

# John Benjamins Publishing Company



This is a contribution from *Revista Española de Lingüística Aplicada* 27:2  
© 2014. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible only to members (students and faculty) of the author's/s' institute. It is not permitted to post this PDF on the internet, or to share it on sites such as Mendeley, ResearchGate, Academia.edu.

Please see our rights policy on <https://benjamins.com/#authors/rightspolicy>

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: [www.copyright.com](http://www.copyright.com)).

Please contact [rights@benjamins.nl](mailto:rights@benjamins.nl) or consult our website: [www.benjamins.com](http://www.benjamins.com)

# Sociolingüística comparada y gramática de construcciones

## Un acercamiento a la pluralización de *haber* presentacional en las capitales antillanas

Jeroen Claes

Universiteit Antwerpen

En este artículo, estudiamos la pluralización de *haber* presentacional (e.g., *Habían fiestas*) en el español de La Habana, Santo Domingo y San Juan. Apoyándonos en la gramática de construcciones de Goldberg (1995), afirmamos que el fenómeno constituye un cambio lingüístico en marcha ‘desde abajo’: la variante pluralizada de la construcción presentacional con *haber* (<SAdv *haber* Sujeto>) está reemplazando la variante impersonal (<SAdv *haber* Objeto>). Mediante un análisis de regresión de efectos mixtos, mostramos que el caribeño pluraliza el verbo en un 41–46% de los casos. Los factores lingüísticos analizados en este estudio (la posición más natural del referente del SN en la cadena de acciones, la polaridad de la oración, el tiempo verbal, el *priming* de la comprensión a la producción y el *priming* de la producción a la producción) corroboran que se trata de una alternancia de estructura argumental. Los análisis sociolingüísticos comparados revelan que dichos factores tienen el mismo efecto y peso relativo en las tres variedades. Finalmente, para las tres comunidades, los resultados para el género y la clase social del hablante soportan que el fenómeno constituye un cambio desde abajo avanzado.

**Palabras clave:** gramática de construcciones, sociolingüística comparada, pluralización de *haber* presentacional, español cubano, español dominicano, español puertorriqueño

In this article, we investigate the pluralization of presentational *haber* (e.g., *Habían fiestas*. ‘There were parties.’) in the Spanish of Havana, Santo Domingo, and San Juan. Drawing on Goldberg’s (1995) Cognitive Construction Grammar, we claim that the phenomenon consists in a language change from below: the pluralized variant of the presentational *haber* construction (<AdvP *haber* Subject>) is replacing the impersonal variant (<AdvP *haber* Object>). Using

a mixed-effects regression analysis, we show that speakers of the Caribbean dialects pluralize the verb in 41–46% of the cases. The linguistic factors that were investigated in this study (typical action-chain position of the noun's referent, clause polarity, verb tense, comprehension-to-production priming and production-to-production priming) argue in favor of considering the variation an argument-structure alternation. The comparative sociolinguistic analyses reveal that these factors have the same effects and relative strengths in the three communities. For the three communities, the results for gender and social class support that the phenomenon constitutes an advanced language change from below.

**Keywords:** cognitive construction grammar, compared sociolinguistics, pluralization of presentational *haber*, Cuban Spanish, Dominican Spanish, Puerto Rican Spanish

## 1. Introducción

En español estándar, *haber* presentacional se usa como un impersonal: el único argumento nominal, *fiestas* en el ejemplo (1), desempeña la función de objeto directo y, por consiguiente, el verbo sólo se conjuga en la tercera persona del singular (Real Academia Española & Asociación de Academias de la Lengua Española, 2009, §41.6, entre otros muchos).

- (1) Entonces, él siempre estaba velando en el periódico donde era que *había*  
fiestas (SJ03H22/ SJ327).<sup>1</sup>

No obstante, los hablantes de muchas variedades del español, tanto de la Península (p.ej., Blas-Arroyo, 1999, cap. 2), como de Canarias (p.ej., Pérez-Martín, 2007) y América Latina (p.ej., DeMello, 1991; Kany, 1951, pp.255–260) tienden a hacer concordar el verbo con el grupo nominal (contémplese el ejemplo (2)), un fenómeno conocido en la bibliografía como la 'pluralización de *haber*'.

- (2) A sí, *han habido* huracanes fuertes aquí, sí (SD01H21/RD73).

En Latinoamérica, la investigación variacionista ha enfocado el español venezolano. En esa variedad, la variante pluralizada se ve apoyada por los SSNN de

1. En este texto, nos atendremos a las convenciones siguientes para indicar el origen de los ejemplos: el primer código identifica la muestra, empezando por una indicación de la ciudad (SJ=San Juan; SD=Santo Domingo; LH=La Habana), seguido por dos dígitos que indican el número del hablante y una letra que indica su género (H=hombre; M=mujer). Después viene un número que representa la edad del hablante (1=21–35 años; 2=mayor de 55 años) y su grado de instrucción (1=menos que universidad; 2=universidad). El código después de la barra identifica el caso en la base de datos.

referencia humana, la ausencia de negación, el imperfecto de indicativo, los tiempos verbales compuestos y las perífrasis (Bentivoglio & Sedano, 1989; D'Aquino-Ruiz, 2004; Díaz-Campos, 2003). Asimismo, la bibliografía revela que la pluralización se correlaciona con los grupos socioeconómicos bajos (Bentivoglio & Sedano, 1989; D'Aquino-Ruiz, 2004; Díaz-Campos, 2003) y el género femenino (Bentivoglio & Sedano, 1989). Igualmente, los estudios sociolingüísticos llevados a cabo en Venezuela han observado que la variante pluralizada está extendiéndose a más contextos lingüísticos (e.g., Díaz-Campos, 2003) y grupos sociales (p.ej., D'Aquino-Ruiz, 2008), lo cual, como Fontanella de Weinberg (1992a) ya había argumentado en base a una investigación documental, sugiere un cambio en curso (D'Aquino-Ruiz, 2008).<sup>2</sup>

