Jeroen Claes

Competing constructions: The pluralization of presentational *haber* in Dominican Spanish

Abstract: Drawing on Cognitive Construction Grammar (CCxG), I present an analysis of the pluralization of *haber* in Dominican Spanish (e.g. *Habían fiestas* 'There were parties') as an ongoing language change from below during which the singular argument-structure construction (<**AdvP** *haber* **Obj**>) is being replaced by a pluralized schema (<**AdvP** *haber* **Subj**>). Using a mixed-effects regression analysis, in which the individual speakers and the NPs' head nouns were included as random intercepts, I show that speakers pluralize presentational *haber* in 47% of the cases and that the variation is conditioned by three general cognitive constraints (markedness of coding, structural priming, and statistical pre-emption). Using a conditional inference tree, I show that the former two cognitive constraints work in tandem to promote the pluralized construction for the encoding of conceptualizations that statistical preemption tends to reserve to the singular construction. The results also unveil that the variation is associated with social class membership. These data confirm the main hypothesis, while at the same time corroborating and strengthening the conclusions of an earlier investigation of *haber* pluralization in Puerto Rican Spanish.

Keywords: Cognitive Construction Grammar, variationist sociolinguistics, Dominican Spanish, pluralization of *haber*

DOI 10.1515/cog-2014-0006
Received April 24, 2014; revised July 9, 2014; accepted July 16, 2014.

1 The pluralization of *haber*

In normative Spanish, presentational *haber* is used in subjectless, impersonal constructions. In other words, the NP arguments, *personas subyugadas* 'oppressed people', *pobres* 'poor people', and *ricos* 'rich people' in example (1) do

Jeroen Claes: Department of Linguistics, University of Antwerp. E-mail: jeroenclaes@gmail.com

not function as subjects, but rather direct objects. This is shown by the fact that the NP clitizes as an accusative pronoun in example (2). As a result, the verb represents default third-person singular agreement.

(1) Y de, por eso siempre *va a haber* personas subyugadas y *va a haber* pobres y *va haber* ricos (SD11H22/RD1324).[1]
'And of, therefore, *there will* always be_{SG} oppressed people and *there will* be_{SG} poor and *there will* be_{SG} rich.'

(2) Author: ¿Y también habían comidas que sólo se preparaban en fiestas por ejemplo?
Subj: Sí, claro y todavía las_{Acc} *hay* (SD19M12/RD2547).
Author: 'And were there also dishes that were only prepared on festivities, for example?
Participant: 'Yes, of course, and $there_{Acc}$ are_{SG} still.'

However, in many varieties of Canarian (e.g., Pérez-Martin 2007), Peninsular (e.g., Blas-Arroyo 1999: Chap. 2), and Latin American Spanish (e.g., Bentivoglio and Sedano 2011), including Dominican Spanish (Alba 2004: 28, 323), variable verb agreement with the noun phrase can be found, as in example (3). This phenomenon is known as 'the pluralization of *haber*'.

(3) Y, e, *han habido* ciertos cambios en, en la sociedad (SD16H22/RD2020).
'And, e, *there have* $been_{PL}$ certain changes in, in society.'

In a recent article (Claes 2014), I have argued against the background of Cognitive Construction Grammar (e.g., Goldberg 1995, 2006; CCxG, henceforth) that, in Puerto Rican Spanish, the pluralization of *haber* amounts to a language change 'from below'[2] during which the argument-structure construction <**AdvP** *haber*

1 The codes at the end of the examples indicate the following: SD = Santo Domingo; 11 = informant number 11; H = male speaker; 2 = 55+ years of age; 2 = with university degree. The code behind the backslash points to the occurrence number in my database. The subscripts in the translation identify the case as either a pluralized variant of presentational *haber* (PL) or as a singular variant (SG).

2 Language changes from below are spontaneous linguistic evolutions that emerge in the middle class (Labov 2001: 188) and spread upward throughout the social hierarchy, below the level of awareness (Labov 1972: 179). As this type of language change occurs without speakers realizing it, changes from below have a high probability of going to completion (Labov 1972: 178–180, 2001: 517–518). In situations of change from below, younger speakers (Labov 1994: 43–72), middle–class speakers (Labov 2001: 188), and female speakers (Labov 2001: 292) use the innovative

Subj> is replacing <**AdvP** *haber* **Obj**> under the influence of three general cognitive factors (markedness of coding, statistical preemption, and structural priming). With this paper, I wish to evaluate whether these results are also obtained in Dominican Spanish, as this would greatly fortify the claim that general cognitive constraints, rather than specific contextual features, constrain the alternation between singular and pluralized presentational *haber*. Additionally, I intend to investigate the way these cognitive factors jointly determine speakers' choice for one of the variants of the presentational construction with *haber*.

The paper is organized as follows. In Section two I present the methods that were used in collecting and analyzing the data. Then, in Section 3, I briefly introduce CCxG and I present a succinct CCxG description of the variants of the presentational *haber* construction. Against this background, I present my hypotheses in Section 4. The results follow in Section 5 and the paper concludes with a brief discussion in Section six.

2 Method

2.1 The sample

The analyses are based on a corpus of approximately 28 hours of recording sessions with 24 native speakers of Dominican Spanish who reside in the Greater Santo Domingo Area. As is shown in Table 1, the data are stratified by three social parameters: age (25–35 years vs. 55+ years), academic achievement (without university degree vs. with university degree), and gender (female vs. male).

To obtain more variable contexts and to investigate whether the variation between singular and pluralized presentational *haber* is sensitive to style-shifting,

Table 1: Configuration of the sample

	25–35 years		55+		Total
	Male	Female	Male	Female	
Without university degree	3	3	3	3	12
With university degree	3	3	3	3	12
Total	12		12		24

forms more frequently. Additionally, the rates of use of the innovative variant do not decrease when formality rises (Labov 1972: 239, 2001: Chap. 3), that is, when the amount of attention that is paid to speech increases (Labov, 1972: 99).

the author, who was also the fieldworker, structured the recordings into three sections:

1) **Interview.** Speakers were interviewed for about 30 minutes on a variety of topics related to their day-to-day life. The interview format was loosely based on the conversational modules proposed by Moreno-Fernández (2003), Quintanilla-Aguilar (2009: App. F), and Tagliamonte (2006: App. B). In addition, a set of questions with presentational *haber*, as in example (4), was included in order to investigate the effect of comprehension-to-production priming (see Section 4.3). In these questions, I tried to use a more or less equal number of singular and pluralized presentational *haber* clauses.

(4) ¿Cuántos habitantes *podía haber/podían haber* durante tu niñez?
'How many inhabitants *could there have been*$_{SG}$/*could there have been*$_{PL}$ in your childhood?'

2) **Reading passage.** After the first section, the speakers were instructed to read out loud a two-page children's story (*Juan Sin Miedo*, '*John Without Fear*'), which includes 31 selection contexts (20 trials, 11 fillers), as in example (5).

(5) En una pequeña aldea, *había/habían* un anciano padre y sus dos hijos ...
'In a small village, *there were*$_{SG}$/*there were*$_{PL}$ 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 description that evokes the right pragmatic context for the interpretation of the trial sentence, as can be seen in example (6). Whenever a participant had difficulties completing the reading or questionnaire task, the author read the sentences to her/him and asked her/him which form s/he preferred.

