A Cognitive Construction Grammar approach to the pluralization of presentational *haber* in Puerto Rican Spanish

JEROEN CLAES
University of Antwerp

ABSTRACT

In this paper, I present an analysis of the pluralization of *haber* 'there is/are' in Puerto Rican Spanish (e.g., *habían fiestas* 'there were parties') as an ongoing language change from below in which the impersonal argument-structure construction (<**AdvP** *haber* **Obj**>) is being replaced by a personal variant (<**AdvP** *haber* **Subj**>). Speakers pluralize presentational *haber* in about 41% of the cases, and linguistic conditioning factors are 'typical action chain-position of the noun's referent,' polarity of the clause, verb tense, comprehension-to-production priming, and production-to-production priming. I argue that the effect of these independent variables can be traced back to three cognitive factors: the preference for unmarked coding, statistical preemption, and structural priming. Social distributions can also be modeled in constructionist frameworks, with results for social class, formality, and gender advocating in favor of considering this variation as an ongoing change from below, which allows speakers to position themselves in terms of gender and social class.

THE PLURALIZATION OF *HABER*

In normative Spanish, presentational *haber* 'there is/are' is used in subjectless, impersonal constructions, exemplified in (1). This means that the nominal argument—in this case, *árboles* 'trees', *mangos* 'mangos', and *grosellas* 'gooseberries'—does not function as a subject, but rather as a direct object, as is shown by the fact that it pronominalizes as an accusative pronoun in example (2). As a result, in both examples, the verb presents default third-person singular agreement (e.g., Real Academia Española & Asociación de Academias de la Lengua Española, 2009:§41.6).

(1) Sí, porque tenía patio y todo. Y *había* árboles y cosas, había una, ha, hab, *había* a, árboles, árboles frutales. Sí, *había* mangos, *había* grosellas (SJ14H22 /SJ1672).[1]
     'Yes, because it had a garden and all. And *there were*$_{Imp}$ trees and stuff, there was a, ther, ther, *there were*$_{Imp}$ t, trees, fruit trees. Yes, *there were*$_{Imp}$ mangos, *there were*$_{Imp}$ gooseberries.'

(2) Y viví tres años en Nueva Orleans, que, digamos, yo no tuve problemas pero sé que
    los$_{Acc}$ *hay*, no (SJ19H22/SJ2231).
    'And I lived in New Orleans for three years, that, let's say, I never had any problems,
    but I know that *there$_{Acc}$ are$_{Imp}$*, right.'

However, speakers of many varieties of both Peninsular (e.g., Blas-Arroyo, 1995–
1996) and Latin American Spanish (e.g., DeMello, 1991) tend to establish verb
agreement with the noun phrase, as in (3), a phenomenon that is known in the
literature as "the pluralization of *haber*."

(3) E, no había, por ejemplo, nos, yo, nosotros vivíamos en una finca de veintisiete
    cuerdas. Era grande, enorme. . . . *Habían* mangos, *habían* piñas (SJ16H21/SJ1949).
    'E, there wasn't, for example, we, I, we lived on a farm of twenty-six acres. It was
    big, enormous. . . . *There were$_{Pers}$* mangoes, *there were$_{Pers}$* pineapples.'[2]

In Hispanic sociolinguistics, the pluralization of *haber* has received much
attention (e.g., Bentivoglio & Sedano, 1989, 2011; Brown & Rivas, 2012;
D'Aquino-Ruiz, 2004, 2008; Díaz-Campos, 2003; Freites-Barros, 2004;
Quintanilla-Aguilar, 2009; Rivas & Brown, 2012). In broad strokes, earlier
studies have shown that the phenomenon covaries with social factors and is
sensitive to the properties of the noun phrase (NP) and the verb tense, but the
cognitive constraints underlying these linguistic factors are yet to be uncovered.
Therefore, to deepen our understanding of *haber* pluralization, in this paper, I
will try to answer the following questions:

   1. What are the cognitive factors that constrain the pluralization of presentational
      haber in the Spanish of San Juan, Puerto Rico?
   2. What is the social distribution of *haber* pluralization in San Juan, Puerto Rico?
   3. How can the phenomenon, these constraints and its social distribution be
      modeled in Cognitive Construction Grammar?

THEORETICAL PERSPECTIVE

Cognitive Construction Grammar

Cognitive Construction Grammar (CCxG) is particularly fit to model
sociolinguistic data for a number of reasons. First, it is a usage-based syntactic
theory, which means that language is taken to represent variability and change at
any given moment in time and that knowledge of language is assumed to
include knowledge of the quantitative patterns in this variation (Bybee &
Beckner, 2010:830). Second, CCxG assumes an encyclopedic/experiential view
of semantics, which means that recurrent aspects of the usage events in which a
particular structure is used are recorded as part of its meaning (Bybee &
Beckner, 2010:846; Goldberg, 2006:10; Langacker, 1987:63). Third, this
theoretical approach includes a commitment to producing corpus (see, e.g.,

Goldberg, 2011), psycholinguistic (see, e.g., Bencini & Goldberg, 2000; Goldberg, 2006:69–92; Goldberg, Casenhiser, & Sethuraman, 2004), and even neurological (see, e.g., Allen, Pereira, Botvinick, & Goldberg, 2012) evidence for the constructs the framework depends upon, which ensures that the model is psychologically plausible.

CCxG claims that language consists entirely of form-meaning pairs of different degrees of abstractness, called "constructions" (Goldberg, 1995:7, 2006:45). As a result, no differences are posited between rules/generalizations (e.g., grammatical patterns such as transitivity) and lexical items/idiosyncratic patterns (e.g., idioms), because both are captured with the same principles (Goldberg, 2009:94; Langacker, 1987:28). In other words, in this framework, the grammar is depicted as a network of interconnected construction schemas that offers speakers symbols to express conceptualizations (Langacker, 1987:409).

For example, in the short utterance given in (4), no less than five constructions can be distinguished: *Juan, coger* 'take', *el* 'the', *vaso* 'glass', and the transitive pattern <**Subject** Verb **Object**>.[3]

(4)   Juan coge el vaso.
         'John takes the glass.'

Contrary to low-level constructions (or words) such as *Juan*, *coger*, *el*, and *vaso*, the transitive pattern possesses a more schematic meaning, which, according to Langacker (1991:302) relates to the conceptualization of a transfer of energy from the agent/subject to the patient/object participant. Based on this abstract meaning, the argument-structure construction determines that there will be two *profiled* argument roles (i.e., without these two participants, events of this type cannot take place): agent and patient. In addition, the schema determines the mapping of these roles onto syntactic functions and specifies the information structure of the clause. In sum, in this framework, argument-structure constructions, rather than individual verbs, are taken to be the determinants of clause structure.

A Cognitive Construction Grammar approach to the pluralization of haber

Within this theoretical setting, I would like to propose the hypothesis that the pluralization of *haber* results from a competition within the grammar[4] between two variants of the presentational construction with this verb, which allows speakers to position themselves in the social landscape. Formally, both are characterized by an adverbial phrase (AdvP) and an NP argument. In the impersonal argument-structure construction, the nominal is a direct object; in the personal variant, the NP functions as a subject.

When it comes to semantics, both construction schemas encode the POINTING-OUT event type proposed by Lakoff (1987:490):

> It is assumed as a background that some entity exists and is present at some location in the speaker's visual field, that the speaker is directing his attention at it, and that the hearer is interested in its whereabouts but does not have his attention focused on it, and may not even know that it is present. The speaker then directs the hearer's attention to the location of the entity (perhaps accompanied by a pointing gesture) and brings it to the hearer's attention that the entity is at the specified location.

Because both variants refer to the same event type, the constructions also assign the same semantic role to the NP slot. That is, as the referent of the NP is simply present in the scene that is being presented through the construction, it is probably safe to assume that it is assigned the "zero" semantic role (Langacker, 1991:288).

As these constructions are used to introduce referents into discourse, the nominal argument cannot be "given" with respect to the hearer's beliefs, consciousness, or world-knowledge (i.e., "hearer-old" in Prince's [1992] terminology). This is not to say that it is impossible to use definite NPs in *haber* clauses. Rather, Ward and Birner (1995) showed that the nominal argument of presentational constructions can be definite and/or discourse-old as long as it can somehow be construed as hearer-new. This possibility is exemplified in (5), where the "reminder" (Lakoff, 1987:561–562; Ward & Birner, 1995:730–731) presentational reading emerges when the speaker treats *condominios* 'condominiums' as hearer-new, although this referent has already appeared in discourse and the interview takes place in such a building.

