiency in both languages.

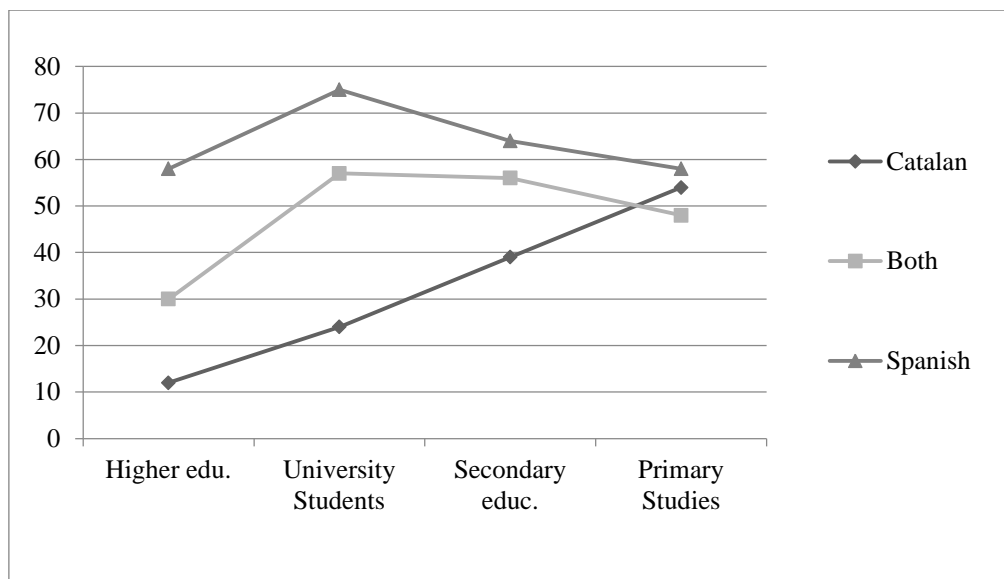
An initial hypothesis led us to suppose that the degree of pluralization would be directly proportional to the level of knowledge and usage of Catalan, which is the language in which the phenomenon also occurs with an especially higher frequency. If this were so, we could expect the bilingual speaker to realize the same variant with a similar frequency in both languages, a possibility that, in contrast, would be reduced among the Spanish-speakers, who are subject to a smaller amount of vernacular input. In fact, in one study conducted on the district of Campanar (Valencia) Blas Arroyo (1996) found that the dominant Catalan-speakers not only pluralized significantly more than the Spanish-speakers in completion tests, but, at the same time, they also displayed attitudes that were especially favourable towards the phenomenon. And differences in the same line have also been observed in a recent sociolinguistic study carried out in the city of Valencia, both in interviews and in an acceptability test (Gómez Molina, 2013).

The results of the present study confirm the significance of this factor group, but, paradoxically, in completely the opposite direction to the expected one. In fact, the greatest numbers of plural agreement appear among speakers with a better command of Spanish (.62; 62%), followed by the balanced bilinguals (.50; 51%) and finally individuals who make greater use of Valencian (.39; 42%).

Does the foregoing mean that the hypothesis of linguistic convergence is therefore discarded, as has been suggested (Gómez Molina, 2013: 133)? In our opinion, not necessarily. In fact, we believe that these results can be explained by the different way in which the normative pressures affect each group, especially in certain social sectors.

One argument to support this hypothesis is provided by the interaction between the dominant language and the speakers' level of instruction. As a matter of fact, the vernacular variants display a significant decrease among the speakers with a higher education, followed by the students who were following a university degree at the time of the interview. In contrast, pluralizations are far more common among the groups without university qualifications, whose linguistic awareness concerning the 'correctness' of this syntactic phenomenon appears to be lower. Yet, a cross-analysis

between this factor and the dominant language, as can be seen in Graph 3, shows a very uneven behaviour among the corresponding subgroups, depending on the informants' language ascription.



**Figure 3.** Cross tabulation between the *level of education* and the *dominant language* of the speakers in the MCSCS (%).