(6) 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
'Inés's car, which she had parked in a dark alley, has just been stolen. 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$_{SG}$ b) There will$_{PL}$

2.2 Post-stratification: social class

In accordance with standard practice in variationist sociolinguistics (e.g., Milroy and Gordon 2003: 42–43), I constructed a composite social class index. To this end, two additional demographic parameters were registered, namely, housing (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), and occupation (0: e.g., unskilled workers; 0.25: e.g., shop–owners, clerks, factory workers; 0.5: e.g., college–educated professionals; 0.75: e.g., university professors; 1: e.g., senior managers/officials; see Moreno-Fernández [2003] for the full scale). Besides these two variables, the index also includes a measure of educational achievement. For this measure, speakers who did not finish high school were separated from those who did, because the former group has greater prestige and easier access to power than the latter.

Since these three variables may not make an equal contribution to individuals' social status (e.g., Milroy and Gordon 2003: 43), I incorporated impact factors in the social class index. In order to establish these, after the recording sessions, I handed the participants a questionnaire with the instruction to rank, on a scale from one to five, educational achievement, housing, and profession by their importance for social status. This provided the relative importance estimates displayed in Table 2.

Afterwards, each speaker's score for the three variables was first multiplied by the relative importance ranking that had been established for that variable in the community, then summed together and, finally, converted into percentages of the maximum possible score. This way, for example, for a university graduate (1 on the education scale) who has a large, comfortable home (1 on the housing scale), and works as a university professor (0.75 on the profession scale), the social status score would be calculated as is shown in (7).

$$(7)\quad \frac{(1\times 4.31)+(1\times 2.56)+(0.75\times 3.63)}{10.5}=91.4\%$$

Table 2: Relative importance to social status of educational achievement, housing, and occupation in Santo Domingo

Factor	Relative importance estimate
Educational achievement	4.31
Housing	2.56
Occupation	3.63

As customary in variationist sociolinguistics, based upon these percentages, speakers were binned together in three social class groups:
– 0–33%: Lower class (9 participants).
– 34–66%: Middle class (6 participants).
– 67–100%: Upper class (9 participants).

2.3 Envelope of variation

Most earlier variationist studies of *haber* pluralization did not document any variation involving the present tense *hay* and its vernacular plural form *hayn* (e.g., Bentivoglio and Sedano 2011). However, as my corpus provides 24 cases of *hayn*, the alternation between *hay* and *hayn* had to be included in the envelope of variation (see, e.g., Labov 1972: 72). In contrast, some previous investigations (e.g., DeMello 1991; Quintanilla-Aguilar 2009) have considered first-person plural *haber* as a case of *haber* pluralization. However, because this form includes the speaker in the *presentatum*, it is not interchangeable with third-person singular *haber,* as is evident from the difference in meaning between example (8) and its constructed variant, example (9). For this reason, the alternation *hay-habemos* was excluded from the envelope of variation.

(8) O sea, *habíamos* nueve personas en un vehículo (SD19M12).
'That is to say, *we were* nine people in one car.'

(9) O sea, *había* nueve personas en un vehículo (constructed example).
'That is to say, *there were*$_{SG}$ nine people in one car.'

2.4 Statistical toolkit

2.4.1 Mixed-effects logistic regression

After transcription, all the cases of *haber* + plural NP were selected from the corpus and coded for linguistic and social factors corresponding to the hypotheses that will be presented in Section 4. Then, I performed a generalized mixed-effects logistic regression analysis with Rbrul (Johnson 2014). In this analysis, I included the individual speakers and the nouns that occur with *haber* as random intercepts. These allow us to model the possibility that some nouns or some speakers might favor or disfavor a particular variant over and above (or under and below) the contextual or social factors they instantiate (Johnson 2009: 365; Tagliamonte 2012: 137). Specifically, when a particular variant occurs more or less often when

a particular fixed factor is present, a generalized mixed-effects model will first evaluate whether this effect is stronger than the variation between the individual speakers or words of the same group. If this is the case, the model will report that the fixed factor has a favorable or an unfavorable effect on the occurrence rate of that variant. In the opposing case, the model will conclude that the fixed factor has no influence whatsoever (Johnson 2009: 365; Tagliamonte 2012: 137). This renders generalized mixed-effects models highly suitable for handling "sociolinguistic data, drawn always from the production of individuals, inevitably from less than ideally distributed datasets, and with innumerable cross–cutting social and linguistic factors" (Tagliamonte 2012: 139).

However, as the random intercepts were collinear, I had to run parallel analyses (e.g., Baayen 2008: 294). In Section 5, I will only report as statistically significant those fixed effects that proved to condition the variation for all speakers and all lexical items. In the tables, the results are provided as log odds. Positive values indicate that a particular level of a factor has a favorable effect on the occurrence rate of pluralized presentational *haber*, negative values indicate the opposite, and 0 is neutral. Summary statistics will also be provided for the models, namely, their deviances, their Akaike's Information Criteria, as well as their Somers' C–index and Dxy. For the former two, lower values are indicative of better model fits. For the latter, higher values indicate a better prediction accuracy.

2.4.2 Conditional inference tree

Although pairwise interactions between predictors can fruitfully be investigated with generalized mixed-effects models, conditional inference trees offer the advantage of visualizing the way all the predictors work together to determine speakers' choice for a particular variant and the complex interactions that may characterize the dataset (Tagliamonte and Baayen 2012: 163). In R (R Core Team 2013), conditional inference trees can be grown with the *ctree()* function of the package *party* (Hothorn et al. 2014). According to Baayen, this type of statistical models

> estimate a regression relationship by means of binary recursive partitioning. The *ctree* algorithm begins with testing the global null hypothesis of independence between any of the predictors and the response variable. The algorithm terminates if this hypothesis cannot be rejected. Otherwise, that predictor is selected that has the strongest association to the response, as measured by a p-value corresponding to a test for the partial null hypothesis of a single input variable and the response. A binary split in the selected input variable is carried out. These steps are recursively repeated until no further splits are supported. (Baayen 2014: 364)

In the conditional inference tree that will follow in Section 5.4, the ovals represent the factors. The higher a node is located in the tree, the stronger it is associated with speakers' choice between the presentational *haber* constructions. The branches that go down from the nodes represent the binary split the algorithm has established in the data. At the bottom, the thermometer graphs represent the proportion of pluralized presentational *haber* in light gray. For this model, Somers' C–index and Dxy will be provided.

3 Theoretical approach

3.1 Cognitive Construction Grammar

CCxG (e.g., Goldberg 1995, 2006) is a usage–based linguistic theory that takes an encyclopedic approach to semantics. Most importantly, CCxG claims that form-meaning pairs, called 'constructions', represent every aspect of language (Goldberg 1995: 7, 2006: 45). That is, both generalizations, such as, for example, transitivity, and highly idiosyncratic patterns such as words or idioms are considered to be represented mentally as constructions (Goldberg 2009: 94; Langacker 1987: 28). In other words, in CCxG, the 'grammar' is reduced to an interconnected network of constructions, which provide speakers with symbolic resources to encode conceptualizations. As such, in example (10), there are no less than six constructions: *John, gave, Beth, a, napkin* and the ditransitive construction <**Subj** V **Obj1 Obj2**>.