De esta breve revisión bibliográfica se deduce que la pluralización de *haber* presentacional ocurre en varios dialectos hispánicos y que las alternancias entre las formas singular y pluralizada del verbo se ven restringidas por dos factores lingüísticos (el tiempo verbal y la referencia del SN) y factores sociales tales como el género y la clase social del hablante. Sin embargo, no contamos con análisis que puedan explicar la influencia de estos factores. Tampoco hemos podido localizar estudios sociolingüísticos recientes que comparen el uso de las variantes según se observa en La Habana, Santo Domingo y San Juan. Por consiguiente, sólo podemos afirmar que la pluralización de *haber* constituye un fenómeno generalizado en esas comunidades de habla (Alba, 2004, p. 323; Domínguez-Hernández, 2007, p. 22; Vaquero, 1978, pp. 135–140, 1996, p. 63). Sin embargo, una comparación pormenorizada de las capitales antillanas podría arrojar nueva luz sobre la pregunta de si la pluralización de *haber* constituye un lento cambio en marcha (*vid.* Wolfram & Schilling-Estes, 2003, p. 713), como afirman D'Aquino-Ruiz (2008) y Fontanella de Weinberg (1992a), o más bien una variación estable, como arguyen Díaz-Campos (2003) y Quintanilla-Aguilar (2009).

Por lo tanto, en este artículo trataremos de modelar las restricciones cognitivas que condicionan la pluralización de *haber* presentacional, así como su distribución social. Además, nos proponemos comparar los resultados obtenidos para cada una de las tres variedades. Para lograr estas metas, lanzamos las siguientes preguntas de investigación:

- I. ¿Qué factores cognitivos restringen la pluralización de *haber* presentacional?
- II. ¿Cuál es la distribución social de la pluralización de *haber* presentacional en La Habana, Santo Domingo y San Juan?

2. Para una revisión bibliográfica más amplia, remitimos al lector al capítulo de Bentivoglio y Sedano (2011).

- III. ¿Cómo pueden modelarse la distribución social y el efecto de los factores lingüísticos en la gramática de construcciones de Goldberg (1995, 2006)?
- IV. ¿Qué diferencias se detectan entre las tres variedades en lo que atañe a dichos factores y distribuciones?

## 2. Acercamiento teórico

Como ya deja entrever la tercera pregunta de investigación, el marco teórico que se asume en este estudio es la gramática de construcciones de Goldberg (1995, 2006). Este modelo cognoscitivo basado en el uso propone que tanto los aspectos idiosincráticos del lenguaje (p.ej., las entidades léxicas o las expresiones idiomáticas) como las generalizaciones lingüísticas (p.ej., la transitividad y otras abstracciones) constituyen pares de forma y significado, llamados *construcciones*. En otras palabras, la gramática de construcciones caracteriza el conocimiento lingüístico como una red jerárquica de esquemas constructivos (Goldberg, 2009, p.99), que le ofrecen al hablante símbolos para codificar conceptualizaciones (Langacker, 1987, p.409).

Considérese el ejemplo (4). En esta oración sencilla, se detectan no menos de siete construcciones, cada una con su significado: *Fulano, da, un, pedazo de, pan, María* y <**Sujeto le Verbo Objeto1 a Objeto2**>. <sup>3</sup>

(4) Fulano le da un pedazo de pan a María.

Contrariamente a las construcciones más específicas, esta última posee una semántica más abstracta y esquemática, que Goldberg (1995, cap. 6) resume como CAUSAR-RECIBIR ('CAUSE-RECEIVE', en inglés). En base a este significado, la construcción determina cuáles de los participantes son necesarios para que el evento pueda tener lugar (i.e., cuáles son los participantes perfilados), cuáles son sus papeles semánticos (en este caso, *agente, paciente y recibidor*), cómo se codifican a nivel sintáctico (en este caso, Sujeto, Objeto1 y Objeto2) y cuál es su estatus informativo. En otras palabras, en este marco teórico, es la construcción la que determina la estructura argumental y, por consiguiente, nada impide que un verbo se use en más de una configuración argumental.

3. Emplearemos las convenciones notacionales propuestas por Goldberg (1995): usaremos la letra negrita para señalar argumentos perfilados. Asimismo, conviene destacar que el esquema constructivo no especifica el orden lineal de los argumentos.

## 2.1 Descripción de las variantes y su distribución social

Sobre este trasfondo teórico, podemos considerar que la variación entre las formas impersonal y pluralizada de *haber* presentacional reviene a una competencia entre dos variantes de la construcción presentacional con este verbo. Como hemos argumentado en otro lugar, en el plano estructural, las dos alternativas incluyen un sintagma adverbial (SAdv; Lyons, 1967; Hernández-Díaz, 2003) y un único argumento nominal, perfilados ambos (Claes, 2014). En la variante pluralizada, el participante nominal funciona como sujeto; en la variante impersonal, desempeña el papel de objeto.

En cuanto al significado de las construcciones, las dos variantes se refieren a la escena SEÑALAR (*'POINTING-OUT'*, en inglés) propuesta por Lakoff (1987):

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, p. 490).