Note how the most diverging behaviour occurs between the groups that are initially more sensitive to the norm. Hence, the differences between university students who have Spanish as their dominant language are vastly superior in number (75%) to the dominant Catalan-speakers (24%), with the balanced bilinguals lying in an intermediate position (57%). The same tendencies can be observed between speakers with a higher education, where the low intensity of agreement among Catalan-speakers is especially striking (12%), ranking a long way behind the balanced bilinguals (30%) and the Spanish-speakers (58%). Conversely, these differences are attenuated to a notable extent among the individuals with a secondary education, whereas they are practically neutralized in the majority sectors of society, that is, among those with only a basic education.

Yet, the most notable aspect of this hypercorrective behaviour is that it is also observed in the Catalan corpus... but now in the opposite direction. Although the differences between the groups are not statistically significant now,<sup>6</sup> it is nevertheless revealing that the direction of effects is completely the opposite. In fact, the dominant Catalan-speakers lead in terms of pluralizations (56%), closely followed by the balanced bilinguals (49%) and, much further behind, by those who habitually speak Spanish (37%).

In the end, the hypercorrection hypothesis would explain some distributional facts that we will outline in the following:

- a) Both in one language and in the other, the balanced bilinguals are always situated in an intermediate position.
- b) Hypercorrection is particularly active in sociolects where linguistic awareness is more prominent. As we have seen in Graph 4, this is especially apparent among dominant Catalan-speakers when they speak Spanish, which would explain the

<sup>6</sup> We cannot rule out a possible contribution of the notable imbalance of the sample existing on this point in the Catalan corpus, where the representation of Spanish-speakers and balanced bilinguals is clearly lower.

unexpected low agreement figures among informants with a higher education. In this sense, the fact that the number of pluralizations of this sociolect when they speak Catalan is nearly four times (40%) more frequent than those realized by this same group in Spanish (12%) does not appear to be due to chance.

c) Equally revealing is the interaction between this language ascription factor group and gender, above all among females, who are generally more susceptible to the pressures exerted by language prestige (Tagliamonte, 2012: 32-34). This would account for the fact that agreement in Catalan rises sharply above all among Catalan-speaking females (55%), followed by balanced bilinguals (49%) and lastly dominant Spanish speakers (36%). Exactly the opposite occurs in the Spanish corpus, where Catalan-speakers females are now the most reluctant to use the agreed solutions (45%; versus 56% of balanced bilinguals and 68% among the Spanish-speakers).

d) Last, but by no means the least revealing, mention should be made of the cross-analysis with the tenor of the interactions, and more particularly with the context that most favours agreement, that is, the more spontaneous conversations. In this regard, tellingly, the most relaxed settings encourage agreement in Catalan significantly more among those who have this as their dominant language (60%) (vs. 35% and 2% for balanced bilinguals and dominant Spanish-speakers, respectively). Again, this is contrary to what happens in the Spanish corpus: this time the pluralizations rise sharply among speakers of Spanish (62%) to a far greater extent than in the other groups (29% and 27%, respectively).

In the end, although firmly rooted among the expressive habits of the two speech communities, in some social sectors there seems to emerge an awareness that, deep down, this is an 'incorrect' variant that should be corrected. Institutions such as schools, universities, etc. are responsible for remarking on this, and it therefore comes as no surprise to find this behaviour with greater frequency in informants who are more concerned about such normative pressures in the languages in which they have less proficiency.

## **5. Final: variation, hypercorrection, koineization in a convergence outcome**

In the preceding pages, we have seen how the pluralization between the verb *haber* and its argument in presentational constructions takes on considerably high proportions in the Spanish spoken in Castellón. These data, which confirm those from earlier studies, are much higher than the results from other peninsular speech communities, where the variation is considerably lower.

Furthermore, the analysis of the same variation in Catalan reveals very similar indices of diffusion of the vernacular variant, both in the Castellón and in the Central Catalan subcorpora. Nonetheless, an even more relevant finding is that the grammar underlying the phenomenon in the two languages displays a notable degree of congruence. Indeed, a comparative study of two independent multivariate analyses has enabled us to confirm the existence of a number of points on which the conditioning of variation coincides in both languages. In no less than 15 of the 18 factor groups analysed, there is a complete coincidence between the factors selected and non-selected, even although two of the former (gender and degree of animacy) display a different direction of effect in each language.