(10) John gave Beth a napkin (constructed example).

However, contrary to low-level constructions such as *John, gave, Beth, a,* and *napkin*, the double object construction possesses a more schematic meaning. According to Goldberg (1995: Chap. 6), this meaning comes down to a conceptualization of the first argument causing the second argument to receive the third argument, or, schematically, CAUSE-RECEIVE. Based on this abstract sense, the ditransitive construction determines that there will be three profiled argument roles: agent, patient, and receiver. In addition, the construction schema specifies how these roles will be mapped onto syntactic functions and how information will be distributed over the arguments. In other words, in this framework, argument-structure constructions, rather than verbs, are taken to be the key determinants of clause-structure.

3.2 A Cognitive Construction Grammar description of presentational *haber*

Within this theoretical setting, I would like to advance the hypothesis that the pluralization of *haber* arises from a competition within the grammar between two variants of the presentational construction with this verb, which allows speakers to position themselves against the background of social categories. As argued in Claes (2014), both variants encode the POINTING-OUT idealized cognitive model proposed by Lakoff:

> 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. (Lakoff 1987: 490)

The semantic similarities between the two presentational constructions with *haber* entail that the variants are also very similar when it comes to their formal characteristics. That is, both the argument-structure constructions possess a single profiled nominal argument, which is assigned the 'zero' argument role (see Langacker 1991: 289) and has to provide new information with reference to the beliefs and/or knowledge of the listener (Ward and Birner 1995). Similarly, both types of presentational *haber* clauses display a profiled adverbial phrase, which functions to set up and describe the mental space (see Fauconnier 2007) wherein the presentational construction situates the NP argument (Lakoff 1987: 542). With this in mind, let us now consider the way *haber* pluralization, including the cognitive and social factors that shape its usage, can be modeled in CCxG.

4 Hypotheses and coding

In this study, the following main hypothesis will be pursued:

In Santo Domingo, the Dominican Republic, the pluralization of presentational *haber* corresponds to a slowly advancing language change from below: the pluralized presentational construction schema with *haber* (<**AdvP** *haber* **Subj**>) is replacing the singular presentational construction with this verb (<**AdvP** *haber* **Obj**>). The variants only differ with regard to the syntactic function of the NP

(singular variant: object; pluralized variant: subject) and the social groups associated with their relative frequencies.[3]

Of course, this is a very abstract hypothesis, which on its own, does not allow for any predictions. However, through reference to three general 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.

4.1 Markedness of coding

Regarding the first cognitive factor, when introducing the notion of 'markedness of coding' Langacker (1991: 298) observes that, ideally, "a notion approximating an archetypical conception is coded linguistically by a category taking that conception as its prototype". 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 pluralized variant more often with NP arguments that are more similar to prototypical subjects and the singular variant with NP arguments that are more similar to prototypical objects.

This hypothesis raises the question as to which features characterize prototypical subjects. In this regard, the typological literature suggests agenthood as a relevant characteristic (e.g., Keenan 1976: 321; Langacker 1991: 294). However, as we have seen in the previous section, presentational *haber* does not assign an agent, but rather a 'zero' semantic role. Still, some entities (say, *lumberjack*) are intrinsically more likely than others (say, *tree*) to take on an agentive role in events. Therefore, with presentational *haber*, which does not explicitly construe the nominal as a patient nor an agent, entities such as *lumberjack* may be perceived as more potential agents, and, thus, as more prototypical subjects (Langacker 1991: 294) than entities like *tree*.

Since the semantic roles 'agent' and 'patient' are defined in relation to what Langacker (1991: 283–285) calls the 'canonical event model' or the 'action-chain

3 In CCxG, boldface indicates profiled portions of event frames. Observe as well that the constructions do not specify the linear ordering of the arguments.

model',[4] I coded the data for the typical action-chain position of the entity designated by the noun. In order to do so, I relied on the answers to the question in (11).

(11) *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 absence/presence of negation. As I will argue in Section 5.1.1, this factor can be connected to subject/objecthood in the sense that, with presentational *haber*, the presence of negation causes the NP argument to receive a non-specific indefinite reading. This interpretation, in turn, is prototypical of direct objects (Croft 2003: 132).

4.2 Statistical preemption

Furthermore, in CCxG, forms that occur in multiple constructions are taken to possess stronger independent cognitive representations than those that only occur in one pattern (Bybee and Beckner 2010: 842). As a matter of fact, 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 more abstract construction (Goldberg 1995: 79; Langacker 1987: 59–60, 1991: 48). As expressions based on this partially prefabricated unit require less constructive effort than expressions involving the joint use 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'.

4 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 settings, such that the event model minimally includes three elements: action–chain head/agent, action–chain tail/patient and setting.

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 pluralized 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 following generations, repetition usually ensures that this distribution and the resulting statistical preemption remain intact (Bybee 2006: 715). This leads to hypotheses 2a–c.

Hypothesis 2, Statistical preemption

Hypothesis 2a: If the third–person singular form of a particular tense of *haber* was frequently used outside of the singular construction schema before the actuation of the change, this verb tense will favor the pluralized variant.

Hypothesis 2b: The other verb tenses will disfavor the pluralized construction, provided the conceptualization can be expressed with an entrenched instance of the singular construction.

Since 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 singular construction, the following prediction follows quite naturally:

Hypothesis 2c: When the need to encode an aspectually or modally more complex conceptualization forces speakers to construct a new expression involving aspectual or modal auxiliary constructions rather than retrieving a partially prefabricated one from long-term memory, the tenses of *haber* that were mainly used in the singular construction schema before the actuation of the change will favor the pluralized variant.

However, hypotheses 2a–c raise two questions. First, when did the variation that affects presentational *haber* start at a community-wide scale and, second, which forms of the verb enjoyed a relatively high token frequency in a variety of constructions before this happened? The answer to the first question can only be speculative, as it is virtually impossible to know for certain when, where and how the variation that affects presentational *haber* started exactly. What we do know is that the alternations between singular and pluralized *haber* are already present in documents written in Buenos Aires during the eighteenth century (Fontanella de Weinberg 1992b) and that there is usually a considerable lag between the appearance of an alternation in the spoken language and its trickling down into written documents. Hence, it seems reasonable to assume that the variation probably emerged somewhere in the seventeenth century, which blends in nicely with research in historical linguistics showing 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: Chap. 1). Therefore, in order 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 sixteenth-century Latin American subset of the Spanish Royal Academy's *Corpus Diacrónico del Español* (Real Academia Española 2008–).

The examination of this database indicates that before the pluralization of presentational *haber* appears on a large scale in American Spanish, the present tense (*hay*) and the preterit (*hubo*) occurred primarily in the presentational pattern, which suggests that their most prominent cognitive representations were <**AdvP** *hay* **Obj**> and <**AdvP** *hubo* **Obj**>. The other tense forms, in contrast, are either spread over more different constructions or are restricted to a very low frequency in the corpus (N < 100). This indicates that their independent forms probably also constituted their strongest cognitive representations. In other words, the distributional pattern that emerges from Appendix A suggests two relevant tense groups: synthetic[5] expressions in present and preterit tense vs. all others.