Como las dos construcciones codifican el mismo evento abstracto, también incluyen las mismas especificaciones pragmáticas y semánticas. O sea, en las dos variantes, el argumento nominal está asignado el rol argumental 'cero' (Langacker, 1991, p. 289) y debe proporcionar información nueva con respecto a la consciencia, las creencias y/o los conocimientos del oyente (Ward & Birner, 1995).

Sin embargo, estas similitudes no significan que no se detecten diferencias de significado entre las variantes. A este respecto, en Claes (2014) argüimos que el uso cuantitativamente diferenciado de las variantes le permite al hablante ubicarse socialmente.<sup>4</sup> En ese artículo, planteamos también que las alternancias construccionales que no impliquen contrastes semántico-pragmáticos incluyen especificaciones sociolingüísticas, que conectan una categoría social específica con el uso de una de las alternativas en una proporción determinada de los contextos donde las dos hubieran podido presentarse con igual probabilidad.

---

4. Silverstein (2003) argumenta que, al observar una de las variantes en determinada proporción de los casos en que hubieran podido presentarse todas con igual probabilidad, el hablante inferirá que esta distribución señala que el hablante forma parte de un grupo social determinado. Por supuesto, esto presupone que la alternancia no vehicula otras diferencias semánticas (p.ej., de organización figura-fondo) o pragmáticas (p.ej., de estatus informativo) (Lavandera, 1984, cap. 1).

### 3. Hipótesis

La revisión bibliográfica y la perspectiva teórica nos llevan a plantear la siguiente hipótesis principal:

HP: En las capitales antillanas, la pluralización de *haber* presentacional consiste en un paulatino cambio desde abajo en curso: la variante pluralizada de la construcción presentacional con *haber* (<SAdv *haber* Sujeto>) está reemplazando la variante impersonal (<SAdv *haber* Objeto>). Las diferencias entre ambas residen en la función sintáctica del grupo nominal (impersonal: objeto, pluralizada: sujeto) y los grupos sociales asociados a sus frecuencias relativas.

Por supuesto, la hipótesis principal es muy amplia y, de por sí, no permite hacer predicciones sobre el comportamiento de nuestro fenómeno. Sin embargo, a través de unos principios fundamentales del uso (la preferencia por la codificación no marcada; Langacker, 1991, p.298), la organización (la disuasión estadística; Goldberg, 2006, p.94), y el cambio lingüísticos (los principios del cambio lingüístico de Labov, 2001), implica una lista de extrapolaciones más detalladas. Por ejemplo, si tomamos en cuenta la preferencia generalizada por la codificación no marcada, es decir, si tomamos en consideración que el hablante tiende a inclinarse más por una función sintáctica determinada si la conceptualización que quiere codificar se asemeja al prototipo de dicha función (Langacker, 1991, p.298), la hipótesis principal implica que:

H1: Los hablantes seleccionarán la variante pluralizada más a menudo con SSNN que se parecen más a un sujeto prototípico.

Además, Bybee (2003, p.153) arguye que las formas que ocurren en más de un esquema construccional disponen de representaciones mentales independientes más fuertes que ítems léxicos que sólo se observan en un patrón. De hecho, si una forma se usa frecuentemente en una sola construcción, ocurriendo sólo esporádicamente en otros patrones, dicha forma se almacenará mentalmente como un caso particular del esquema más abstracto, cuya representación mental es mucho más accesible que tanto la de la forma independiente como la del esquema abstracto (Goldberg, 1995, p.79; Langacker, 1987, pp.59–60). El fácil acceso a esta expresión parcialmente prefabricada, a su vez, desfavorece el uso de la forma en otros patrones para codificar conceptualizaciones que podrían expresarse mediante el trozo de habla almacenado. A esta tendencia Goldberg (2006, p.94, 2011) denomina ‘disuasión estadística’ (*‘statistical pre-emption’*, en inglés). Además de explicar por qué el hablante no sobregeneraliza (Goldberg, 2011), esta tendencia también acarrea que, cuando surgió la variante pluralizada como una alternativa

convencional, las expresiones basadas en <SAdv haber Sujeto> no se hubieran empleado frecuentemente para codificar conceptualizaciones que involucraban alguna forma temporal de *haber* que se empleaba mayoritariamente en el esquema presentacional en tiempos anteriores. Como este patrón se suele conservar por repetición (Bybee, 2006, p. 715), obtenemos las predicciones siguientes:

- H2a: Si la forma terciopersonal de un tiempo verbal de *haber* se usaba frecuentemente fuera del esquema impersonal antes de la actualización del cambio, la variante pluralizada se verá favorecida por ese tiempo verbal.
- H2b: Los otros tiempos verbales desfavorecerán la variante pluralizada, por lo menos, siempre y cuando la conceptualización pueda codificarse mediante una forma afianzada (*'entrenched'*, en inglés) de la variante impersonal.

Ya que estas hipótesis atribuyen el efecto del tiempo verbal al grado de afianzamiento (*'entrenchment'*, en inglés) de la forma de *haber* en la estructura presentacional impersonal, la hipótesis siguiente se impone:

- H2c: Cuando la necesidad de codificar una conceptualización modalmente/aspectualmente más compleja obligue al hablante a construir una nueva expresión en vez de usar la forma afianzada del esquema impersonal, la variante pluralizada ocurrirá tan frecuentemente en los tiempos verbales indicados en la H2b como en los tiempos verbales descritos en la H2a.