There are also three other factor groups that have been selected in one language, but not in the other (verb category, sentence modality and dominant language). Yet, a

thorough analysis of the envelope of variation has allowed us to determine the origin of those differences. Sometimes they are related to certain imbalances in the samples of the two languages. Accordingly, we have seen how, in the Catalan corpus, the presentational construction occurs on a massive scale in the present indicative, with very few instances of the variable in other verb paradigms. This fact rules out the possibility of really comparing this factor with the Spanish corpus, where such paradigms are spread out far more uniformly (despite the fact that the imperfect indicative tense is prominent in this latter language). Furthermore, the profile that initially disfavours the pluralized variant in contexts with a negative polarity in Catalan literally comes apart on observing that almost all its scant instances are produced by individuals from the high sociolects, among whom the frequency of agreement drops significantly.

More complex, and intriguing, is the result obtained on comparing the diffusion of the phenomenon among different language ascription groups in each speech community. Thus, we have seen how pluralizations are directly related to the degree of proficiency and usage of the languages. This explains why, in Spanish, the dominant Spanish-speakers lead in the vernacular variant (followed by balanced bilinguals and habitual Catalan-speakers, in that order), whereas in Catalan quite the opposite occurs: it is the dominant Catalan-speakers who pluralize more, followed by balanced bilinguals, with the dominant Spanish-speakers closing the cycle.

From a sociolinguistic point of view, this phenomenon thus differs from other variation facts related with language contact, as occurs with the pronunciation of the /d/ in several phonic settings (Blas Arroyo, 2007), or the unusual vitality in the Catalan speech communities of grammatical variants that are undergoing a severe process of erosion in other dialects, such as the morphological future (Blas Arroyo, 2008; Illamola, 2015) or the modal uses of *haber de* + infinitive (Sinner, 2004; Blas Arroyo, 2015). In all these cases, we have seen the existence of a considerable correlation between the usage of several variants and the ethnolinguistic ascription of the speakers.

This is not exactly the case, however, in the plural agreement with the presentational verb *haber*. Why? In our opinion, there may be a number of different – although not necessarily exclusive – reasons. On the one hand, and unlike the linguistic variables considered in the previous paragraph, in this case it is not clear where the possible origin of the phenomenon lies: is Spanish to blame for the fact that pluralization is performed in Catalan? Conversely, is this last language mainly responsible for a phenomenon in Spanish that is, for the time being, significantly scarcer in other peninsular speech communities? In the late 19th century, the works on Spanish grammar published in Catalonia harshly condemned the agreement that "els catalans feien parlant castellà", as Solà (1994: 23) reminded us. Yet, this does not necessarily mean that the origin of it was the direct influence of Catalan. In any case, what it did trigger was an equally condemning movement by the Catalan grammars that has continued down to the modern day.

It is in this crossfire of normative pressures where we find an eventual explanation to understand the behaviour of different social groups in relation to the vernacular variant in each language. It is a fact that pluralizations have been condemned by the normative grammar since at least the mid-19th century. Both the grammars in Spanish and those in Catalan consider these sentences with *haber* or *haver-hi* to be impersonal and, accordingly, they deem the agreed solutions to be incorrect. Moreover, although very frequently the pluralizations go unnoticed, the pressure against the use of the vernacular solutions is still present in both language communities. This pressure is carried out in the form of warnings by grammarians, secondary school and university

teachers, the style guides of the mass media, and other standardizing institutions that watch over the 'purity' of the language.

But what sectors do these normative pressures tend to reach most often? The results of this study prove the existence of a certain tendency towards hypercorrection in speakers who express themselves in a language that, although familiar to them, is not their dominant tongue. Hence, Catalan-speakers with a higher education, who often employ the agreed form when expressing themselves in their own language, may feel unsure when it comes to doing so in Spanish, guided perhaps by the assumption that pluralization in Spanish is a 'Catalanism' that should be avoided.<sup>7</sup> And vice versa, a habitual speaker of Spanish who is a frequent 'pluralizer' in his or her own language but not too sure of the ins and outs of Catalan grammar may avoid this vernacular variant when speaking Catalan.

Does this mean that the outcome of linguistic convergence is oblivious to the phenomenon we are dealing with here? To provide an answer to this question, let us return to some of the requirements that, according to Poplack *et al.* (2012: 205), are necessary to endorse the influence of language contact in the phenomena of linguistic variation and change. Firstly, the evolution of the phenomenon in each of the languages seems, without a doubt, to be an outcome of linguistic change, which was barely visible for centuries, but nevertheless gained momentum as of the second half of the 19th century.