4.3 Structural priming

Additionally, investigations in the variationist and psycholinguistic traditions support that speakers tend to recycle abstract patterns they have (heard) used in earlier stretches of discourse (e.g., Goldberg 2006: 120–125; Labov 1994: Chap. 20; Pickering and Ferreira 2008). In the psycholinguistic literature, this is called 'structural priming' (Pickering and Ferreira 2008: 427–428). Therefore, in describing the pluralization of *haber* as an argument–structure construction alternation, the main hypothesis entails hypothesis 3.

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. That is, there will be structural priming effects at the argument–structure level regardless of variations in tense, aspect, or mood.

Psycholinguistic experiments have shown priming effects to last for at least ten intervening clauses (Bock and Griffin 2000: 186; Bock et al. 2007: 452; Pickering and Ferreira 2008: 447) and to be modality-independent (Bock et al. 2007: 454; Pickering and Ferreira 2008: 440–441). Therefore, the data were coded for the type of last token that was provided by the author (comprehension–to–

5 I use the term 'synthetic expressions' to indicate expressions where *haber* appears without aspectual or modal auxiliary constructions.

production priming) and the speaker (production–to–production priming) and the number of conjugated verbs that occur between these tokens and the case at hand. While coding, the occurrences were binned together in 5–clause lag groups up until reaching a 20–clause lag and the occurrences in which speakers repeated the verb form and the presentational *haber* construction were separated from those in which they only repeated the construction. This resulted in a total of seventeen levels for both factors. However, as the initial results displayed a similar priming effect up until a 20–clause lag, independently of whether speakers would repeat the same verb form or not, the factors were collapsed into the broader categories listed in (12).

(12) – First occurrence/distance 20+ clauses
– Primed with the pluralized presentational *haber* construction
– Primed with the singular presentational *haber* construction

4.4 Principles of Linguistic Change (Labov 2001)

As the main hypothesis describes *haber* pluralization as a change from below, it also predicts that the alternations between singular and pluralized *haber* will covary with social factors along the lines described by Labov's (2001) Principles of Linguistic Change. A first clue of an ongoing language change is that younger speakers make more use of the innovative variant than older speakers (Labov 1994: 43–72). This leads to hypothesis 4.

Hypothesis 4, Apparent Time: The youngest speakers will favor the pluralized variant, whereas older speakers will make more use of the singular variant.

Nevertheless, earlier research suggests that *haber* pluralization progresses too slowly to be observed in apparent time (Fontanella de Weinberg 1992b), if it progresses at all (Quintanilla-Aguilar 2009). Therefore, more evidence will be needed to test the change-in-progress hypothesis contained in the main claim. In this regard, Labov's (2001: Chap. 8) Gender Principle establishes that "[i]n linguistic change from below, women use higher frequencies of innovative forms than men do" (Labov 2001: 292). Therefore, if the pluralization of *haber* amounts to such a language change, I expect to find the pattern described by hypothesis 5.

Hypothesis 5, Gender Principle: In comparison to men of the same social characteristics, women will use the pluralized variant more often.

Yet, since gender-differentiated behavior is also found in changes from above (Labov 2001: 274) and because differences between generational groups can also

be due to age grading,[6] 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 linguistic change from below seems to be the social class distribution described by the Curvilinear Principle: "[l]inguistic change from below originates in a central social group, located in the interior of the socioeconomic hierarchy" (Labov 2001: 188). This entails hypothesis 6.

Hypothesis 6, Curvilinear Principle: The middle class will show higher frequencies of use of the pluralized construction than the groups of lower and higher social status.

Research in sociolinguistics has also shown that, in changes from below, the incoming variants display no style-shifting at all or increase in frequency when more attention is explicitly turned to speech (i.e., when formality rises; Labov 2001: Chap. 3). This leads to hypothesis 7.

Hypothesis 7, Formality: When formality increases, the rates of pluralization will not decrease.

Finally, Labov (1972: 138) observes that university education triggers an important change in speakers' sensitivity to linguistic variation and its association to social types. As a result, speakers who have enjoyed an extensive formal education, regardless of their social class, tend to conform less to community-based standards of conduct but all the more to supralocal prestige norms. For *haber* pluralization, Freites-Barros (2003: 380) has shown that this bias towards normative usage causes university-educated speakers to judge pluralized *haber* as incorrect more frequently. Therefore, I expect to find the pattern described by hypothesis 8.

Hypothesis 8, Educational achievement: Higher educational achievement will favor the singular variant whereas less formal education will promote the pluralized variant.

5 Results

My dataset contains 1,861 instances of *haber* followed by a plural NP, of which 46.7% correspond to pluralized occurrences. Although these numbers show that

6 Labov (1994) describes age grading as "a regular change of linguistic behavior with age that repeats in each generation" (Labov 1994: 45).

the pluralization of presentational *haber* is a wide-spread variation in Dominican Spanish, they do not match the high frequencies that have been documented in earlier studies of Latin American Spanish (54–82%; Bentivoglio and Sedano 2011: 173). Yet, these differences appear to be due to the fact that I did not exclude the

Table 3: Parallel by-speaker and by-noun generalized mixed-effects logistic regression models of the pluralization of *haber*: Numbers, percentages, and Log odds for the pluralized presentational *haber* construction

Fixed effects	**N**	**%**	**Log odds (by noun)**	**Log odds (by speaker)**
(intercept)			–0.20	–0.25
Verb tense				
All others	720	65.3	1.39	1.17
Expressions in the present and preterit tense, without aspectual or modal auxiliaries	140	18.9	–1.39	–1.17
Production–to–production priming				
Pluralized variant	484	68.1	0.85	0.77
No earlier use/last use 20+ clauses removed	123	36.5	–0.15	–0.24
Singular variant	253	31.9	–0.70	–0.53
Typical action–chain position				
Typical heads	439	53.9	0.43	0.44
Typical tails and settings	421	41.0	–0.43	–0.44
Comprehension–to–production priming				
Pluralized variant	151	57.2	0.48	0.50
Singular variant	63	34.1	–0.16	–0.14
No earlier use/last use 20+ clauses removed	646	46.4	–0.33	–0.35
Social class				
Middle class	248	53.8	0.37	0.41
Lower class	301	45.5	–0.16	–0.18
Upper class	311	43.3	–0.21	–0.23
Random intercepts			**Nouns**	**Speakers**
Variance			0.59	0.11
Standard Deviation			0.77	0.33
Model summary			**By–noun model**	**By–speaker model**
Deviance			1837.24	1889.31
Akaike's Information Criterion			1857.24	1909.31
Somers' C			0.86	0.83
Somers' Dxy			0.72	0.66

present tense *hay-hayn*. Without these forms, the frequency of pluralized *haber* rises as high as 61.2% (N = 835/1320). In what follows, I will discuss the results by order of the hypotheses they are meant to test, beginning with those relative to the typical action-chain position of the noun's referent.