Asimismo, la investigación sico- y sociolingüística ha revelado que el hablante tiende a reciclar las estructuras abstractas que han aparecido en el discurso precedente, un fenómeno conocido en la bibliografía como *'priming estructural'* (p.ej., Goldberg, 2006, pp. 120–125; Labov, 1994, cap. 20). Por lo tanto, si la pluralización de *haber* presentacional constituye una alternancia de estructura argumental, como establece la hipótesis principal, hallaremos la distribución siguiente:

- H3: La ocurrencia de una de las variantes en el discurso precedente a un contexto variable favorecerá el uso de la misma variante en dicho contexto.

La hipótesis principal implica también que la variación dará cuenta de los patrones de estratificación social y estilística que se han mostrado típicos de los cambios desde abajo. Más específicamente, la hipótesis del tiempo aparente (Labov, 1994, pp. 43–72) nos lleva a afirmar que:

- H4: Mientras más joven sea el hablante, más se inclinará por la variante pluralizada.

El Principio del Género propuesto por Labov (2001, p. 292), a su vez, nos lleva a especular que:

- H5: Las féminas respaldarán la variante pluralizada.



Además, Labov (2001, p. 188) arguye que el Patrón Curvilíneo es el indicio más claro de los cambios desde abajo. Acarrea la hipótesis siguiente:

H6: La clase media dará cuenta de las más elevadas tasas de uso de la variante pluralizada.

De la misma manera, en situaciones de cambio desde abajo, las variantes innovadoras suelen, o bien, usarse más en contextos formales, o bien, no correlacionarse del todo con la formalidad del encuentro (Labov, 2001, cap. 3; Silva-Corvalán, 2001, pp. 157–158), lo cual implica la hipótesis 7:

H7: Según vaya subiendo el grado de formalidad de la interacción, la frecuencia de la variante pluralizada no bajará.

Adicionalmente, se sabe que los hablantes cultos tienden a conformarse más con las normas de prestigio supralocales (Labov, 1972, p. 138), lo que sugiere que:

H8: Una formación académica más prolongada desfavorecerá el uso de la variante pluralizada, mientras que una instrucción formal más limitada lo beneficiará.

Finalmente, si la variación constituye un cambio en marcha, cabe suponer (Labov, 2010, p. 5) que:

H9: Las variedades darán cuenta de etapas diferentes de la evolución hacia la variante pluralizada.

Para poner a prueba estas hipótesis, realizamos tres análisis de regresión logística de efectos mixtos y comparamos nuestros resultados. Consideremos ahora la metodología del presente estudio.

## 4. Metodología

### 4.1 La muestra

Nuestros análisis se fundamentan en tres muestras muy recientes –la recogida de datos se llevó a cabo entre marzo y junio del 2011– de aproximadamente tres veces 24 horas de grabaciones a hablantes nativos del español antillano, residentes de, respectivamente, la Provincia de Ciudad de La Habana y las aglomeraciones urbanas conocidas como Gran Santo Domingo y el Área Metropolitana de San Juan. En las tres capitales, el muestreo se realizó mediante cuotas (24 hablantes por ciudad ó 3 hablantes por celda) a partir de los mismos tres parámetros sociales: el género (hombre vs. mujer), la edad (21–35 años vs. mayor de 55 años) y el nivel

de instrucción (universitarios vs. no universitarios). Además, con el fin de recoger más contextos variables e investigar si las alternancias se ven condicionadas por la formalidad del encuentro, estructurábamos las sesiones de grabación en tres partes:

1. Entrevista semidirigida sobre las experiencias de los participantes. Para investigar el *priming* de la comprensión a la producción lingüística, incluimos en el formato de las entrevistas (basado en Tagliamonte, 2006, ap. B; Moreno-Fernández, 2003, pp. 11–15; Quintanilla-Aguilar, 2009, ap. F) algunas preguntas con la variable, ejemplificadas en (5). En dichas preguntas, alternábamos el singular y el plural del verbo.

(5) Entrevistador: ¿Cuántos estudiantes *podía(n) haber* en tu época?

2. Lectura en voz alta de un cuento infantil en que se habían insertado 31 contextos de elección (20 de *haber* presentacional, 11 de un verbo distractor) (véase el ejemplo 6).

(6) En una pequeña aldea, *había/habían* un anciano padre y sus dos hijos...

3. Lectura en voz alta de un cuestionario de 45 ítems (32 de *haber* presentacional y 13 ítems distractores) precedidos de una historieta que evocaba el contexto pragmático apropiado (ver (7)).<sup>5</sup>

(7) A Inés le acaban de robar el carro, que tenía aparcado en algún callejón oscuro. 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

#### 4.2 Post-estratificación: clase social

Luego, post-estratificamos la muestra según el índice de clase social, que definimos como una función de las condiciones de vivienda, el nivel de instrucción y la ocupación del informante. Con el fin de realizar estimaciones realistas de las contribuciones de esos factores al prestigio social, pedíamos a los informantes que jerarquizaran las variables según su importancia para el estatus social. Luego, multiplicamos las cotizaciones obtenidas para los factores individuales con la posición promedio de la variable en la escala y sumamos estos números. Después, los convertimos en porcentajes del valor máximo y procedimos a juntar a nuestros

5. Si las destrezas de lectura o la vista del informante no le permitían leer las pruebas, el entrevistador se las leía, preguntándole por su preferencia.

informantes en tres grupos iguales: 0–33%: clase baja, 34–66%: clase media y 66–100%: clase alta.