Furthermore, the variant does not appear to any significant extent in other varieties of the language without contact, a fact that is almost perfectly demonstrated in our case, as we have seen on comparing its diffusion with other speech communities of the Peninsula, in which it reaches very low levels. Indeed, this variation is widespread in Atlantic Spanish (Latin America and the Canary Islands), whose historical relationships with one another are sufficiently proven, unlike what happens in those addressed here. Consequently, there is nothing to ensure that the seed from which their diffusion arose on both sides of the ocean must be the same.

Lastly, linguistic convergence requires that the phenomenon must take place "in the same way as its putative borrowed counterpart in the source variety [and] it differs in non-trivial ways from superficially similar constructions in the host language" (Poplack *et al.*, 2012: 205). The multivariate comparative analysis shows that the envelope of variation is particularly coincident in the two languages, although this does not mean that the linguistic and sociolinguistic matrices of the variation necessarily have to be identical.

In our opinion, there are therefore enough arguments to defend the particular extension of pluralizations in these speech communities as a fact arising from linguistic convergence. In the absence of diachronic studies that can provide a more definitive answer to this question (which are difficult to conduct given the insufficient representation of pluralization in written texts) we postulate a socio-historical explanation for this particular ('non-grammatical') outcome.

In a recent study, Claes (2014) suggests koineization as the main process to explain the vitality of this vernacular phenomenon in the speech of many Latin American

---

<sup>7</sup> Deep down, this behaviour is similar to that driving the *dequeísta* solutions in the Spanish spoken by educated Catalan-speakers and of which many examples can be found in the mass media on a daily basis. Perhaps being aware of the fact that the absence of a preposition, which represents the norm in Catalan syntax, does not coincide in many cases with that of Spanish ('*soy consciente de que.... / sóc conscient que....*') leads to this over-representation of *dequeísta* realizations among these speakers when they express themselves in Spanish.

communities.<sup>8</sup> In his opinion, the diffusion of plural agreement represents a clear case of simplification, understood to mean the victory of more transparent structural solutions in situations of intense interdialectal contact such as those experienced by Spaniards during the conquest of America in the 16th and 17th centuries.

The case we are dealing with here presents an identical structural cause (in both communities the agreed solutions represent a far more logical and simple ‘possibility of the language’ than non-agreement), which was equally aided by the notable vitality of the interdialectal contact that characterized the regions of Central Catalonia and Valencia from the late 19th century through most of the 20th. During this long period of time, both communities (especially Catalonia) have been the destination of millions of Spaniards who emigrated from other regions – above all in the south – to these lands in search of a better life. So much was this the case that, in the 1970s, almost half the Catalan population was of a non-Catalan origin.

All in all, in the Valencian-Catalan case, the intense interdialectal contact is perhaps a condition that is necessary, but not sufficient, to account for the koineization and the particular diffusion of agreed solutions in these regions. In fact, it has a far more limited presence in other areas of Spain, which are also traditionally receivers of immigrants, such as Madrid or the Basque Country. What makes the phenomenon addressed here more singular is the mark left by centuries of interlinguistic contact between two Romance languages, the speakers of which exploit the convergent ‘possibilities’ offered by the two languages in order to reach common, as well as simpler, linguistic outcomes.