5.1 Markedness of coding

Earlier studies of *haber* pluralization have (nearly) consistently found that the pluralized forms are more frequent with human- or animate-reference NPs (Bentivoglio and Sedano 1989; D'Aquino-Ruiz 2004; Quintanilla-Aguilar 2009). In addition, Brown and Rivas (2012) have shown that speakers pluralize *haber* more often when the noun is frequently used as subject in Spanish generally. Rivas and Brown (2012), on the other hand, have argued that singular *haber* is more common with 'stage–level' NPs (nouns that refer to events or entities that can be imagined as having a beginning and an end; e.g., *barrios* 'rough neighboorhoods' in example 13), whereas pluralized *haber* tends to co-occur with 'individual-level' NPs (nouns that refer to entities that have no perceivable beginning or end; e.g., *características* 'characteristics' in example 14).

(13) *Hay* barrios que tú no puedes entrar con una cartera porque sabes que tú entras con una cartera y vas a salir sin nada, segurito que sí (SD12M11/RD1482).
'*There are*$_{SG}$ rough neighborhoods that you cannot enter with a purse because you know that you enter with a purse and you will go out without anything, sure as hell.'

(14) Pero dentro de la, dentro de la gran mayoría, e, *hay* características que hay que reconocer que valen la pena que tiene este determinado sector (SD19M12/RD2505).
'But on a, on a whole, eh, *there are*$_{SG}$ characteristics that one has to recognize that they are worth it, which this particular sector has.'

However, in my study of *haber* pluralization in Puerto Rican Spanish (Claes 2014), I have shown that all of these results can be subsumed under the term 'typical action-chain position'. Indeed, Table 3 shows that in Dominican Spanish, as in Puerto Rican Spanish, nouns that can easily be imagined as heads in the action chain model (see examples 15–16) favor pluralization, as is evident from the positive log odds values.

(15) E, hubo un tiempo que uno podía decir que: "Esto es un barrio muy tranquilo", que uno salía a cualquier hora de la noche, no *habían* atracadores (SD18M22/RD2375).
'Eh, there was a time that one could say: "This is a very quite neighborhood", that one went out at whatever time of night, *there weren't*$_{PL}$ assaulters."'

(16) Pero después, sí *hubieron* muchos ciclones, como diez ciclones *hubieron* más (SD01H21/RD81–RD82).
'But afterwards, yes, *there were*$_{PL}$ a lot of hurricanes, about ten more hurricanes *there were*$_{PL}$'

In contrast, Table 3 shows that nouns that refer to more typical endpoints of the action chain (i.e., typical patients, exemplified in 17–18) or to typical settings of events (e.g., *aeropuertos* 'airports' in example 19) are more common with singular *haber*, as is shown by the negative log odds values.

(17) Nosotros salíamos a pasear por los parques, a los parques así como esto, antes *habían* columpios, *habían* cosas de cre, recreación, sí (SD01H21/RD62).
'We went for walks in parks, to parks like this here, back then *there were*$_{PL}$ swings, *there were*$_{PL}$, re, recreational things.'

(18) Sí, *hubo* muertos, a esos, a esos lados de, de las, las cañales y esa vaina, de ríos por los lados de ríos siempre *hay* ahogados (SD09H11/RD1067).
'Yes, *there were*$_{SG}$ dead people, at those, at those banks of, of the, the channels and shit, of rivers, on the river banks *there are*$_{SG}$ always drowned people.'

(19) Ahora *hay* aeropuertos nuevos, modernos (SD18M22/RD2391).
'Nowadays *there are*$_{SG}$ new, modern airports.'

Finally, Claes (2014) reports that Puerto Ricans tend to use the singular variant more often with presentational *haber* expressions involving negation. However, Rbrul detects no such effect in Dominican Spanish.

5.2 Statistical preemption

For the verb tense, Table 3 shows that, in Santo Domingo, the distributional pattern presented in Section 4.2 (i.e., synthetic expressions in the present and pret-

erit tense vs. all others) is the strongest overall constraint on the variation, as is shown by the fact that for this factor the highest/lowest log odds are obtained. Particularly, the tenses for which an entrenched instance of the singular variant was posited (i.e., present and preterit) disfavor the use of the pluralized presentational *haber* construction strongly, provided the conceptualization can be coded using <**AdvP** *hubo* **Obj**> or <**AdvP** *hay* **Obj**> (i.e., whenever coding the conceptual import does not call for aspectual or modal auxiliaries). In contrast, all other types of expressions are more frequently formed with pluralized presentational *haber*, as is shown by the positive log odds. This supports hypotheses 2a–2b.

Additionally, Hypothesis 2c claims that expressions involving aspectual or modal auxiliary constructions, as in example (20), would favor the pluralized presentational *haber* construction, because these bypass the entrenched instances of *hay* and *hubo*.

(20) Yo creo que *podrían haber* como algunos quinientos chicos (SD05H11/RD567).
'I think *that there could have been*$_{PL}$ like some five hundred boys.'

As can be derived from Table 4, this is the case, because in expressions with aspectual or modal auxiliaries, pluralized presentational *haber* is used as frequently with the present and the preterit tense as with other tenses, as was already observed in earlier investigations (e.g., Quintanilla-Aguilar 2009: 164–165).

Additionally, although present- and preterit-tense presentational *haber* expressions not involving aspectual or modal auxiliaries were consistently binned together in the tables, this is not to say that both types display similar rates of pluralization. Rather, Table 5 indicates that, in synthetic expressions, *haber* pluralization occurs significantly more often with the preterit (see example 21), than with the present tense, as in example (22).

Table 4: Present- and preterit-tense tokens of presentational *haber*, by absence/presence of aspectual or modal auxiliary constructions: Numbers and percentages for the pluralized presentational *haber* construction (Pearson's Chi–square: 127.35; df: 1; p < 0.0001).

	N	%
Presentational *haber* expressions in the present and preterit tense without auxiliary constructions	140	18.9
Presentational *haber* expressions in the present and preterit tense involving auxiliary constructions	92	64.3

Table 5: Present- and preterit-tense tokens of presentational *haber* without aspectual or modal auxiliary constructions: Numbers and percentages for the pluralized presentational *haber* construction (Pearson's Chi–square: 236.45; df: 1; p < 0.0001).

	N	%
Present tense	24	4.6
Preterit tense	116	53.2

(21) Sí, claro, bastantes, *hubieron* unos cuantos muertos, sí (SD22M12/RD2928).
Yes, of course, a lot, *there were*$_{PL}$ a number of casualties, yes.

(22) ¿E, qué otras cosas puedo yo decir que *hayn* ahora que antes no habían? (SD12M11/RD1420).
'Er, what other things can I say that *there are*$_{PL}$ now that there weren't before?'

This pattern is readily accounted for by the analysis that is being presented here: although the preterit *hubo* rarely occurs outside of the singular presentational *haber* schema in spontaneous discourse, every native speaker of Spanish will have observed it a limited number of times in four patterns in other genres, namely, the singular presentational *haber* construction, <*hubo de* infinitive> '<*have to* infinitive>', <*hubo que* infinitive> '<*have to* infinitive>', 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 singular presentational *haber* construction and the singular deontic modal <*hay que* infinitive> '<*have to* infinitive>'. Consequently, speakers possess more evidence that the preterit of *haber* can occur in isolation of the singular presentational *haber* construction than they have for the present tense. As a result, the preempting effect that goes out from the latter is stronger than the one that goes out from the former.