### 4.3 El contexto variable

La mayoría de los estudios anteriores han optado por no incluir las formas del presente de indicativo. Sin embargo, como nosotros recogimos un puñado de ejemplos de la forma *hayn* para cada una de las tres variedades, no pudimos excluir este tiempo verbal de nuestros análisis. Algunas investigaciones anteriores (p.ej., DeMello, 1991; Quintanilla-Aguilar, 2009) han extendido el contexto variable a la forma *habemos* y sus variantes temporales (contémplese el ejemplo 8). Sin embargo, hay razones para suponer que este uso de *haber* no corresponde con las variantes que describimos en el apartado 2, sino con una tercera construcción emparentada, pues se observa un contraste semántico muy marcado entre los ejemplos (8) y (9). O sea, mientras que el ejemplo (8) implica que el hablante era uno de los niños de la cuadra, en la variante (9), el hablante desaparece del escenario.

(8) *Habíamos* muy pocos niños en la cuadra (LH19M11).

(9) *Habían* muy pocos niños en la cuadra.

### 4.4 Herramientas estadísticas

Después de transcribir las muestras y de buscar y codificar todos los casos de *haber* + SN plural, realizamos tres análisis de regresión logística de efectos mixtos mediante Rbrul (Johnson, 2009). Este tipo de análisis, a diferencia de los modelos lineales generalizados (i.e., el modelo de regresión implementado en, por ejemplo, GoldVarb X; véase Johnson, 2009; Tagliamonte & Baayen, 2012), nos permite asegurarnos de que los patrones estadísticos reflejen el efecto de las variables independientes y no la variación aleatoria introducida por algunos hablantes o ítems léxicos (Johnson, 2009, p. 365).

Como el GoldVarb X, el Rbrul arroja los resultados en formato de pesos probabilísticos o índices de probabilidad. En la Tabla 2, los índices de probabilidad mayores del .50 apuntan a que el factor favorece la aparición de la variante, mientras que los menores del .50 indican que la detiene. Cuanto más se aleja el valor del peso de este número, más el factor restringe las alternancias. Además, en la nota a esa misma tabla, proporcionamos las desviaciones de los modelos, que expresan cuán bien el modelo se ajusta a los datos. Idóneamente, los valores de estos índices se acercan más a cero.

Para evaluar las hipótesis, investigamos el efecto de las siguientes variables independientes, que discutiremos más adelante: la posición más natural del referente

del SN en la cadena de acciones, el tiempo verbal, la polaridad de la oración, el *priming* de la comprensión a la producción, el *priming* de la producción a la producción y el estilo de habla, así como la clase social, la edad, el género y el nivel de instrucción del hablante. Además de estos factores fijos, también incluimos una variable aleatoria para los hablantes y los lemas de los SSNN. Sin embargo, ya que gran parte de nuestros datos proviene de conversaciones semidirigidas, estos dos parámetros aleatorios resultaban ser colineales, por lo que nos vimos obligados a efectuar análisis paralelos para estos dos factores. A continuación, sólo discutiremos aquellos patrones que resultaron ser significativos para todos los lemas y todos los hablantes de la muestra en cuestión.

#### 4.5 Sociolingüística comparada

Estableceremos comparaciones entre nuestros resultados según la metodología que Tagliamonte (2002, p. 729) denomina ‘sociolingüística comparada’. En concreto, esta autora destaca tres niveles en los que se pueden encontrar contrastes entre dos o más variedades. La primera oposición radica posiblemente en el hecho de que el análisis de regresión no selecciona las mismas variables independientes para todas las variedades bajo estudio. El segundo contraste posible lo constituye el efecto de los niveles individuales de las variables independientes, el cual se puede comparar jerarquizando los pesos arrojados para los grupos de factores. Finalmente, el tercer contraste posible se revela contraponiendo la importancia relativa de los parámetros puestos a prueba. A este respecto, Tagliamonte (2002) propone jerarquizar las variables independientes según el rango, es decir, según el intervalo comprendido entre el valor probabilístico más elevado y el más bajo. Sin embargo, este procedimiento resulta no ser el más adecuado, ya que el rango obtenido para un grupo de factores depende de la distribución de los datos, la que podría ser muy desigual en las muestras contrastadas, y el número de niveles por variable, siendo mayor el rango de los grupos que incluyen más factores. Para remediar esto, estableceremos la importancia relativa de las restricciones calculando con cuánto se reduce la precisión de un modelo de Bosque Aleatorio (*‘Random Forest’*, en inglés) al sustituir un nivel de una variable independiente por otro aleatorio del mismo grupo de factores.<sup>6</sup> Como los bosques aleatorios no hacen ninguna suposición acerca de la distribución de los datos (Tagliamonte & Baayen, 2012) y el peso relativo de la variable se calcula sin tomar en cuenta el número de niveles por variable, este procedimiento debería corregir los fallos del método basado en el rango.

6. Véase el Apéndice A de Tagliamonte y Baayen (2012) para el procedimiento completo y los comandos *R* necesarios para llevar a cabo las operaciones estadísticas.

## 5. Resultados

Como se desprende de la Tabla 1, los hablantes de las tres variedades pluralizan *haber* presentacional en un 40–45% de los contextos variables. En lo que se refiere a la tasa de pluralización, pues, no se detectan diferencias tajantes entre los dialectos antillanos.