(5)   Inv: ¿Este, cuando se mudó aquí, habían cosas a la que tuvo que acostumbrarse?
      Subj: ¿Tales cómo?
      Inv: No sé por ejemplo. E, este, no sé, cosas.
      Subj: Este, bueno, no, me tuve que acostumbrarme a vivir en un <u>condominio</u> cuando yo viví en una casa. . . .
      Inv: ¿Y usted recuerda como la ciudad era antes? O sea, cuando era niña.
      Subj: Cuando yo era niña, sí. No *habían* estos <u>condominios</u>, desde luego (SJ01M22/ SJ07).
      Inv: 'Eh, when you moved here, were there things to which you had to get used to?'
      Subj: 'Such as?'
      Inv: 'I don't know, for example. E, eh, I don't know, things.'
      Subj: Eh, well, no, I had to get used to living in a <u>condominium</u>, when I had lived in a house. . . .
      Inv: And do you remember how the city was like before? That is to say, when you were a child.
      Subj: 'When I was a girl, yes. *There weren't$_{Pers}$* these <u>condominiums</u>, of course.'

Concerning the adverbial phrase that frequently appears with presentational expressions (e.g., Lyons, 1967), Lakoff (1987:542–543) argues that in the English presentational *there* construction, the adverbial denotes the nature of the "mental space," that is, "partial assemblies constructed as we think and talk for purposes of local understanding and action" (Fauconnier, 2007:351), that is set up by the expletive (the "space builder"). Examples (6) to (8) give reason to

believe that in *haber* clauses the adverbial phrase fulfills a similar function. However, the difference between English and Spanish presentationals appears to be that, as the *haber* construction does not involve an expletive, the adverbial does not only specify the nature of the mental space, it also functions as the space builder (i.e., it sets up the mental space).

(6) <u>En mi salón</u> *habían* treinta o treinta y un niños, sí (SJ12M12/SJ1404).
     '<u>In my classroom</u> *there were_{Pers}* thirty or thirty-one children, yes.'

In contrast, without the adverbial phrase, as in (7), the expression localizes the referent of the NP in the current "base space," that is, "the mutually known world of the interlocutors" (Croft & Cruse, 2004:33), or a previously built mental space, such as the time when the speaker was in high school in (8).

(7) *Hay* veces que tú sales a la calle y ves personas jóvenes hablando con palabras feas (SJ01M22/SJ60).
     '*There are_{Imp}* times that you go out on the streets and you see young people talking with ugly words.'
(8) E, pero yo diría que, que las electivas, las, los cursos electivos probablemente <u>en la escuela superior</u> eran los más que me gustaban. . . . *Podrían haber* días en que yo tenía dos horas libres entremedio (SJ13H11/SJ1566).
     'E, but I would say that, that the electives, the, the elective courses, probably <u>in high school</u> were what I liked most. . . . *There could be_{Pers}* days that I had two hours of free time in between.'

However, the grammatical status of the adverbial phrase does not imply that it is completely optional, because its absence makes the expression uninformative when the NP argument cannot be located in the base space or a previously evoked mental space (Hernández-Díaz, 2006:1130–1132; Lyons, 1967; Meulleman & Roegiest, 2012). If we consider, for example, the last utterance of (8) in isolation of the evocation of the "high school" mental space, we cannot help but wonder against which setting we have to interpret the assertion. Therefore, the adverbial as well is most accurately described as *profiled*.

A Cognitive Construction Grammar approach to the social distribution of alternating constructions

Before we start exploring how the social distribution of alternating constructions can be incorporated in CCxG, first, it is important to recall that constructional variation will only be interpreted in social terms if the variants refer to exactly the same event type (Lavandera, 1978). With these limitations, because CCxG is a usage-based framework, it is hardly controversial to claim that speakers notice that, in the absence of other differences of meaning between two formally distinct options to refer to a given event type, individuals instantiating a particular social type use a certain alternative more often and that they store this social distribution as part of the meaning of the relevant alternating constructions

(Bybee & Beckner, 2010:830, 846; Goldberg, 2006:10; Langacker, 1987:63). Therefore, it is rather unproblematic to suppose that if two constructions refer to exactly the same event type, each will include a social distribution specification, such that the use of one of these constructions at a certain rate where the constructional semantics is observed will be interpreted as signaling social subgroup membership ("1st-order index" in Silverstein's [2003] terms) and, potentially, everything there is to know about this social subgroup ("1 + *n*th order index" in Silverstein's [2003] terms).

All of this knowledge will not be activated every time a particular distribution is observed in a particular context. Rather, the context will activate or background certain things language users know about a particular social group (see Langacker, 1987:154–166), which explains how the multiple context-dependent interpretations of sociolinguistic variables arise (e.g., Eckert, 2008). In other words, CCxG can readily accommodate the social meaning of alternating constructions that refer to the same event type. In this paper, I will only be concerned with the links Puerto Ricans establish between the distribution of presentational *haber* and (knowledge on) social groups.

HYPOTHESES AND CODING

In this study, I will explore the following main hypothesis:

> In San Juan, Puerto Rico, the pluralization of presentational *haber* corresponds to a slowly advancing language change from below: the personal presentational schema with *haber* (<**AdvP** *haber* **Suj**>) is replacing the impersonal presentational construction with this verb (<**AdvP** *haber* **Obj**>). The variants only differ with regard to the syntactic function of the NP (impersonal variant: object; personal variant: subject) and the social groups associated with their relative frequencies.

This is a very abstract hypothesis, which on its own, does not allow for any predictions. However, through reference to three cognitive factors (markedness of coding, statistical preemption, and structural priming) and Labov's (2001) principles of linguistic change, a list of more detailed extrapolations can be drawn up.

Markedness of coding

Against the background of the main hypothesis, a first cognitive factor that can be expected to condition *haber* pluralization is the preference for unmarked coding, that is, the tendency for a "notion approximating an archetypical conception [to be] coded linguistically by a category taking that conception as its prototype" (Langacker, 1991:298). This leads to the first hypothesis:

> *Hypothesis 1. Markedness of coding.* A more prototypical subject will more likely be coded as a subject. Conversely, a more prototypical object will more likely be coded

as an object. This will lead speakers to select the personal variant more often with NP arguments that are more similar to prototypical subjects and the impersonal variant with NP arguments that are more similar to prototypical objects.

This hypothesis raises the question as to which features can portray the NP of presentational *haber* as a more prototypical subject or object. In this regard, the typological literature suggests that prototypical subjects are agents in events (Comrie, 1989:66; Keenan, 1976:321; Lakoff, 1987:64; Langacker, 1991:294). However, as we have seen in the previous section, the nominal argument of presentational *haber* is clearly not an agent, because it is merely present in a static situation. Still, it is inarguably the case that some entities (say, *driver*) are intrinsically more likely than others (say, *victim*) to be agents in events. Therefore, with constructions such as presentational *haber*, which do not explicitly construe the nominal argument as an agent or a patient, entities such as *driver* may be perceived as more potential agents, and, thus, as more prototypical subjects (Langacker, 1991:294) than entities such as *victim.*

In cognitive linguistics, the semantic roles "agent" and "patient" are defined in relation to what Langacker (1991:283–285) called the "canonical event model" or the "action-chain model": the head initiates physical activity, resulting "through physical contact, in the transfer of energy to an external object" (Langacker, 1991:285) and an internal change of state of that entity, the tail of the chain. The semantic roles of agent and patient, in turn, are defined as, respectively, "action-chain head," and "action-chain tail." Additionally, events take place in a particular setting, such that the event model minimally includes three elements: action-chain head/agent, action-chain tail/patient, and setting.

To test the first hypothesis, I coded the data for the typical action-chain position of the entity indicated by the noun, for which I relied on the answers to the question in (9).

(9) Is the referent of the noun highly likely to cause an internal change of state to a second entity without being affected by a third entity first?
     Yes: Typical action-chain head (i.e., more potential agent; e.g., *temblor* 'earth quake', *madre* 'mother', *carro* 'car')
     No: Typical action-chain setting or tail (i.e., more potential setting or patient; e.g., *actividad* 'activity', *víctima* 'victim', *daño* 'damage')

I also evaluated the influence of the polarity of the clause. As we will see in the section dealing with the results for markedness of coding, this factor can be connected to subject/objecthood because, with presentational *haber*, negative polarity originates a nonspecific indefinite reading for the NP argument, which is prototypical of direct objects (Croft, 2003:132).

Statistical preemption

A second likely cognitive constraint is statistical preemption. That is, in a usage-based framework such as CCxG, forms that occur in more than one construction

are taken to have stronger independent cognitive representations than those that only occur in one construction (Bybee, 2003:619; Croft & Cruse, 2004:292). In particular, if a form presents high token frequency in one construction schema, but only occurs sporadically in other patterns, it is taken to be stored as a partially lexically filled instance of this construction with a much stronger representation than both the independent form and the abstract pattern (Goldberg, 1995:79; Langacker, 1987:59–60, 1991:48). As expressions based on this partially prefabricated unit require less constructive effort than do expressions involving the fusing of more abstract construction patterns and words, the prefab disfavors the use of a novel expression based on a competing construction schema that shares the same pragmatic and semantic constraints (Goldberg, 2006:94, 2009:102–103, 2011). This is called "statistical preemption."