5.3 Structural priming

In Table 3, the log odds values for production-to-production priming and comprehension-to-production priming indicate that speakers are more likely to use pluralized presentational *haber* when they have used or processed a pluralized presentational *haber* expression in the past twenty clauses. Similarly, they are less likely to use pluralized presentational *haber* when they have used or processed a singular presentational *haber* instance in the past twenty clauses.

This finding argues in favor of hypothesis 3 and the main hypothesis, because we would not expect plurals to prime plurals regardless of variations in verb form if speakers did not repeat overarching argument-structure constructions.

Other than being good predictors for speakers' choice between the variants of the presentational construction with *haber*, priming effects also seem to account for the cases in which the verb agrees with a direct object pronoun, as is exemplified in (23).

(23) Author: ¿Este, tú piensas que *pueden haber* diferencias entre las regiones del país en cuanto a comida?
Participant: Bueno, *tienen que haberlas*, porque, por ejemplo, en el Sur se comen más granos (SD19M21/RD2551).
Author: 'Er, do you think that *there can be*$_{PL}$ differences between the regions of the country regarding food?'
Participant: 'Well, *there*$_{Acc}$ *have to be*$_{PL}$, because, for example, in the South they eat more grains.'

That is, a look at Table 6 informs that speakers are significantly more likely to produce the object-verb agreement when they have just used or processed an expression based on <**AdvP** *haber* **Subj**>. As a matter of fact, 80% (N = 24) of the examples in which the verb agrees with a direct object pronoun occur in contexts primed with the pluralized variant. Hence, rather than constituting strong evidence arguing against the main hypothesis, these results may suggest that priming effects cause individual speakers to reanalyze the direct object pronoun (a syntactically motivated class) as a hearer–new subject pronoun (i.e., as a pragmatically motivated class). Still, this appears to be an online phenomenon, because, if the reanalysis of the pronoun were a change in progress, one would expect to find clear social patterning. This is not the case.

Table 6: Presentational *haber* tokens that co-occur with object pronouns, by production–to–production priming and comprehension-to-production priming: Numbers and percentages for the pluralized presentational *haber* construction (a two–tailed Fisher's exact test shows $p < 0.0001$).

	N	%
No priming	1	50.0
Last variant: pluralized construction	24	44.4
Last variant: singular construction	5	10.2
Total	30	28.6

5.4 Interaction between cognitive factors

Up until now, the discussion has been concerned with the way the individual cognitive constraints shape *haber* pluralization when they are considered jointly with the other cognitive factors, the social factors, and the random variation due to individual speakers and nouns. What has not been considered is the way these factors work in tandem to promote one of the variants or, conversely, interact to cancel each other's effect. Therefore, following Tagliamonte and Baayen (2012), in this section, I will use a conditional inference tree to visualize how the three cognitive constraints considered in this study jointly determine speakers' use of singular and pluralized *haber*. In this model, I only included the linguistic factors that turned out to be significant in the mixed–effects models.

Like the mixed–effects models, Figure 1 suggests that the verb tense is most strongly associated with speakers' choice between singular and pluralized presentational *haber*. For both tense groups (i.e., for both branches that go down from the Tense node), production-to-production priming is the second-most important factor. Additionally, the right–hand side of Figure 1 displays an interaction between the verb tense, production-to-production priming, and action–chain position. Specifically, with non-present, non-preterit expressions or expressions involving auxiliary constructions, the noun's typical action-chain position is only significant in contexts primed by the speaker with pluralized presentational *haber*. Still, for both tense groups, the inference tree shows that the rates of *haber* pluralization are highest when the speaker has used a pluralized presentational *haber* construction in earlier discourse and the noun refers to a typical action-chain head.

This suggests an antagonistic relationship in (this) language change between statistical preemption and the other two cognitive factors. That is, whereas the first encourages speakers to stick to what they have observed, the other two incite speakers to extend the pluralized presentational *haber* construction to more (and new) conceptual territories. More specifically, every time the preference for unmarked coding and structural priming tip the balance in favor of the pluralized presentational *haber* construction for the encoding of a present- or preterit-tense POINTING-OUT conceptualization without aspectual or modal nuances, the computation of an expression based on this construction weakens the strength of the representations of the entrenched singular instances. This, in turn, debilitates the preemptive effect inherent to strong cognitive representations, which, eventually, may result in the less constrained use of <**AdvP** *hayn* **Subj**> and <**AdvP** *hubieron* **Subj**>.

Finally, the right-hand side of Figure 1 also suggests an interaction between the verb tense, production-to-production priming, and comprehension-to-

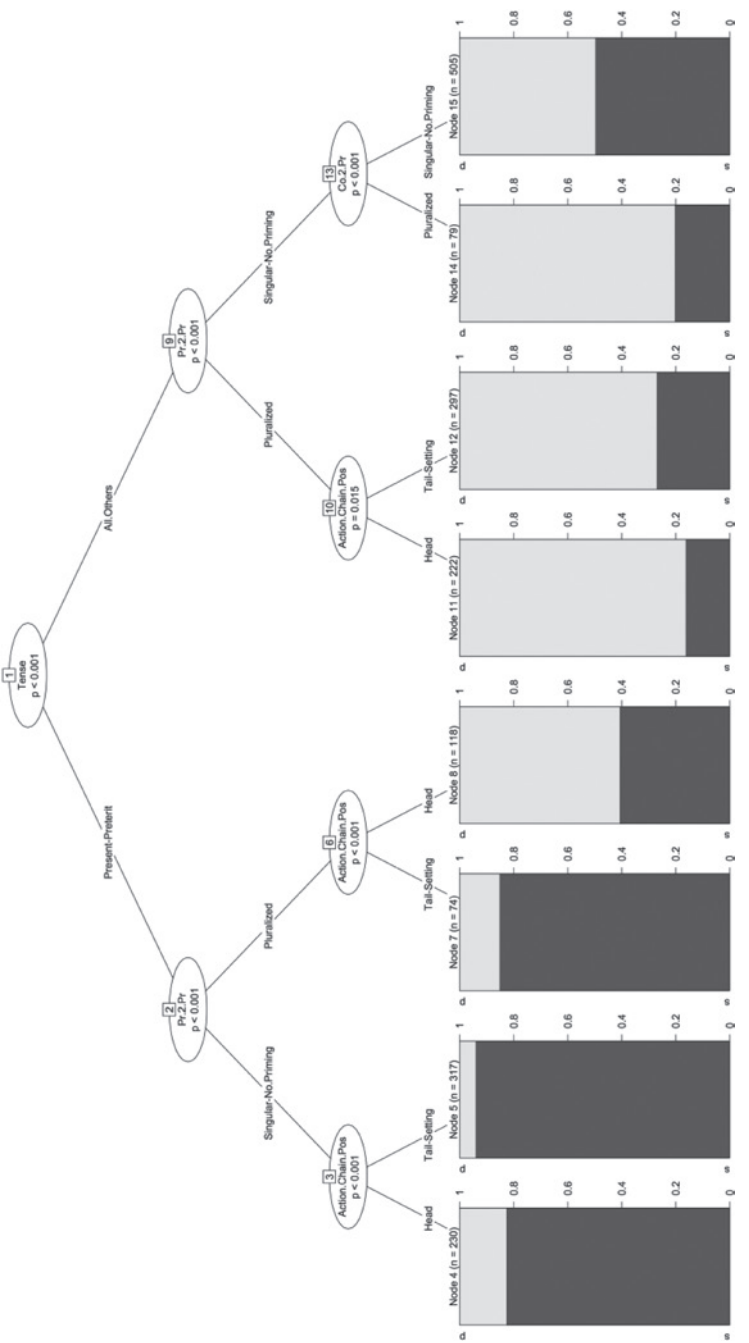

Fig. 1: Conditional inference tree showing the interaction between markedness of coding, statistical preemption, and structural priming
Note: *S* indicates singular presentational *haber*; *P* indicates pluralized presentational *haber*. Co.2.Pr means 'comprehension–to–production' priming; Pr.2.Pr means 'production–to–production' priming. Somers' C = 0.81; Somers' Dxy = 0.62.