**Tabla 1.** Distribución de las variantes impersonal y pluralizada de la construcción presentacional con *haber*

	<i>La Habana</i>		<i>Santo Domingo</i>		<i>San Juan</i>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Variante pluralizada	934	44.6	859	46.7	684	41.3
Variante impersonal	1159	55.4	982	53.3	971	58.7
Total	2093	100	1841	100	1655	100

Al parangonar estos resultados con los de otros estudios anteriores, observamos que en las capitales caribeñas se pluraliza significativamente menos que en, por ejemplo, Caracas (50–60%; Bentivoglio & Sedano, 1989; D’Aquino-Ruiz, 2004; Díaz-Campos, 2003), San Cristóbal de Los Andes (82%; Freites-Barros, 2008), San Salvador (79.6%; Quintanilla-Aguilar, 2009) o Valencia de Venezuela (79.3%; Navarro-Correa, 1992). Sin embargo, la diferencia entre nuestros resultados y los demás parece estribar en que nosotros optamos por no excluir las formas *hay* — *hayn*. Sin éstas, la frecuencia de la variante pluralizada alcanza el 60.6% (N=926/1527) en La Habana, el 61.2% (N=835/1320) en Santo Domingo y el 54.3% (N=661/1217) en San Juan. A continuación, discutiremos los resultados presentados en la Tabla 2 por orden de las hipótesis, empezando con el efecto de la posición más natural del referente del SN en la cadena de acciones.

**Tabla 2.** Frecuencia absolutas, relativas y pesos probabilísticos para la variante pluralizada de la construcción presentacional con *haber*

	<i>La Habana</i> <sup>a</sup>				<i>Santo Domingo</i> <sup>b</sup>				<i>San Juan</i> <sup>c</sup>			
	<i>N</i>	%	<i>P(L)</i>	<i>P(H)</i>	<i>N</i>	%	<i>P(L)</i>	<i>P(H)</i>	<i>N</i>	%	<i>P(L)</i>	<i>P(H)</i>
<b>Tiempo verbal</b>												
Todas las demás	819/ 1298	63.1	.79	.76	720/ 1103	65.3	.80	.76	622/ 1014	61.3	.82	.81
Expresiones sintéticas en presente y pretérito	115/ 795	14.5	.21	.24	140/ 739	18.9	.20	.24	62/ 641	9.7	.18	.19
Rango			58	52			60	52			64	62

Tabla 2. (continúa)

	<i>La Habana</i>				<i>Santo Domingo</i>				<i>San Juan</i>				
	N	%	P (L)	P(H)	N	%	P(L)	P(H)	N	%	P(L)	P(H)	
<b>Priming de la producción a la producción</b>													
Variante pluralizada	556/ 817	68.1	.68	.67	484/ 711	68.1	.70	.68	352/ 558	63.1	.66	.65	
Primera ocurrencia/dis- tancia 20+ oraciones	83/ 297	27.9	.44	.42	123/ 337	36.5	.47	.44	88/ 246	35.8	.46	.45	
Variante impersonal	295/ 979	30.1	.37	.41	253/ 794	31.9	.33	.37	244/ 851	28.7	.38	.39	
Rango			31	26			37	31			28	26	
<b>Priming de la comprensión a la producción</b>													
Variante pluralizada	113/ 239	47.3	.62	.58	151/ 264	57.2	.61	.62	92/ 175	52.6	.63	.64	
Variante impersonal	73/ 204	35.8	.47	.49	63/ 185	34.1	.47	.46	30/ 125	24.0	.44	.43	
Primera ocurrencia/dis- tancia 20+ oraciones	748/ 1650	45.3	.41	.43	646/ 1393	46.4	.42	.41	562/ 1355	41.5	.44	.43	
Rango			21	15			19	21			19	21	
<b>Posición típica del referente del SN en la cadena de acciones</b>													
Cabeza	467/ 925	50.5	.57	.56	439/ 815	53.9	.61	.61	338/ 730	46.3	.58	.58	
Cola y circunstancia	467/ 1168	40.0	.44	.44	421/ 1027	41.0	.39	.39	346/ 925	37.4	.42	.42	
Rango			13	12			22	22			16	16	
<b>Clase social</b>													
Clase baja	374/701	53.4	.67	.61	301/ 662	45.5	.47	.46	No significativo				
Clase media	330/ 700	47.1	.53	.52	248/ 461	53.8	.60	.60					
Clase alta	230/ 692	33.2	.31	.37	311/719	43.3	.43	.44					
Rango			36	24			17	16					
<b>Polaridad de la oración</b>													
Afirmativa		No significativo				No significativo				559/ 1225	45.6	.57	.57
Negativa										125/ 430	29.1	.43	.43
Rango											14	14	
<b>Género</b>													
Femenino		No significativo				No significativo				375/ 836	44.9	.55	.55
Masculino										309/ 819	37.7	.45	.45
Rango											10	10	

Tabla 2. (continúa)

	La Habana		Santo Domingo		San Juan	
	N	% P (L) P(H)	N	% P (L) P(H)	N	% P (L) P(H)
<b>Edad: Clase social (interacción)</b>						
20–35 años: clase baja					125/278	45.0 .56 .56
20–35 años: clase media		No significante		No significante	121/273	44.3 .47 .49
20–35 años: clase alta					132/329	40.1 .46 .47
Rango						10 .9
55+ años: clase media					104/179	58.1 .71 .70
55+ años: clase baja					46/114	40.4 .47 .46
55+ años: clase alta					156/482	32.4 .33 .32
Rango						38 .38

*Nota:* P significa 'Peso'. (L) significa 'con los lemas'. (H) significa 'con los hablantes'.

<sup>a</sup> Con los lemas: desviación: 2019.93; Con los hablantes: desviación: 2049.45

<sup>b</sup> Con los lemas: desviación: 1829.93; Con los hablantes: desviación: 1889.31

<sup>c</sup> Con los lemas: desviación: 1519.56; Con los hablantes: desviación: 1549.19

## 5.1 Factores cognitivos

### 5.1.1 La preferencia por la codificación no marcada.

La revisión bibliográfica presentada en la introducción informa que la variante pluralizada se ve favorecida por la referencia humana (p.ej., Bentivoglio & Sedano, 1989). Además, los estudios de Esther Brown y Javier Rivas han mostrado que favorecen la pluralización los nombres que se usan frecuentemente como sujetos en español (Brown & Rivas, 2012) y los SSNN que se refieren a entidades no delimitadas temporalmente, como *computadores* en (10) ('los SSNN de nivel individual'; Rivas & Brown, 2012).