In other words, if certain verb forms of *haber* occurred mainly in the <**AdvP** *haber* **Obj**> pattern before <**AdvP** *haber* **Subj**> emerged as a conventional alternative for referring to POINTING-OUT, upon actuation of the change, the personal variant would not have been used frequently to refer to a conceptualization that involved this event type in the temporal setting encoded by those tenses. In subsequent generations, repetition usually ensures that a given distribution remains intact (Bybee, 2006:715). This leads to:

> *Hypothesis 2. Statistical preemption*
> *Hypothesis 2A.* If the third-person singular form of a particular tense of *haber* was frequently used outside of the impersonal construction before the actuation of the change, this verb tense will favor the personal variant.
> *Hypothesis 2B.* The other verb tenses will disfavor the personal construction, provided the conceptualization can be expressed with an entrenched instance of the impersonal construction.

Because these hypotheses assume that the preempting effect of certain verb tenses is a function of the degree of entrenchment of a particular form in the impersonal construction, the following prediction follows quite naturally:

> *Hypothesis 2C.* When the need to encode an aspectually/modally more complex conceptualization forces speakers to construct a new expression involving a compound verb form rather than retrieving a partially prefabricated expression from long-term memory, the tenses that were predominantly used in the impersonal construction before the actuation of the change will favor the personal variant.

The testing of these hypotheses raises two questions: When did the variation that affects presentational *haber* emerge as a community-wide phenomenon? Which tense forms of the verb enjoyed a relatively high frequency in a variety of constructions before this happened? The answer to the first question can only be tentative, as it is virtually impossible to know for a fact when and how the variation that affects presentational *haber* started exactly in Puerto Rican Spanish. For Buenos Aires, Fontanella de Weinberg (1992b) showed that the

alternations between singular and pluralized presentational *haber* are already present in written discourse from the 18th century onward. Because there is usually a considerable lag between the actuation of a linguistic change and its trickling down into writing, the variation probably emerged somewhere in the 17th century, which blends in nicely with research in historical linguistics that shows the most prominent features of American Spanish to stem from a *koiné* variety that emerged through language and dialect contact during that century (Fontanella de Weinberg, 1992a:42–54).

Therefore, to formulate an answer to the second question, in Appendix A, I tabulated the distribution of the third-person singular forms of *haber* in a 16th-century Latin American subset of the Spanish Royal Academy's *Corpus Diacrónico del Español* (Real Academia Española, 2008). The results show that before presentational *haber* became involved in large-scale variation in American Spanish, the present and preterit tense forms occurred primarily in presentational clauses, which suggests that the most salient representations of these forms were <**AdvP** *hay* **Obj**> and <**AdvP** *hubo* **Obj**>. The other tense forms, on the other hand, were either used more productively (spread over more different constructions) or are restricted to a very low frequency in the corpus ($n < 100$), which indicates that their independent forms probably also constituted their strongest cognitive representations. This distribution suggests two relevant tense groups: synthetic expressions in present and preterit tense versus all others.

Structural priming

Furthermore, sociolinguistic and psycholinguistic research has shown that once speakers have used or processed a constructional pattern, they tend to recycle it in the following stretches of discourse (Bock, Dell, Chang, & Onishi, 2007; Goldberg, 2006:120–125; Labov, 1994:550), which is known in the psycholinguistic literature as "structural priming." Therefore, if the variation amounts to a competition between two argument-structure constructions, as the main hypothesis claims, I expect that:

> *Hypothesis 3. Structural priming.* The earlier mention of one of the variants in discourse will promote the use of the same variant in the next occurrence, regardless of variations in tense, aspect or mood.

To test this hypothesis, all tokens were coded for the last variant of the *haber* structures that was used by the speaker (production-to-production priming) and the investigator (comprehension-to-production priming). Additionally, I binned together the occurrences in 5-clause lag groups (e.g., lag: 0–4 clauses; 5–9 clauses, etc., up to a 20-clause lag) and I separated the occurrences in which speakers repeated the verb form and the argument structure from those in which they only repeated the argument structure. This resulted in a total of 17 factors for both factor groups. However, as the initial results displayed a similar priming effect until reaching a

20-clause lag, independently of whether speakers would repeat the specific verb form or not, I collapsed the factors into the following broader categories: first occurrence/ distance 20+ clauses, primed with the personal presentational *haber* construction, primed with the impersonal presentational *haber* construction.

Principles of linguistic change

The claim that the variation reflects a slowly advancing language change from below also predicts that the alternations will display patterns of social and stylistic covariation typical of this type of linguistic evolution. More specifically, the "apparent-time" construct (Labov, 1994:43–72) predicts that:

> *Hypothesis 4.* The youngest speakers will favor the personal construction, whereas older speakers will make (slightly) more use of the impersonal construction.

By the same token, Labov's Gender Principle (Labov, 2001:292) predicts that:

> *Hypothesis 5.* In comparison to men of the same social characteristics, women will use the personal construction more often.

However, because gender-differentiated behavior is also found for changes from above (Labov, 2001:274) and because the possibility of age-graded behavior always exists for apparent-time distributions (Labov, 1994:45), more evidence will be needed before we can confidently conclude that this alternation constitutes a linguistic change from below. In this regard, the most conclusive indication of an ongoing change from below seems to be the curvilinear pattern (Labov, 2001:188), which entails the following description:

> *Hypothesis 6.* The middle class will show higher frequencies of use of the personal construction than will the groups of lower and higher social status.

In addition, in change from below, the innovative variants usually show no style-shifting or increase in frequency when formality rises (Labov, 2001:95–101), which leads to the following hypothesis:

> *Hypothesis 7.* When formality increases, the frequency of the personal variant will not decrease.

Finally, Labov (1972:138) observed that highly educated speakers tend to conform to the supralocal prestige norms, which suggests that:

> *Hypothesis 8.* Higher educational achievement will favor the impersonal variant, whereas a shorter formal education will promote the personal variant.

To test these predictions, a mixed-effect logistic regression analysis was performed. However, before we can go on to discuss the results, a few words are in order

concerning the methods that were used in compiling the corpus from which this investigation draws.

METHOD

The sample

The analyses are based on a 200,000-word corpus of sociolinguistic interviews (about 24 hrs. of speech) that was recorded in March and April 2011 with 24 native speakers of Puerto Rican Spanish who reside in the San Juan metropolitan area. As is shown in Table 1, the corpus is stratified by three social parameters: academic achievement (university degree vs. no university degree), age (20–35 years vs. 55+ years), and gender (male vs. female).

To have access to more variable contexts and to investigate whether the variation is sensitive to style-shifting, as is predicted by Hypothesis 7, the author, who was also the fieldworker, structured the recordings into three sections:

1. *Interview*. Speakers were interviewed for about 30 mins. on a variety of topics related to their day-to-day life. The interview format was loosely based on the conversational modules that are found in Moreno-Fernández (2003), Quintanilla-Aguilar (2009), and Tagliamonte (2006:App. B). In addition, a set of questions containing the variable, such as (10), was included to investigate comprehension-to-production priming effects. In these questions, the variants were used randomly.

   (10) ¿Cuántos estudiantes *podía haber/podían haber* en tu época?
   'How many students *could there have been$_{Imp}$/could there have been$_{Pers}$* in your time?'

2. *Reading passage.* After the first part, speakers were instructed to read out a two-page children's story (*Juan Sin Miedo*, *'John Without Fear'*) in which 31 selection contexts, as in (11), had been inserted (20 trials, 11 fillers).

   (11) En una pequeña aldea, *había/habían* un anciano padre y sus dos hijos . . .
   'In a small village, *there were$_{Imp}$/there were$_{Pers}$* an old father and his two sons . . .'

3. *Questionnaire.* Finally, the interviewees were instructed to read out loud a questionnaire consisting of 45 items (32 trials, 13 fillers) preceded by a

TABLE 1. *Configuration of the sample*

Academic achievement	20–35 Years		55+ Years		Total
	Male	Female	Male	Female	
Without university degree	3	3	3	3	12
With university degree	3	3	3	3	12
Total	6	6	6	6	24

description that evoked the right pragmatic context for the interpretation of the test sentence, as can be seen in (12). For both the reading passage and the questionnaire, whenever a speaker had difficulties reading, the author read the tests to him/her and asked him/her which form he/she preferred.