## References

- Acadèmia Valenciana de la Llengua. 2006. *Gramàtica normativa valenciana*. València: Publicacions de la AVL.
- Badia i Margarit, A. 1994. *Gramàtica catalana. Descriptiva, normativa, diatòpica i diastràtica*. Barcelona: Enciclopèdia Catalana.
- Bello, A. 1981 [1847]. *Gramática de la lengua castellana destinada al uso de los americanos, ed. crítica de R. Trujillo*. Santa Cruz de Tenerife: Instituto Universitario de Lingüística / Andrés Bello / Cabildo Insular de Tenerife.
- Benito, C. and Octavio de Toledo y Huerta, A. 2016. *En torno a haber: construcciones, usos y variación desde el latín hasta la actualidad*. Bern: Peter Lang.
- Bentivoglio, P. and Sedano, M. 1989. *Haber. ¿Un verbo impersonal? Un estudio sobre el español de Caracas*. In *Estudios sobre el español de América y lingüística afroamericana: Ponencias presentadas en el 45 Congreso internacional de americanistas*. Bogotá, Colombia, July 1985. Bogotá: Instituto Caro y Cuervo. 51-81.
- Blas Arroyo, J. L. 2016. Entre la estabilidad y la hipercorrección en un antiguo ‘cambio desde abajo’: *Haber* existencial en las comunidades de habla castellanenses. *Lingüística Española Actual* 38(1): 69-108.
- Blas Arroyo, J. L. 2015. The scope of language contact as a constraint factor in language change: the periphrasis *haber de + infinitive* in a corpus of language immediacy in modern Spanish. *International Journal of Bilingualism* 19(5): 499-524.
- Blas Arroyo, J. L. 2011. Deber (de) + infinitivo: ¿un caso de variación libre en español? Factores determinantes en un fenómeno de alternancia sintáctica. *Revista de Filología Española* 91(1): 9-42.

---

<sup>8</sup> About other linguistic outcomes of koineization in Spanish, see also Fontanella de Weinberg (1992), Granda (1994), Lipski (1994), Parodi (2001), Tuten (2003), among others.



- Blas Arroyo, J. L. 2008. The variable expression of future tense in Peninsular Spanish: The present (and future) of inflectional forms in the Spanish spoken in a bilingual region. *Language Variation and Change* 20(1): 85-126.
- Blas Arroyo, J. L. 2007. El contacto de lenguas como factor de retención en procesos de variación y cambio lingüístico. Datos sobre el español en una comunidad bilingüe peninsular. *Spanish in Context* 4(2): 263-291.
- Blas Arroyo, J. L. 1996. A propósito de un caso de convergencia gramatical por causación múltiple en el área de influencia lingüística catalana. Estudio sociolingüístico. *Cuadernos de Investigación Filológica* 21: 175-200.
- Bouzouita, M. and Pato, E. 2016. La pluralización de haber en español peninsular. *Proceedings of 21th Sociolinguistics Symposium*. Murcia, Spain, 15-18 June 2016. Universidad de Murcia (in press).
- Casals, D. and Cervià, C. 1999. Aproximació al fenomen de la concordança del verb haver-hi. *Xarxa Tècnica de Política Lingüística*. Barcelona: Generalitat de Catalunya. 11-21.
- Castillo Trelles, C. 2007. La pluralización del verbo haber impersonal en el español yucateco. *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics*. Somerville, MA: Cascadilla Proceedings Project. 74-84.
- Catalán, D. 1989. *El español: Orígenes de su diversidad*. Madrid: Paraninfo.
- Claes, J. 2014. A cognitive construction grammar approach to the pluralization of presentational *haber* in Puerto Rican Spanish. *Language Variation and Change* 26: 219-246.
- Claes, J. (in press). Cognitive and geographic constraints on morphosyntactic variation: The variable agreement of presentational *haber* in Peninsular Spanish. *Belgian Journal of Linguistics*.
- Company, C. 2006. *Sintaxis histórica de la lengua española. Primera parte: la frase verbal*. Vol. 2. México DF.: Universidad Nacional Autónoma de México/Fondo de Cultura Económica.
- D'Aquino Ruiz, G. 2004. *Haber* impersonal en el habla de Caracas. Análisis sociolingüístico. *Boletín de Lingüística* 21: 3-26.
- Díaz-Campos, M. 2003. The pluralization of *haber* in Venezuelan Spanish: A sociolinguistic change in real time. *IU Working Papers in Linguistics* 3(5): 1-13.
- Domínguez, C., Guzmán, B., Moros, L., Pabón, M., and Vilain, R. 1998. Personalización de *haber* en el español de Mérida. *Lengua y Habla* 3(1): 23-36.
- Fabra, P. 1956. *Gramàtica catalana*. Barcelona: Teide.
- Fontanella de Weinberg, M. 1992. *El español de América*. Madrid: Función MAPFRE.
- Freites-Barros, F. 2008. Más sobre la pluralización de *haber* impersonal en Venezuela. El estado Táchira. *Lingua Americana* 22: 36-57.
- Garachana, M. 1997. Acerca de los condicionamientos cognitivos y lingüísticos de la sustitución de *aver* por *tener*. *Verba* 24: 203-235.
- Gómez Molina, J.R. 2013. *El español de Valencia, Estudio sociolingüístico*. Bern: Peter Lang.
- Granda, G. 1994. *Español de América, español de África y hablas criollas hispánicas. Cambios, contactos y contextos*. Madrid: Gredos.
- Hernández, A. 2006. Posesión y existencia: la competencia de *haber* y *tener* y *haber existencial*, C. Company (ed), 1055-1160. México DF.: Universidad Nacional Autónoma de México/Fondo de Cultura Económica.
- IEC. 1992. Declaració arran de la llei sobre l'autoritat lingüística de l'IEC. *Documents de la Secció Filològica II*. Barcelona: Institut d'Estudis Catalans. 11-16.
- Illamola, C. 2015. *Contacto de lenguas y la expresión de la posterioridad temporal en el español de Cataluña*. PhD Thesis, Universidad de Barcelona.
- Lipski, J. 1994. *Latin American Spanish*. New York: Longman.
- Moll, F. 1968. *Gramàtica catalana. Referida especialmente a les Illes Balears*. Palma de Mallorca: Moll.
- Montes de Oca, M. P. 1994. La concordancia con *haber* impersonal. *Anuario de Letras* 32: 7-35.
- Parodi, C. 2001. Contacto de dialectos y lenguas en el Nuevo Mundo: La vernacularización del español en América. *International Journal of the Sociology of Language* 149: 33-53.