- (10) E, eso, e, de seguro no había televisión y, e, no *habían* computadores  
(SD04M22/ RD437).

Sin embargo, no existe una relación directa entre, por un lado, la función de sujeto y, por otro, las características que modelan estos parámetros. Más bien, en Claes (2014) hemos mostrado que estas tres variables independientes modelan facetas del rasgo más saliente de los sujetos prototípicos, a saber, su agentividad (Comrie, 1989, p. 120; Keenan, 1976, p. 321). Por lo tanto, en este estudio, codificamos los datos según la posición más natural del referente del SN en la cadena de acciones, representada en la Figura 1.

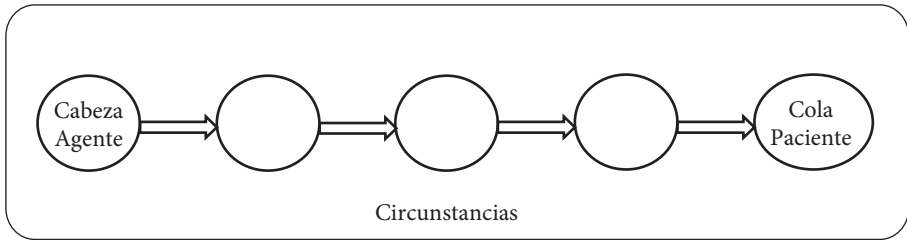


Figura 1. El modelo canónico de eventos (adaptado de Langacker, 1991, pp. 283, 284)

Tocante a ello, la Tabla 2 indica que, en las tres variedades, propician la construcción pluralizada los nombres cuyos referentes pueden ser imaginados fácilmente como cabezas de una cadena de acciones (p.ej., *ciclón* en 11).

- (11) Pero después, *sí hubieron* muchos ciclones, como diez ciclones *hubieron* más (SD01H21/RD81).

En cambio, claramente detienen esta variante los nombres cuyos referentes se sitúan más bien hacia la cola de la cadena de acciones (p.ej., *muertos* en 12) o los nombres que se refieren a las circunstancias de una acción (p.ej., *restaurantes* en (13)).

- (12) *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).
- (13) Este, aquí *hay* desde restaurantes que son, o sea que no son ba, son baratos, hasta restaurantes bien caros (SJ01M22/SJ04).

Además, aunque en la bibliografía tipológica no se conecta la presencia o ausencia de negación con ninguna función sintáctica determinada, la polaridad de la oración podría ser un factor que explique por qué el hablante se inclina más por la variante pluralizada o la impersonal. Más específicamente, Prince (1992, pp. 299–300) apunta que los SSNN de las oraciones presentacionales afirmativas no pueden recibir interpretaciones genéricas. Más bien, la construcción presentacional lleva al oyente a interpretar el argumento nominal como un *token* específico desconocido, o, en términos de la jerarquía de la definitud/especificidad presentada en (14), como un nombre indefinido específico (Croft, 2003, p. 132). En oraciones de polaridad negativa, el operador *no* suspende la referencia del SN (Keenan, 1976, p. 318), por lo que se interpreta como un tipo o, en otras palabras, una entidad indefinida no específica (Croft, 2003, p. 132).



- (14) Definido > indefinido específico > indefinido no específico  
(Langacker, 1991, p. 308).

En relación con esto, la bibliografía tipológica (p.ej., Croft, 2003, p. 132; Langacker, 1991, p. 308) indica que las lecturas indefinidas no específicas son prototípicas de los objetos directos, lo que explica por qué el boricua tiende a elegir la variante con objeto directo para codificar conceptualizaciones que incluyen negación (contéplense el ejemplo (15) y la Tabla 2).

- (15) O sea, otra escuela superior y ésa me gustó mucho y no *hubo* problemas, y es, y estaba cerca de mi casa, así que yo podía ir a pie (SJ10H12/SJ1186).

No obstante, esto no aclara por qué la polaridad de la oración no resultó ser una restricción significativa en La Habana y Santo Domingo. Sin embargo, como estas dos comunidades de habla también dan cuenta de tasas de pluralización un poco más altas, esto podría apuntar a que la pluralización de *haber* “progresivamente pierde restricciones de empleo y se generaliza” (Company-Company, 2003, p. 26), es decir, constituye un cambio en marcha.

### 5.1.2 *La disuasión estadística.*

La evaluación de las hipótesis 2a–c presupone las respuestas a dos interrogantes adicionales. Primero ¿cuándo fue que la variación empezó a darse a nivel comunitario? y, segundo ¿cuál era la distribución de las formas temporales de *haber* en el periodo justamente anterior? En lo que atañe a la primera interrogante, hay indicios documentales de que ya para finales del siglo XVIII el fenómeno ocurría con cierta frecuencia en español bonaerense (Fontanella de Weinberg, 1992a). Como suele haber un periodo considerable entre la actualización de un cambio lingüístico en la lengua hablada y su reflejo en fuentes escritas, la variación –como la mayoría de los rasgos más prominentes del español latinoamericano (De Granda, 1994, cap. 1; Fontanella de Weinberg, 1992b, cap. 1)– probablemente surgiría durante el siglo XVII. Por consiguiente, para contestar la segunda pregunta, mejor enfoquemos el periodo comprendido entre 1492 y 1600.