(12)   A Inés le acaban de robar el carro, que tenía aparcado en algún callejón obscuro. Aunque no es la cosa más sensata que se pueda hacer, una amiga trata de consolarla diciendo: "No es culpa tuya, es que siempre _____ unas personas malas."

a. habrá   b. habrán

'They've just stolen Ines' car, which she had parked in a dark alley. Although this is not the most intelligent thing to do, a friend tries to comfort her, saying: "It is not your fault, _____ always be a few bad people."'

a. There will$_{Imp}$ b. There will$_{Pers}$

Poststratification: Social class

After transcription, this sample was poststratified by social class, a combination of the speaker's weighted scores for the following variables:

1. Academic achievement: 0: less than high school; 0.5: high school; 1: university.
2. Housing conditions: 0: house/apartment in poor condition; 0.5: house/apartment in good condition with up to two bedrooms; 1: house/apartment in good condition with three or more bedrooms.
3. Profession: 0: for example, street vendors, unskilled workers; 0.25: for example, shop-holders, secretaries, mechanics; 0.5: for example, college-educated professionals, teachers; 0.75: for example, liberal professions, mid-level managers, university professors; 1: senior managers/officials (see Moreno-Fernández, 2003, for the full scale).

Rather than relying on my own intuitions, in order to establish the relative importance of each factor for social prestige, after the recording sessions, I presented the informants with an additional questionnaire with the instruction to rank academic achievement, housing, and profession by their importance for social status. Subsequently, each speaker's score for a given variable was first multiplied by the average impact factor that had been established for that variable, and these weighted scores were summed together. Then, the raw numbers were converted into percentages of the maximally possible score. Finally, speakers were grouped together in three equal bins by their score on the prestige index: those who scored 0% to 33% were considered to be members of the lower class (7 speakers), those who scored 34% to 66% were considered to belong to the middle class (7 speakers), and those who scored 67% to 100% were considered to be members of the higher classes (10 speakers).

The variable context

Earlier studies did not find any variation with present-tense *hay*. However, my corpus provides a handful of tokens of *hayn*, such that the alternation between *hay* and *hayn* cannot be excluded from the scope of this investigation. In contrast, some previous surveys (e.g., DeMello, 1991; Quintanilla-Aguilar, 2009) have included cases in which *haber* agrees with first-person plural NPs. Nevertheless, as examples such as (13) and (14) display a clear difference in meaning, personalized first-person plural presentational *haber* was considered to be outside the envelope of variation.

(13)   Y *habían* bastantes, bastantes estudiantes en, e, los salones de clase (constructed
        example).
        'And *there were$_{Pers}$* plenty, plenty of students in, e, the classrooms.'
(14)   Y *habíamos* bastantes, bastantes estudiantes en, e, los salones de clase (SJ03H22).
        'And *we were* plenty, plenty of students in, e, the classrooms.'

Statistical tools

After selecting and coding all the cases of presentational *haber* + plural NP, I performed a mixed-effect logistic regression analysis with Rbrul (Johnson, 2009). In order to evaluate the hypotheses, I tested the statistical effect of 11 factor groups, which either instantiate the three cognitive constraints referred to in Hypotheses 1 to 3 (markedness of coding, statistical preemption, and structural priming) or try to establish the social distribution of *haber* pluralization. Additionally, the individual speakers and the lemmas of the NPs' head nouns were included in the runs as grouping factors. However, as these were collinear, I ran parallel analyses for the grouping factors, and I only report as statistically significant those fixed effects that proved to condition the variation for all speakers and all lexical items.

RESULTS

Table 2 shows that, using these methods, I collected 1655 observations of presentational *haber* + plural NP, of which 41% are pluralized. This number is significantly lower than the 50% to 60% or even 80% that is observed in, respectively, Caracas (Bentivoglio & Sedano, 1989; D'Aquino-Ruiz, 2004; Díaz-Campos, 2003) and San Salvador (Quintanilla-Aguilar, 2009). However, the difference between the figures that are reported here and the findings of earlier investigations is due to the fact that I did not exclude the present-tense forms *hay-hayn*. Without these two forms, the frequency rises to 54% ($n = 661$ of 1217), which is significantly higher than the 44% ($n = 83$ of 190) of plural forms obtained by Brown and Rivas (2012:329) for Caguas, Cayey, and San Juan, but these fluctuations could be due to methodological differences and/or

TABLE 2. *Distribution of the singular and pluralized forms of* haber

	n	%
Plural	684	41.3
Singular	971	58.7
Total	1655	100

the fact that these investigators base their conclusions on a rather limited number of tokens.

In what follows, I will discuss the results in the order of the hypotheses they are meant to test, starting with those relative to the properties of the NP (i.e., typical action-chain position and polarity, as well as two additional factors).

Cognitive factors

The preference for unmarked coding. Table 3 indicates that, as hypothesized, nouns that refer to typical action-chain heads/more potential agents (i.e., entities that are highly likely to cause an internal change of state to a second entity without being affected by a third entity first), as in (15), favor *haber* pluralization, whereas nouns that refer to typical tails of an action chain, as in (16), or its setting (17) clearly disfavor the personal presentational *haber* construction.

(15)   For example, humans such as *madre* 'mother', natural phenomena such as *huracán* 'hurricane', self-propelling objects such as *carro* 'car'.

(16)   For example, tangible objects such as *libro* 'book', animate beings that undergo an action, e.g., *víctima* 'victim', *invitado* 'invitee'.

(17)   For example, *lugar* 'place', *año* 'year', event nouns such as *actividad* 'activity', *discusión* 'discussion'.

In contrast, earlier studies on *haber* pluralization have claimed that human versus nonhuman reference (e.g., Bentivoglio & Sedano, 1989, 2011) or the noun's proportion of subject use are the relevant constraints connected to the nominal argument (Brown & Rivas, 2012). More particularly, Bentivoglio and Sedano (1989), among others, have shown that human-reference NPs favor *haber* pluralization. In turn, Brown and Rivas (2012) have shown that *haber* pluralization occurs more often with nouns that are frequently used as subjects in Spanish. However, I would like to argue here that these results are actually epiphenomenal of differences in typical action-chain position.

In this regard, although the typological literature shows that animate-reference nouns are more likely to be used as subjects (Ashby & Bentivoglio, 1993:71; Croft, 2003:130; Du Bois, 1987:840; Keenan, 1976:321; Langacker, 1991:306–307), it should also be observed that animate reference is only connected to

TABLE 3. *Significant constraints on the pluralization of* haber: *Number, percentage, and weights for the personal presentational construction with* haber

Factor groups	n/n	%	Weight[a] (Nouns)	Weight[b] (Speakers)
Verb tense				
All others	622/1014	61.3	.82	.81
Expressions in the present or preterit tense without modal or aspectual auxiliaries	62/641	9.7	.18	.19
Range			*64*	*62*
Production-to-production priming				
Personal variant	352/558	63.1	.66	.65
First occurrence/Distance 20+ clauses	88/246	35.8	.46	.45
Impersonal variant	244/851	28.7	.38	.39
Range			*28*	*26*
Comprehension-to-production priming				
Personal variant	92/175	52.6	.63	.64
First occurrence/distance 20+ clauses	562/1355	41.5	.44	.43
Impersonal variant	30/125	24.0	.44	.43
Range			*19*	*21*
Typical action-chain position of the noun's referent				
Head	338/730	46.3	.58	.58
Tail and setting	346/925	37.4	.42	.42
Range			*16*	*16*
Polarity of the clause				
Positive	559/1225	45.6	.57	.57
Negative	125/430	29.1	.43	.43
Range			*14*	*14*
Gender				
Female	375/836	44.9	.55	.55
Male	309/819	37.7	.45	.45
Range			*10*	*10*
Speaker age: Social class (interaction group)				
20–35 years				
lower class	125/278	45.0	.56	.56
middle class	121/273	44.3	.47	.49
upper class	132/329	40.1	.46	.47
Range			*10*	*9*
55+ years				
middle class	104/179	58.1	.71	.70
lower class	46/114	40.4	.47	.46
upper class	156/482	32.4	.33	.32
Range			*38*	*38*

Note: AIC = Akaike information criterion.
[a]Log likelihood: −759.778; deviance: 1519.556; AIC: 1549.556; input probability: 0.35; [b]Log Likelihood: −774.559; deviance: 1549.19; AIC: 1579.19; input probability: .32.

subjecthood through the tendency for animate beings to be agents in events (Langacker, 1991:322). In other words, most animate-reference nouns also typical action-chain heads. Similarly, chances are high that nouns of high proportion of subject use also refer to more typical agents.

    To gain more insight in this matter, I ran two additional, parallel models besides the model presented in Table 3: one in which the noun's typical action-chain

position had been replaced with, respectively animacy (i.e., animate vs. inanimate reference) and one in which it had been replaced with Brown and Rivas' (2012) factor group "proportion of noun use as subject" (low to mid vs. high), which was established by tracing the grammatical distributions of the nouns' lemmas in 200-token samples drawn from Davies' (2002) *Corpus del español*.

   Although the results in Table 4 show that pluralized *haber* indeed occurs more often with animate-reference NPs and nouns that are frequently used as subjects, this does not necessarily mean that animacy (e.g., Bentivoglio & Sedano, 1989) or the proportion of subject use of a noun (Brown & Rivas, 2012) condition *haber* pluralization. Rather, as I suspected, a close comparison of the lemmas of the corpus (see Appendix B and examples (18) and (19)) unveils that most animate-reference nouns and nouns that are frequently used as subjects refer to more typical action-chain heads.