production priming, because the latter factor is only significant for non-present, non-preterit expressions or expressions involving auxiliary constructions that occur in unprimed contexts or contexts primed by the speaker with the singular variant. Particularly, in these cases, comprehension-to-production priming is able to cancel production-to-production priming. This is also evident from contexts such as the one provided in example (24), where the speaker appears to be insensitive to the priming effect that one would expect to go out from her earlier use of *hay*.

(24) Author: ¿Y *han habido*, o sea, cuando usted, t, o sea, me podría nombrar cinco cosas que existen hoy y que no *habían* cuando usted era niña? ¿Acá en la ba, en el barrio?
Participant: ¿Cómo así? ¿Cómo así?
Author: Este, como por ejemplo que en, a, edificios que, que poní, que, que pusieron, remodelaciones, e, restaurantes?
Participant: Aja okay, que no *habían* cuando yo era niña. Okay. *Hay* muchas cosas que no *habían* (SD10M21/RD1151–RD1153).
Author: And *have there been*$_{PL}$, that is, when you, that is, could you name me five things that exist today and that *there weren't*$_{PL}$ when you were a child, here in the nei, in the neighborhood?
Participant: Like what? Like what?
Author: Er, like, for example, that in, a, buildings that that they pu, that, that they have put, remodeling, er, restaurants?
Participant: Aha, okay, that *there weren't*$_{PL}$ when I was a child. Okay, *there are*$_{SG}$ a lot of things that *there weren't*$_{PL}$.

5.5 Principles of linguistic change

Turning now to the social factors that are considered in this study, Table 3 shows that Rbrul attributes no significance to academic achievement, age, gender, and speech style/formality. For age, this result was already anticipated in the discussion of the hypotheses, because earlier investigations (Claes 2014; D'Aquino-Ruiz 2008; Fontanella de Weinberg 1992b: 44) have shown the pluralization of *haber* to be a slowly advancing process, which may be invisible in apparent time. By the same token, hypothesis 7 already anticipates that the alternations between the two constructions may not display any style shifting. As the effect of prolonged formal education is usually explained in terms of a better control of formal styles and a greater sensitivity to the formality of the usage event (Labov 1972: 138), the

absence of a stylistic value for the pluralization of *haber* might also explain why education does not condition the phenomenon.

In contrast, the results for gender do not conform to the expectations set forward in hypothesis 5. Yet, this does not necessarily mean that the pluralization of presentational *haber* constitutes a stable variable or a change from above, which would imply that women disfavor the pluralized presentational *haber* construction (Labov 2001: Chap. 3). Additionally, Table 3 shows that the middle class favors the pluralized presentational *haber* construction, as is expected by hypothesis 6.

Therefore, the findings for gender, social class, and style seem to support that, in the Dominican variety, the pluralization of *haber* constitutes a slowly progressing, advanced language change from below, for which Labov (2001: 308–309) observes that gender differences tend to become smaller or may disappear altogether. In this sense, the results discussed in this section seem to corroborate those of earlier investigations of Latin American Spanish (Claes 2014; D'Aquino-Ruiz 2008; Fontanella de Weinberg 1992b), while at the same time supporting the main hypothesis.

6 Conclusion

Let us now return to the hypotheses for some concluding remarks. The data that were reported for the factor typical action-chain position of the referent of the NP indicate that speakers are more likely to use pluralized *haber* with NPs that refer to more typical agents. In contrast, negation does not constitute a significant restriction on the variation. Yet, as agenthood is among the most salient characteristics of subjects, the results still support the view that the preference for unmarked coding conditions the use of the singular and the pluralized presentational *haber* constructions.

The distribution of the tense-forms of *haber* in a sixteenth-century corpus proved to be the strongest overall predictor for the choice between the two variants of the presentational construction with *haber*. The verb tenses for which an entrenched instance of the singular presentational *haber* construction was posited disfavor the pluralized presentational *haber* construction, provided the conceptualization can be encoded with <**AdvP** *hay* **Obj**> or <**AdvP** *hubo* **Obj**>. In contrast, whenever the encoding of the conceptual import calls for aspectual or modal auxiliaries, pluralized *haber* occurs as frequently in these tenses as in others. This seems to confirm hypotheses 2a–c, support the main hypothesis, and suggest that the most influential general cognitive factor that constrains this variation is statistical preemption.

Regarding hypothesis 3, both for comprehension-to-production and production-to-production priming the type of long-lasting effect that is usually found in the psycholinguistic literature was documented (see, e.g., Pickering and Ferreira 2008). This suggests that speakers do not repeat the specific verb-forms and the argument-structure configurations that are stipulated by these, but rather that they repeat argument-structure constructions, namely, <**AdvP** *haber* **Subj**> and <**AdvP** *haber* **Obj**>. In turn, the interaction observed between the two priming modalities suggest that in non-experimental settings, production-to-production priming has a deeper impact than comprehension-to-production priming. This contradicts the results of earlier laboratory studies, which found the magnitude of the priming effect to be comparable (Bock et al. 2007: 452). On a methodological note, the importance of structural priming in this variation strongly suggests that priming effects should not be neglected in analyses of language variation and change, even more so because variationist (e.g., Labov 1994: 559; Poplack 1984: 213–214) and psycholinguistic inquiry (e.g. Pickering and Ferreira 2008: 429) have shown that virtually all levels of linguistic analysis (including phonology) display priming–like phenomena.[7]

The examination of the interaction between the cognitive factors has also unveiled an antagonistic relationship between, on the one hand, statistical preemption and, on the other, structural priming and the preference for unmarked coding. This is reminiscent of the roles these cognitive factors play in language acquisition and innovation. That is, in language acquisition, statistical preemption has been shown to be the mechanism that prevents children from overgeneralizing (Goldberg 2006: Chap. 5), whereas structural priming has been argued to promote the extension of perceived structures to new conceptualizations of the same type (Goldberg 2009: 107; Pickering and Ferreira 2008: 449–450). Regarding language innovation, Croft (2000: Chap. 5) argues that the tendency to maximize unmarked coding is the prime motivation for form-function reanalysis, which reforms established constructions or, put differently, overrules their preemptive effect.