- Pato, E. 2016. La pluralización de *haber* en español peninsular. C. Benito and A. Octavio de Toledo y Huerta (eds), 357-391. Bern: Peter Lang.
- Pérez Martín, A. M. 2007. Pluralización de *había* en el habla de El Hierro: datos cuantitativos. *Revista de Filología de la Universidad de La Laguna* 25: 505-513.
- Pickering, M., and Ferreira, V. 2008. Structural priming: A critical review. *Psychological Bulletin* 134(3): 427-459.
- Poplack, S., Zentz, L. and Dion, N. 2012. Phrase-final preposition in Quebec French: An empirical study of contact, code-switching and resistance to convergence. *Bilingualism: Language and Cognition* 15(2): 203-225.
- Poplack, S. and Tagliamonte, S. 2001. *African American English in the diaspora: Tense and aspect*. Oxford: Blackwell.
- Quilis, A. 1983. *La concordancia gramatical en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas.
- Quintanilla, J. R. 2009. *La (des)pluralización del verbo haber existencial en el español salvadoreño: ¿un cambio de progreso?* PhD Thesis. Miami: University of Florida.
- Ramos-Alfajarín, J. R. 2001. El verb *haber-hi*: Evolució dels usos sintàctics. *Estudis Romànics* 23: 123-146.
- Real Academia Española and Asociación de Academias de la Lengua Española. 2009. *Nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- Rigau, Gemma. 1993. El comportamiento sintáctico de los predicados existenciales en catalán. *Revista de llengües i literatura catalana, gallega y vasca* 3: 33-53.
- Rivas, J. and Brown, E. 2012. Stage-level and individual-level distinction in morphological variation: An example with variable *haber* agreement. *Borealis: An International Journal of Hispanic Linguistics* 1(2): 73-90.
- Samper, J. A. 2012. La sociolingüística en España. *Español Actual* 98: 39-69.
- Samper Padilla, J. and Hernández Cabrera, C. 2012. En torno a los usos personales de *haber* en el español de Las Palmas de Gran Canaria. *Cum corde et in nova grammatica: Estudios ofrecidos a Guillermo Rojo*. Santiago de Compostela: Universidad de Santiago de Compostela. 743-754.
- Sanchis-Guarner, M. 1993. *Gramàtica valenciana*. Barcelona: Alta Fulla.
- Sinner, C. 2004. *El castellano de Cataluña*. Tübingen: Max Niemeyer Verlag.
- Solà, J. 1994. *Sintaxi normativa: estat de la qüestió*. Barcelona: Empúries.
- Universitat de València. 2011. *Gramàtica Zero*. València: Universitat de València.

José Luis Blas Arroyo  
 Universitat Jaume I  
 Av. de Vicent Sos Baynat s/n  
 12071 Castelló de la Plana (Spain)