(18) Hubo una vez que salí del trabajo mi esposo me fue a recoger y veníamos por el expreso y *habían* muchas <u>motoristas</u> al f, al frente y uno fue a esquivar al otro y se dio, le dieron con la goma posterior y se cayó (SJ05M12/SJ660).
      'There was one time when I left the office and my husband came to get me and we were driving on the express road and *there were$_{Pers}$* a lot of <u>motor cycle drivers</u> in f, in front and one went to evade the other and he hit, they hit him with the back tire and he fell.'
(19) Inv: E, bueno, y ahora tú estás viviendo en una casa de madera. ¿Nunca te da miedo con los huracanes? Porque esto no es . . .
      Subj: ¿Seguro, seguro?
      Inv: ¿No?
      Subj: Bueno, este, fíjate ahora, ahora no estoy pensando en huracanes porque me mudé en una época que no es de huracán. Pero sí, *han habido* <u>terremotos</u> (SJ12M12/SJ1412).
      Inv: 'E, well, and now you are living in a wooden house. Aren't you ever afraid of

TABLE 4. *Results for animacy and proportion of noun use as subject in parallel models: Numbers, percentages, and weights for the personal presentational construction with* haber

Factor groups	*n/n*	%	Weight (Nouns)	Weight (Speakers)
Animacy[a]				
Animate	308/692	44.5	.58	.56
Inanimate	376/963	39.0	.42	.44
Range			*16*	*12*
Proportion of noun use as subject[b]				
High	165/321	51.4	.58	.55
Low to mid	519/1334	38.9	.42	.46
Range			*16*	*9*

Note: As signaled in text, I permuted action-chain position for either animacy or noun proportion of subject use and I ran the model again. AIC = Akaike information criterion.
[a]Nouns: deviance: 1520.113; AIC: 1550.113; input probability: .35; speakers: deviance: 1557.986; AIC: 1587.986; input probability: .32. [b]Nouns: deviance: 1524.264; AIC: 1554.264; input probability: .38; speakers: deviance: 1567.728; AIC: 1597.728; input probability: .33.

hurricanes? Because it is not . . . '
Subj: 'Safe-safe?'
Inv: 'No?'
Subj: 'Well, eh, you figure, now, now I'm not thinking of hurricanes because I
moved at a nonhurricane time of the year. But *there have been*$_{Pers}$ earthquakes.'

Moreover, the model fit indicators displayed in Table 5 show that the fit of the
model with typical action-chain position is significantly better than that of the
other two (see Baayen, 2008:275–276 for the procedure and the necessary R
code). These and the other findings presented in this section strongly suggest
that neither proportion of noun use as subject nor animacy is a relevant
restriction. Rather, as I have argued when introducing Hypothesis 1, the true
constraint connected with the NP appears to be the typical position of its referent
on the action chain and the tendency to encode more potential agents as subjects
(i.e., the preference for unmarked coding).[5] Analogously, in lieu of supporting
Rivas and Brown's (2012:87) claims that temporal persistence (in terms of
independent existence and reference) is a feature of prototypical subjects and that
"stage-level nouns" (i.e., nouns that refer to entities with a conceivable
beginning and end, such as *parrandas* 'surprise-party visits' in (20)) disfavor
haber pluralization because they are not temporally persistent, the results of this
study suggest that these authors' findings as well are epiphenomena of
differences in typical action-chain position, as such nouns (e.g., event nouns,
temporal nouns, deverbal nouns) (Rivas & Brown, 2012:81) refer to typical
action-chain settings rather than heads.

(20)   Sin embargo, en los pueblos *hay* parrandas siempre (SJ04M22/SJ446).
        'Nevertheless, in the smaller towns, *there are*$_{Imp}$ surprise-party visits all of the
        time.'

Finally, Prince (1992:299–300) observed that the NP of an affirmative
presentational clause can only be interpreted as referring to "a specific instance

TABLE 5. *Comparison of model fit indicators for the three models*

	Akaike's information criterion		Bayesian information criterion		Deviance	
	Nouns	Speakers	Nouns	Speakers	Nouns	Speakers
Model with typical action-chain position of the noun's referent	1549.6	1579.2	1630.7	1660.4	1519.6	1549.2
Model with animacy	1550.1	1588.0	1631.3	1669.2	1520.1	1558.0
Model with proportion of noun subject use	1554.3	1597.7	1635.4	1678.9	1524.3	1567.7

Note: Likelihood ratio tests indicate $p = .001$ for the differences in model fit.

or token whose identity is unknown to the hearer and possibly also the speaker" (Croft, 2003:132) or, in terms of the definiteness/specificity hierarchy given in (21), as being specific indefinite. Under negative polarity, in contrast, the NPs reference is suspended (Keenan, 1976:318), such that it becomes "identifiable only as a type, not as a specific instance or token" (Croft, 2003:132), in other words, it is interpreted as a nonspecific indefinite NP.

(21) Definite > specific indefinite > nonspecific indefinite (Langacker, 1991:308).

As nonspecific indefinite is the pragmatic definiteness/specificity value associated with prototypical direct objects (Croft, 2003:132; Du Bois, 1987:847; Keenan, 1976:319; Langacker, 1991:308), negation may cause speakers to categorize the nominal argument more often as such, which could explain why Table 3 shows speakers to disfavor <**AdvP** *haber* **Subj**> in negative clauses.[6]

Statistical preemption.   Table 3 shows that the distributional pattern presented in the section concerned with Hypothesis 2 (i.e., synthetic expressions in present and preterit tense vs. all others) turns out to be the strongest overall predictor for the variation. In addition, the verb tenses for which an entrenched instance of <**AdvP** *haber* **Obj**> was posited (i.e., present and preterit) disfavor <**AdvP** *haber* **Subj**> strongly whenever the conceptualization speakers wish to express matches the entrenched instance (i.e., whenever coding the conceptual import does not call for temporal, aspectual, or modal periphrases). However, Table 6 indicates that, although the personal construction is significantly less frequent in the preterit tense (excluding periphrastic expressions), as in (22), it is already used in about 20% of cases, whereas *hayn,* in (23), only appears in about 5%.

(22) Pero entonces tuve tantos vecinos que fueron como mis hermanos y en adición a mi
        madre *hubieron* tantas madres de otros amigos míos que fueron matriarcales para
        mí (SJ09H12/SJ1051).
        'But, then I had so many neighbors who were, like, my brothers and in addition to
        my mother, *there were*$_{Pers}$ (Preterit) so many mothers of other friends of mine who
        were mother-like for me.'
(23) Pero, este, sí, *hayn* platos como que es, específicos de diciembre (SJ05M12/
        SJ655).
        'But, e, yes, *there are*$_{Pers}$ dishes like sp, specific of December.'

TABLE 6. *Frequencies of the forms* hayn *and* hubieron

	n/n	%
Hayn	21/433	4.8
Hubieron	41/208	19.7

Note: $p < .000$.

This pattern is readily accounted for by the analysis that is being developed here. Although *hubo* rarely appears in isolation from the presentational schema in spontaneous discourse, every speaker will have observed it a limited number of times in four patterns: the impersonal presentational construction, *hubo de* + infinitive 'have to', *hubo que* 'have to', and the preterit perfect construction (e.g., *hubo hablado* 'had spoken'). In contrast, in no matter what type of discourse, *hay* only appears in two patterns: the impersonal presentational construction and the impersonal obligation modal *hay que* + infinitive 'have to + infinitive'. Consequently, speakers possess more evidence that the preterit of *haber* can occur independently of <**AdvP** *haber* **Obj**> than they have for the present tense and, as a result, the preempting effect that is exercised by the latter is stronger than the one that goes out from the former.

Additionally, Hypothesis 2C claimed that complex conceptualizations, that is, with verbal periphrases, as in (24), would favor the personal variant because these would bypass the entrenched instances of *hay* and *hubo*, as the coding of such more elaborate conceptualizations results in the calculation of a novel expression involving aspectual or modal auxiliaries.

(24)   Tú sabes como Estados Unidos, cuando *empezaron a haber* huracanes era como que se llevó todo, pero estamos acostumbrados (SJ09H12/ SJ1088).
        'You know, like the United States, when *there began to be*$_{Pers}$ hurricanes it was like it took away everything, but we are used to them.'

As can be derived from Table 7 this seems to be the case, because the rates of pluralization are more than six times higher in clauses that involve a modal or aspectual auxiliary than in synthetic expressions, which was already observed in earlier investigations (Hernández-Díaz, 2006:1150; Quintanilla-Aguilar, 2009:164–165).