The discussion on social factors, on the other hand, was mainly concerned with trying to establish the associations Dominicans establish between *haber* pluralization and social groups and with evaluating whether the variation corresponds to an ongoing linguistic change from below, as is claimed by hypotheses 4–8. The fact that educational achievement and speech style impose no significant constraints seems to suggest that the variation does not have a conven-

7 For example, Poplack (1984: 213–214) observes that speakers tend to repeat the same variant of Spanish word–final /–s/ in successive tokens.

tionalized meaning of stylistic appropriateness. Rather, the data indicate that the frequencies of the argument structure constructions signal primarily social class membership. In this regard, the fact that Dominicans associate high rates of *haber* pluralization with middle-class membership, together with the fact that *haber* pluralization is not associated to gender or style seems to support the hypothesis that the phenomenon constitutes an advanced ongoing language change from below.

In sum, the data that were reported in this article confirm the main hypothesis: in Santo Domingo the pluralization of presentational *haber* corresponds to a slowly progressing language change from below during which the argument–structure construction <**AdvP** *haber* **Subj**> is replacing the <**AdvP** *haber* **Obj**> pattern. The fact that the same results were obtained in an earlier investigation on San Juan Spanish strongly suggest that the statistical patterns reported in this and that study are not epiphenomenal. Rather, the consistency of the results supports that the language change affecting the presentational *haber* construction is constrained by three domain-independent general cognitive factors: markedness of coding, statistical preemption, and structural priming.

References

Alba, Orlando. 2004. *¿Cómo hablamos los dominicanos? Un enfoque sociolingüístico.* Santo Domingo: Grupo León Jimenes.

Baayen, Harald R. 2008. *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Baayen, Harald R. 2014. Multivariate statistics. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 337–372. Cambridge: Cambridge University Press.

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

Bentivoglio, Paola & Mercedes Sedano. 2011. Morphosyntactic variation in Spanish–speaking Latin America. In Manuel Díaz-Campos (ed.), *The handbook of Hispanic sociolinguistics,* 123–147. Oxford: Blackwell.

Blas-Arroyo, José L. 1999. *Lenguas en contacto. Consecuencias lingüísticas del bilingüismo social de las comunidades de habla del este peninsular.* Frankfurt am Main & Madrid: Vervuert/Iberoamericana.

Bock, Kathryn J. & Zenzi M. Griffin 2000. The persistence of structural priming: Activation or implicit learning? *Journal of Experimental Psychology: General* 129(2), 177–192.

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

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

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4), 711–733.
Bybee, Joan & Clyde Beckner. 2010. Usage-based theory. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis,* 827–856. Oxford: Oxford University Press.
Claes, Jeroen. 2014. A cognitive construction grammar approach to the pluralization of presentational *haber* in Puerto Rican Spanish. *Language Variation and Change* 26(2), 219–246.
Croft, William. 2000. *Explaining language change: An evolutionary perspective.* London & New York: Longman.
Croft, William. 2003. *Typology and universals.* Cambridge: Cambridge University Press.
D'Aquino-Ruiz, Giovana. 2004. *Haber* impersonal en el habla de Caracas. Análisis sociolingüístico. *Boletín de Lingüística* 21, 3–26.
D'Aquino-Ruiz, Giovana. 2008. El cambio linguistico de *haber* impersonal. *Núcleo* 20(25), 103–124.
DeMello, George. 1991. Pluralización del verbo *haber* impersonal en el español hablado culto de once ciudades. *Thesaurus* 46, 445–471.
Fauconnier, Gilles. 2007. Mental spaces. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 351–376. Oxford: Oxford University Press.
Fontanella de Weinberg, María B. 1992a. *El español de América.* Madrid: Mapfre.
Fontanella de Weinberg, María B. 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. 2003. Actitudes lingüísticas en torno a la pluralización de *haber* impersonal en los Andes venezolanos. *Interlingüística* 14, 375–382.
Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure.* Chicago: Chicago University Press.
Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.
Goldberg, Adele E. 2009. The nature of generalization in language. *Cognitive Linguistics* 20(1), 93–127.
Goldberg, Adele E. 2011. Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics* 22(1), 131–153.
Hothorn, Torsten, Kurt Hornik, Carolin Strobl, & Achim Zeileis. 2014. *Party: A laboratory for recursive partytioning*. http://cran.r-project.org/web/packages/party/index.html (accessed 10 May 2014).
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.
Johnson, Daniel E. 2014. *Rbrul version 2.22–Cotton Candy*. http://www.danielezrajohnson.com/Rbrul.R (accessed 10 April 2014).
Keenan, Edward. 1976. Towards a universal definition of subject. In Charles N. Li (ed.), *Subject and topic*, 305–333. New York: Academic Press.
Labov, William. 1972. *Sociolinguistic patterns.* Philadelphia: University of Pennsylvania Press.
Labov, William. 1994. *Principles of linguistic change. Vol. 1: Internal factors.* Oxford: Blackwell.
Labov, William. 2001. *Principles of linguistic change. Vol. 2: Social factors.* Oxford: Blackwell.
Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: Chicago University Press.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar. Vol. 1: Theoretical prequisites*. Stanford: Stanford University Press.
Langacker, Ronald W. 1991. *Foundations of cognitive grammar. Vol. 2: Descriptive application*. Stanford: Stanford University Press.
Milroy, Lesley & Matthew Gordon. 2003. *Sociolinguistics: Method and analysis*. Oxford: Blackwell.
Moreno-Fernández, Francisco. 2003. *Metodología del proyecto para el estudio sociolingüístico del español de España y América (PRESEEA)*. www.linguas.net/portalpreseea (accessed 10 March 2008).
Pérez-Martín, Ana 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, Martin J. & Victor S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin* 134(3), 427–459.
Poplack, Shana. 1984. Variable concord and sentential plural marking in Puerto Rican Spanish. *Hispanic Review* 52(2), 205–222.
Quintanilla-Aguilar, José R. 2009. *La (des)pluralización del verbo* haber *existencial en el español salvadoreño: ¿Un cambio en progreso?* Miami, FL: University of Florida dissertation.
R Core Team. 2013. *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org/ (accessed 20 February 2013).
Real Academia Española. 2008–. *Corpus diacrónico del español*. Retrieved from: http://corpus.rae.es/cordenet.html (accessed 10 October 2010).
Rivas, Javier & Esther Brown. 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.
Tagliamonte, Sali. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
Tagliamonte, Sali. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Oxford: Wiley-Blackwell.
Tagliamonte, Sali & Harald R. Baayen. 2012. Models, forests and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2), 135–178.
Ward, Gregory & Betty Birner. 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) (Real Academia Española, 2008–)

Constructions	Había		Hubo		Habría		Habrá		Haya		Hubiera		Hay		Ha habido	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
<*haber* past participle>	1806	52.4	38	6.5	6	10.5	41	13.4	202	32.3	110	50.0	0	0.0	0	0.0
<*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	0.0
<*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	0.0
Possessive *haber*	54	1.6	113	19.4	4	7.0	17	5.5	89	14.2	6	2.7	0	0.0	0	0.0
Presentational *haber*	870	25.2	406	69.8	43	75.4	173	56.4	295	47.2	78	35.5	3440	98.7	6	100.0
<*ha* time expression>	67	1.9	1	0.2	0	0.0	59	19.2	2	0.3	2	0.9	0	0.0	0	0.0
Total	3449	100	582	100	57	100	307	100	625	100	220	100	3485	100	6	100

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, since by that time speech communities had yet 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 use that had already largely decayed by that time in other types of sources.