In sum, the true constraint imposed by the verb tense seems to be the generally observed pattern that more specific knowledge (i.e., the entrenched instances of the impersonal variant) preempts the generalization (i.e., the computation of a novel expressions based on the personal variant) for conceptualizations that could be coded equally well using both patterns (Goldberg, 2006:94–98, 2009:102).

TABLE 7. *Pluralized cases of* haber *in the present and preterit tense, by presence/absence of modal or aspectual auxiliaries*

	n/n	*%*
Presentational *haber* expressions in the present and preterit tense without modal or aspectual auxiliaries	62/641	9.7
Presentational *haber* expressions in the present and preterit tense involving modal or aspectual auxiliaries	84/124	67.7

Note: *p* < .000.

Structural priming.   Regarding this factor, Table 3 indicates that production-to-production (i.e., the earlier use of a variant by the speaker) and comprehension-to-production (i.e., the earlier processing of a variant by the speaker) priming effects are among the most important cognitive constraints that condition this variation. In addition, the results indicate that whenever speakers have used a personal presentational *haber* clause, they are more prone to use another one, whether or not they repeat the specific verb form, if the next variable context is situated within a 20-clause range. The same results were obtained for the impersonal presentational *haber* construction. Similarly, when speakers have processed a personal clause, they are more likely to utter an expression based on this pattern and vice versa.

However, as a reviewer points out, it could be the case that some of the speakers are strongly biased toward one of the variants, such that, their singular or pluralized tokens following a singular or pluralized instance could lead Rbrul to attribute too much importance to the structural priming variables. In addition, because we have just seen that *haber* pluralization only occurs rather limitedly in synthetic expressions in the present and preterit tenses, it could be the case that the tokens of *hay* and *hubo* have a similar effect in contexts following an impersonal variant. To control for these possible sources of error, I first excluded the synthetic present and preterit tense tokens and, subsequently, the six speakers whose rates of pluralization in this subcorpus were either below 30% or above 75%. As Table 8 shows, when I run the model on this subset of the data, the priming effect remains intact for production-to-production priming, with almost identical factor weights. However, as this procedure effectively eliminates about 50% of the tokens ($n = 866$ of 1655), for comprehension-to-production priming, unfortunately, the number of observations for the primed conditions drops to the point that no statistically relevant conclusions can be drawn up (personal variant: 77%, $n = 59$ of 77; impersonal variant: 62%; $n = 23$ of 37). Still, the robustness of the production-to-production priming effect advocates strongly in favor of treating *haber* pluralization as a competition between two argument-structure constructions, because, if there were not an overarching schema that could be repeated, we would not expect plurals to prime plurals and singulars to prime singulars regardless of variations in verb form.

TABLE 8. *Production-to-production priming effects for speakers representing robust variation, without synthetic expressions in present and preterit. Numbers, percentages, and weights for the personal presentational construction with* haber

Factor group	n/n	%	Weight (Nouns)	Weight (Speakers)
Production-to-production priming				
Personal variant	207/301	68.8	.63	.64
First occurrence/distance 20+ clauses	50/93	53.8	.44	.43
Impersonal variant	180/395	54.4	.43	.43

Finally, structural priming also seems to offer an explanation for cases such as (25), in which the verb agrees with a direct object pronoun.

(25) Un abuelo está contándoles a sus nietos de su niñez. Uno de ellos, ansioso de saber de esos tiempos pasados, pregunta: "¿Papi, cuando usted era niño, acá ya *habían* carros?" Contesta el abuelo: "¡Claro que los_{Acc} *habían*, no soy tan viejo!" (SJ02M12/SJ231).
'A grandfather is telling his grandsons about his childhood. One of them, desiring to know about those past times, asks: "Granpa, when you were a child, *were therePers* already cars here?" The grandfather answers: "Sure that *there_{Acc} werePers*, I'm not that old!"'

That is to say, a look at Table 9 shows that speakers are more likely to produce the object-verb agreement when they have just used or processed an expression based on <**AdvP** *haber* **Subj**>, which is also evident from the example. As a matter of fact, 71% ($n = 22$ of 31) of the cases in which the verb agrees with a direct object pronoun occur in contexts primed with this variant. Therefore, rather than constituting counterexamples to the main hypothesis, the results seem to suggest that the high level of activation of the personal presentational *haber* construction might cause individual speakers to reanalyze the object pronoun online as a hearer-new subject pronoun. This characterization finds some support in the fact that certain speakers use the unexpected agreement multiple times, whereas others do not use it at all, without there being any clear social pattern.

Social factors

Concerning Hypotheses 4 to 8, Table 3 shows that Rbrul does not consider age, educational achievement, and style to be significant constraints. For age, this result is already implicit in the main hypothesis, which describes the phenomenon as a slowly advancing change from below. By the same token, Hypothesis 7, which states that the rates of pluralization will not decrease as formality rises, already anticipates that the alternations between the two constructions will not display any style shifting. For educational achievement, in contrast, the results do not confirm the expectations formulated in Hypothesis 8.

TABLE 9. *Pluralized cases of* haber *that agree with a direct object pronoun, by production-to-production priming and comprehension-to-production priming*

	n/n	%
Unprimed	0/2	0.0
Primed with the impersonal variant	9/47	16.1
Primed with the personal variant	22/48	45.8
Total	31/106	29.2

Note: $p = .003$.

Still, the findings seem to be supported by earlier studies on Puerto Rican Spanish, which have shown *haber* pluralization to occur frequently in educated speech (e.g., DeMello, 1991; López-Morales, 1992:147; Vaquero, 1978:135–140).

In addition, Table 3 shows that social class is only significant when considered jointly with speaker age, as for older speakers, frequent *haber* pluralization signals middle-class identity, whereas younger speakers associate high frequencies of pluralized *haber* with the lower class. This pattern could suggest that some sort of stigma is developing, which, however, would not be all too salient, because *haber* pluralization is still associated with female behavior and no significant interaction could be found between this factor and speaker age. The absence of a stylistic dimension also seems to support this interpretation of the results.

Finally, let us zoom in on the question whether the variation constitutes a change from below, as is argued by the main hypothesis. In this regard, although the phenomenon does not display any progress in apparent time, the facts that *haber* pluralization does not show any significant style shifting, that older speakers display the social class pattern that is characteristic of ongoing changes from below, and that women favor the personal construction suggest that in San Juan, as in Buenos Aires (Fontanella de Weinberg, 1992b) and Caracas (D'Aquino-Ruiz, 2008; Díaz-Campos, 2003), the variation constitutes a very slowly progressing change from below, which, however, might be evolving into a stable variable.

DISCUSSION AND CONCLUSIONS

Let us now take stock of the results for some concluding remarks. The findings that were obtained for the factor groups that sought evidence for the claim that subjectlike NPs would favor the personal construction indicate that speakers are only mildly influenced by such conditions when they decide between singular and plural presentational *haber*. Nonetheless, I have shown that the tendency to pluralize *haber* more often with animate nouns and nouns that are frequently used as subjects is better stated in terms of typical action-chain position: nouns that refer to entities that are easily conceptualized as action-chain heads favor the personal presentational *haber* construction. In addition, conceptualizations that involve negation have been shown to disfavor the personal variant, as was already observed in Caracas (D'Aquino-Ruiz, 2004:18) and San Salvador (Quintanilla-Aguilar, 2009:175), and I have suggested that this could be due to the fact that, in negative contexts, the NP receives a nonspecific indefinite reading, which, in turn, renders it more objectlike. These findings support the analysis that presentational *haber* occurs in two argument-structure constructions and that, as formulated in Hypothesis 1, the variation between these two patterns is constrained by the generalized preference for unmarked coding (Langacker, 1991:288).

With regard to Hypotheses 2A to 2C, the distribution of the tense forms of *haber* in a 16th-century corpus has proven to be the strongest overall predictor for the

choice between the two variants of the presentational construction with *haber.* Yet, the results that were presented for this factor group are not readily explained by an analysis that takes the unit of change to be the verb and focuses on token frequency, as, for example, in Waltereit and Detges (2008:27). That is to say, without recurring to two construction schemas, entrenched instances and differing strengths of statistical preemption, it is difficult to explain why the personal variant is used less in the preterit vis-à-vis other nonpresent tenses. By the same token, if the effect of the verb tense were somehow due to a differing degree of morphophonologic contrast between the singular and plural forms across tenses (e.g., Bentivoglio & Sedano, 2011:174; Hernández-Díaz, 2006:1151) rather than to statistical preemption, we would not expect to find that expressions involving modal or aspectual auxiliaries, which display the same contrasts, favor the personal construction. In other words, the results seem to confirm Hypotheses 2A to 2C and support the main hypothesis.

Similarly, Hypothesis 3 turns out to be correct, as comprehension-to-production and production-to-production priming are important predictors for speakers' choice between singular and plural presentational *haber*. Most importantly, the results for these independent variables suggest, in line with earlier research at this regard (e.g., Bock et al., 2007), that speakers do not repeat the specific tense forms and their argument structures, but rather that they repeat argument-structure constructions (Goldberg, 2009:107). Therefore, these findings also contribute to the characterization of this phenomenon as a competition between two conventionalized argument-structure schemas, rather than a token-by-token reanalysis of *haber* as an intransitive verb (e.g., Waltereit & Detges, 2008) or of the NP slot of the presentational construction as a subject, as was argued in recent papers (Brown & Rivas, 2012; Rivas & Brown, 2012).

The discussion of social factors (Hypotheses 4 to 8), on the other hand, has been concerned with the answers to two questions: Do Puerto Ricans use the pluralization of *haber* to position themselves within social categories of gender, class, and educational achievement? Does the variation constitute an ongoing linguistic change from below? In this regard, the results show that high rates of pluralization correlate with female gender and, for older speakers, middle-class membership, which suggests an ongoing change from below. Although this is supported by the findings for speech style, the fact that for the group 20 to 35 years of age a linear social class alignment was observed might point to incipient stigmatization and/or stabilization of the change.

In sum, this discussion allows for two final, broader conclusions. First, regarding the phenomenon studied in this paper, the data confirm the main hypothesis to a large extent. In San Juan, Puerto Rico, the pluralization of *haber* corresponds to an ongoing language change from below in which the argument-structure construction <**AdvP** *haber* **Subj**> is replacing the canonical <**AdvP** *haber* **Obj**> pattern, but it might be evolving toward a situation of stable variation. The effect that goes out from the independent variables that were tested in this investigation was shown to come down to the same three general principles of language organization and use that may constrain any type of

linguistic encoding (the preference for unmarked coding, statistical preemption, and structural priming) and speakers' desire to position themselves in the social landscape. This, in turn, shows the potential Cognitive Construction Grammar possesses for the exploration/explanation of ongoing syntactic changes. It is in this framework that syntactic change and the variation that is inherent to it can be modeled for what they are: a competition within the system, constrained by general principles of language change, organization, and use that allows speakers to position themselves within social categories.

NOTES

1. The code at the end of the examples identifies the case in my database. The part before the backslash identifies the informant: SJ: San Juan; 14: informant number 14; H: male informant; 2: age 55+ years; 2: university graduate. The code behind the slash points to the identifier of the case in the database: SJ: San Juan; 1672: case 1672. The subscript *Imp* indicates the impersonal variant of the presentational *haber* construction (i.e., the variant *without* verb agreement). The subscript *Pers* indicates the personal presentational *haber* construction (i.e., the variant *with* verb agreement).

2. This is not a mistake; the Puerto Rican land measure *cuerda* equals 0.97 acres. Twenty-seven *cuerdas* are 26.19 acres.

3. Following CCxG notational conventions, boldface indicates the profiled portions of the event frame (Goldberg, 1995:59). Note that the constructions do not specify the linear ordering of the arguments, which is captured by specific word-order constructions.

4. This claim is not to be confused with other approaches (e.g., Kroch, 2001) that model variation as a competition between two grammars. In CCxG, the grammar provides both alternatives.

5. In principle, nothing prohibits that a noun's typical action-chain position is expressed mentally as a "grammatical relation probability," which, in turn, conditions the variation. Yet, the results seem to argue against this idea, because if the variation were conditioned by such a probability, we would expect speakers to decide on an item-by-item basis whether to encode the NP argument as a subject or not, rather than depend on the generalization (Goldberg, 2006:94–102; Lakoff, 1987:147) that agentlike nouns tend to be used as subjects. In statistical terms, such item-by-item decisions would imply that the influence of the random intercept (in this case, the individual nouns) would rise above that of the fixed factor typical actionchain position of the noun's referent and, hence, that this independent variable would be removed from the model (Johnson, 2009:365). This is not the case.

6. Note, however, that this does not mean that the definiteness/specificity hierarchy is a constraint on the variation, as a change in the definiteness/specificity status of the NP presupposes negative polarity. Moreover, due to the fact that the NP is hearer-new, it can never be pragmatically definite (which would require its identity to be known to both the speaker and the hearer) (Croft, 2003:132; Langacker, 1991:98). Rather, the formally definite NPs that appear in presentational *haber* expressions receive a specific indefinite reading, which is the only interpretation available in affirmative presentational clauses (Prince, 1992:299–300).

REFERENCES

Allen, Kachina, Pereira, Francisco, Botvinick, Matthew, & Goldberg, Adele. (2012). Distinguishing grammatical constructions with fMRI pattern analysis. *Brain & Language* 123(3):174–182.

Ashby, William J., & Bentivoglio, Paola. (1993). Preferred argument structure in French and Spanish. *Language Variation and Change* 5(1):61–76.

Baayen, Rolf H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Bencini, Giulia M., & Golberg, Adele. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language* 43:640–651.

Bentivoglio, Paola, & Sedano, Mercedes. (1989). *Haber*: ¿Un verbo impersonal? Un estudio sobre el español de Caracas. In G. de Granda (ed.), *Estudios sobre el español de América y lingüística afroamericana: Ponencias presentadas en el 45 Congreso internacional de americanistas*. Bogotá: Instituto Caro y Cuervo. 59–81.

———. (2011). Morphosyntactic variation in Spanish-speaking Latin America. In M. Díaz-Campos (ed.), *The handbook of Hispanic sociolinguistics*. Oxford: Blackwell. 123–147.

Blas-Arroyo, José L. (1995–1996). A propósito de un caso de convergencia gramatical por causación múltiple en el área de influencia lingüística catalana. Análisis sociolingüístico. *Cuadernos de investigación filológica* 21–22:175–200.

Bock, Kathryn, Dell, Garry S., Chang, Franklin, & Onishi, Kristine. (2007). Persistent structural priming from language comprehension to language production. *Cognition* 104:437–458.

Brown, Esther L., & Rivas, Javier. (2012). Grammatical relation probability: How usage patterns shape analogy. *Language Variation and Change* 24(3):317–341.

Bybee, Joan. (2003). Mechanisms of change in grammaticization. In B. D. Joseph & R. D. Janda (eds.), *The handbook of historical linguistics*. Oxford: Blackwell. 602–623.

————. (2006). From usage to grammar: The mind's response to repetition. *Language* 82(4):711–733.

Bybee, Joan, & Beckner, Clyde. (2010). Usage-based theory. In B. Heine & H. Narrog (eds.), *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press. 827–856.

Comrie, Bernard. (1989). *Language universals and linguistic typology*. Chicago: The University of Chicago Press.

Croft, William. (2003). *Typology and universals*. Cambridge: Cambridge University Press.

Croft, William, & Cruse, Alan. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

D'Aquino Ruiz, Giovana. (2004). *Haber* impersonal en el habla de Caracas. Análisis sociolingüístico. *Boletín de lingüística* 21:3–26.

————. (2008). El cambio lingüístico de *haber* impersonal. *Núcleo* 25:103–123.

Davies, Mark. (2002). *Corpus del español*. Available at http://www.corpusdelespanol.org. Accessed June 10, 2013.

DeMello, George. (1991). Pluralización del verbo *haber* en el español hablado culto de once ciudades. *Thesaurus* 46(3):445–471.

Díaz-Campos, Manuel. (2003). The pluralization of *haber* in Venezuelan Spanish: A sociolinguistic change in real time. *IU Working Papers in Linguistics* 3(5):1–13.

Du Bois, John. (1987). The discourse basis of ergativity. *Language* 63(4):805–855.

Eckert, Penelope. (2008). Variation and the indexical field. *Journal of Sociolinguistics* 12(4):453–476.

Fauconnier, Gilles. (2007). Mental spaces. In D. Geeraerts & H. Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*. Oxford: Oxford University Press. 351–376.

Fontanella de Weinberg, María B. (1992a). *El español de América*. Madrid: Mapfre.

————. (1992b). Variación sincrónica y diacrónica de las construcciones con *haber* en el español americano. *Boletín de filología de la universidad de Chile* 33:35–46.

Freites-Barros, Francisco. (2004). Pluralización de *haber* impersonal en el Táchira: Actitudes lingüísticas. *Boletín de lingüística* 22:32–51.

Goldberg, Adele. (1995). *Constructions: A Construction Grammar approach to argument structure*. Chicago: Chicago University Press.

————. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

————. (2009). The nature of generalization in language. *Cognitive Linguistics* 20(1):93–127.

————. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics* 22(1):131–153.

Goldberg, Adele, Casenhiser, Devin, & Sethuraman, Nitya. (2004). Learning argument structure generalizations. *Cognitive Linguistics* 15(3):289–316.

Hernández-Díaz, Axel. (2006). Posesión y existencia: La competencia de *haber* y *tener* y *haber* existencial. In C. Company-Company (ed.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal*. Vol. 2. Mexico City: Universidad Nacional Autónoma de México/Fondo de Cultura Económica. 1055–1160.

Johnson, Daniel E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1):359–383.

Keenan, Edward. (1976). Towards a universal definition of subject. In C. N. Li (ed.), *Subject and topic*. New York: Academic Press. 305–333.

Kroch, Anthony. (2001). Syntactic change. In M. Baltin & C. Collins (eds.), *The handbook of contemporary syntactic theory*. Oxford: Blackwell. 699–729.

Labov, William. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

————. (1994). *Principles of linguistic change*. Vol. 1. *Internal factors*. Oxford: Blackwell.

————. (2001). *Principles of linguistic change*. Vol. 2. *Social factors*. Oxford: Blackwell.

Lakoff, George. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Langacker, Ronald W. (1987). *Foundations of cognitive grammar*. Vol. 1. *Theoretical prequisites*. Stanford: Stanford University Press.

———. (1991). *Foundations of cognitive grammar.* Vol. 2. *Descriptive application*. Stanford: Stanford University Press.

Lavandera, Beatriz. (1978). Where does the sociolinguistic variable stop? *Language in Society* 7 (2):171–182.

López-Morales, Humberto. (1992). *El español del Caribe*. Madrid: Mapfre.

Lyons, John. (1967). A note on possessive, existential and locative sentences. *Foundations of Language* 3:390–396.

Meulleman, Machteld, & Roegiest, Eugène. (2012). Los locativos en la valencia de la construcción existencial española. ¿Actante o circunstante? *Zeitschrift für romanische Philologie* 128:57–70.

Moreno-Fernández, Francisco. (2003). *Metodología del proyecto para el estudio sociolingüístico del español de España y América (PRESEEA).* Available at www.linguas.net/portalpreseea. Accessed April 9, 2008.

Prince, Ellen. (1992). The ZPG letter: Subjects, definiteness, and information-status. In W. C. Mann & S. A. Thompson (eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text*. New York: John Benjamins. 295–326.

Quintanilla-Aguilar, José R. (2009). *La (des)pluralización del verbo* haber *existencial en el español salvadoreño: ¿un cambio en progreso?* PhD dissertation. Miami: University of Florida.

Real Academia Española. (2008). *Corpus diacrónico del español (CORDE).* Available at http://corpus. rae.es/cordenet.html. Accessed June 10, 2010.

Real Academia Española & Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española*. Madrid: Espasa-Calpe.

Rivas, Javier, & Brown, Esther L. (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.

Silverstein, Michael. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23:193–229.

Tagliamonte, Sali A. (2006). *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.

Vaquero, María. (1978). ¿Enseñar español, pero qué español? *Boletín de la Academia Puertorriqueña de la Lengua Española* 6:127–146.

Waltereit, Richard, & Detges, Ulrich. (2008). Syntactic change from within and from without syntax: A usage-based analysis. In U. Detges & R. Waltereit (eds.), *The paradox of grammatical change: Perspectives from romance*. Amsterdam: John Benjamins. 13–30.

Ward, Gregory, & Birner, Betty. (1995). Definiteness and the English existential. *Language* 71(4):722–742.

APPENDIX A
Distribution of the third-person singular forms of haber *across constructions in American texts from* CORDE *(1492–1600)*

Constructions	Había		Hubo		Habría		Habrá		Haya		Hubiera		Hay		Ha habido		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Auxiliary for the compound tenses	1806	52.4	38	6.5	6	10.5	41	13.4	202	32.3	110	50	0	0	0	0	2203	25.2
Haber de+infinitive	644	18.7	23	4.0	3	5.3	16	5.2	35	5.6	23	10.5	0	0	0	0	744	8.5
Haber que+infinitive	8	0.2	1	0.2	1	1.8	1	0.3	2	0.3	1	0.5	45	1.3	0	0	59	0.7
Haber with a possessive reading	54	1.6	113	19.4	4	7.0	17	5.5	89	14.2	6	2.7	0	0	0	0	283	3.2
Presentational construction	870	25.2	406	69.8	43	75.4	173	56.4	295	47.2	78	35.5	3440	98.7	6	100	5311	60.8
Años ha+time expression	67	1.9	1	0.2	0	0	59	19.2	2	0.3	2	0.9	0	0	0	0	131	1.5
Total	3449	100	582	100	57	100	307	100	625	100	220	100	3485	100	6	100	8731	100

Note: The following parameters were used for the collection of the instances of *haber*: *CORDE*, 1492–1600, *Lírica, Narrativa, Breve, Relato breve tradicional*, and *otros*. As initial searches within the Caribbean section of the corpus did not yield enough results, the searches were extended to all of Latin America (Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Uruguay, and Venezuela). However, this could not really be considered problematic, because by that time speech communities yet had to be formed and throughout the continent a comparable situation of dialect and language contact existed. I did not take administrative and legal documents into account, because these typically contain a very archaic type of language, which, in the case of the subjunctive present *haya,* results in abundant use of this form as an imperative with its possessive reading, a usage that had already largely decayed by that time in other types of sources.
Source: Based on data from Real Academia Española (2008).

APPENDIX B

Lemmas of the nouns that occur with haber, *by proportion of use as subject in 200-token samples*

Low to mid	%	Low to mid	%	Low to mid	%	Low to mid	%	Low to mid	%	Low to mid	%	High	%	High	%
cuido	0	*salón*	6.5	*evolución*	10.5	*cambio*	13	*tos*	17	*pez*	22	*invitado*	26	*motorista*	43.5
lo	0	*tanto*	6.5	*tienda*	10.5	*carretera*	13	*modalidad*	17.5	*traductor*	22	*persona*	26	*madre*	44
año	0.5	*contacto*	7.5	*caña*	11	*parque*	13	*pelea*	17.5	*corporación*	22.5	*vecino*	26.5	*María*	44.5
vez	0.5	*discusión*	7.5	*daño*	11	*tradición*	13	*gente*	18	*crítica*	22.5	*deambulante*	27	*temblor*	44.5
día	1.5	*madera*	7.5	*fiesta*	11	*vianda*	13	*inundación*	18	*gran sociedad*	22.5	*pájaro*	27.5	*señor*	51.5
mes	2	*maltrato*	7.5	*muerte*	11	*gallina*	13.5	*problema*	18	*maremoto*	22.5	*monja*	28	*papá*	62.5
semana	3	*pantalón*	7.5	*pastel*	11	*amistad*	14	*posibilidad*	18.5	*variación*	23	*estudiante*	29.4		
sillón	3	*parranda*	7.5	*piña*	11	*huevo*	14	*sentimiento*	18.5	*factor*	23.5	*niño*	30		
butaca	3.5	*hueso*	8	*sector*	11	*regla*	14	*comida*	19	*nieto*	23.5	*tormenta*	30		
casa	3.5	*negocio*	8	*animal*	11.5	*árbol*	14.5	*lobo*	19	*paloma*	23.5	*asiático*	30.5		
corral	3.5	*baño*	8.5	*asalto*	11.5	*carro*	15	*monstruo*	19	*grupo de*	24	*ladrón*	30.5		
momento	3.5	*fruta*	8.5	*carnaval*	11.5	*grosella*	15	*recuerdo*	19	*indígena*	24.5	*compañero*	31		
peso	3.5	*reunión*	8.5	*conocimiento*	11.5	*víctima*	15	*ley*	19.5	*terremoto*	25	*ardilla*	31.5		
sitio	4	*apartamento*	9	*pescado*	11.5	*iglesia*	15.5	*león*	20			*huracán*	32		
cuarto	4.5	*clase*	9	*almuerzo*	12	*matrimonio*	15.5	*medio de comunicación*	20			*joven*	32		
restaurante	4.5	*condominio*	9	*beneficio*	12	*pala*	15.5	*sonido*	20			*alumno*	34.5		
porche	5	*juego*	9	*curso*	12	*pino*	15.5	*indicio*	20.5			*vegetariano*	34.5		
lugar	5.5	*jueyes*	9	*llave*	12	*tipo*	15.5	*diferencia*	21			*maestro*	35		
paso de río	5.5	*castigo*	9.5	*vecindario*	12	*truco*	15.5	*hijo*	21			*corpulento*	35.6		
uniforme	5.5	*escuela*	9.5	*bandera*	12.5	*libro*	16	*influencia*	21			*tonto*	35.6		
mango	6	*faceta*	9.5	*cosas*	12.5	*muerto*	16	*edificación*	21.5			*pelotero*	38.5		
actividad	6.5	*plato*	9.5	*experiencia*	12.5	*frase*	16.5	*habitante*	21.5			*padre*	39		
entidad	6.5	*china*	9.83	*piso*	12.5	*palabra*	16.5	*sapo*	21.5			*mujer*	39.5		
nivel	6.5	*ciudad*	10.5	*talento*	12.5	*peligro*	16.5	*situación*	21.5			*hermana*	40		
quiosco	6.5	*estilo*	10.5	*café*	13	*visita*	16.5	*familia*	22			*muchacho*	41.5		

Source: The token samples are from Davies (2002).