En el apéndice A, presentamos la distribución de unas 9000 formas tercio-personales de *haber* recogidas mediante búsquedas sistemáticas en las secciones latinoamericanas del CORDE (Real Academia Española, 2008-) que cubren ese periodo. Este análisis revela que antes de la actuación del cambio, *hay* y *hubo* se usan sobre todo en oraciones presentacionales, lo que sugiere que sus representaciones más accesibles serían: <SAdv *hay* Objeto> y <SAdv *hubo* Objeto>. Las demás formas de *haber*, o bien, se emplean frecuentemente en dos o más construcciones, o bien, son muy poco frecuentes, lo que sugiere que sus representaciones más prominentes serían sus formas independientes.

Ahora bien, esta distribución resulta ser el mejor predictor del comportamiento de los hablantes de las tres comunidades de habla. Además, la Tabla 2 deja ver que el pretérito y el presente de indicativo, para los que planteamos las representaciones <SAdv hay Objeto> y <SAdv hubo Objeto>, desfavorecen la variante pluralizada, por lo menos siempre y cuando la complejidad de la conceptualización permita el uso de sus formas sintéticas. En cambio, según indica la Tabla 3, cuando la conceptualización es aspectual o modalmente demasiado compleja como para codificarse mediante estas formas afianzadas de la variante impersonal, se observa la variante pluralizada tan frecuentemente en el pretérito y el presente de indicativo como en otros tiempos verbales (contémplese el ejemplo 16).

- (16) O sea que pese a que *pueden haber* muchas cosas que andan mal como n, no la gente, me incluyo, no, no reaccionamos (SD24M12/RD3175).

**Tabla 3.** Frecuencias absolutas y relativas para la variante pluralizada de la construcción presentacional con *haber* en el pretérito y el presente de indicativo, según forma verbal sintética vs. perifrástica

	<i>La Habana</i>		<i>Santo Domingo</i>		<i>San Juan</i>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Expresiones sintéticas en presente y pretérito de indicativo	115/795	14.5	140/739	18.9	62/641	9.7
Expresiones perifrásticas en presente y pretérito de indicativo	133/206	64.6	92/143	64.3	84/124	67.7

Nota:  $p=0.000$  para el contraste entre las primeras dos filas.

### 5.1.3 *El priming estructural.*

Para evaluar la hipótesis 3, codificamos todas las ocurrencias del corpus según la distancia hasta el último caso (0–20 verbos conjugados vs. más de 20), la variante que aparece en dicho ejemplo, y si fue el informante (el *priming* de la producción a la producción) o el investigador (el *priming* de la comprensión a la producción) quien lo proporcionó. Ahora bien, al contemplar la Tabla 2, observamos que los resultados para ambas modalidades del *priming* apuntan hacia la misma dirección: las variantes pluralizadas impulsan las pluralizadas y viceversa. Estos hallazgos constituyen un argumento fuerte a favor de la hipótesis principal, pues sugieren que el hablante no repite la forma verbal específica y la estructura argumental detallada por la misma, sino que repite una entidad más abstracta, a saber, la construcción de estructura argumental.

Además, el *priming* parece ofrecer una explicación para contraejemplos aparentes como (17), en los que el hablante establece la concordancia con un pronombre acusativo.

- (17) Y, entonces, *los<sub>Acc</sub> habían* disciplinados, pero *los<sub>Acc</sub> habían* que eran la candelita (LH22H11/LH2833).

O sea, en la Tabla 4 observamos que los hablantes de las tres comunidades de habla son más susceptibles a producir la concordancia objeto-verbo cuando acaban de producir o procesar una variante pluralizada. En otras palabras, antes de invalidar la hipótesis principal, estas ocurrencias parecen ejemplificar *blends* construccionales que surgen en línea debido al alto nivel de activación de <SAdv *haber* Sujeto>.

**Tabla 4.** Frecuencias absolutas y relativas de las formas pluralizadas de *haber* que concuerdan con un pronombre acusativo, según el *priming* de la comprensión a la producción y el *priming* de la producción a la producción

	<i>La Habana</i>		<i>Santo Domingo</i>		<i>San Juan</i>	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Primera ocurrencia/no hay distancia 20+ oraciones	0/2	0.0	1/2	50.0	0/2	0.0
Variante impersonal	13/70	18.6	5/49	10.2	9/47	16.1
Variante pluralizada	36/59	61.0	24/54	44.4	22/48	45.8
Total	49/131	37.4	30/105	28.6	31/106	29.2

Nota:  $p = 0.000$

## 5.2 Jerarquía de los factores cognitivos

Comparemos ahora la importancia relativa de los factores lingüísticos considerados en este estudio. Aunque en la Tabla 6, sólo tabulamos el peso de estos factores, para su cómputo, también incluimos las variables sociales en los respectivos modelos de bosque aleatorio. Optamos por dejar fuera los factores aleatorios, pues los bosques aleatorios tienden a sobreestimar el peso de este tipo de variables independientes (Baayen, 2014, p. 366).

En términos generales, no se detecta ninguna diferencia entre el peso de las variables en las tres comunidades de habla, por lo menos si hacemos caso omiso por el momento de la polaridad de la oración, que, como hemos visto anteriormente, no alcanza significancia estadística en La Habana y Santo Domingo. Como las hipótesis 1–3 y 9 pretenden que las tres variedades están pasando por el mismo cambio lingüístico que, además, se ve restringido por los mismos tres factores cognitivos, este resultado es completamente esperado (Tagliamonte, 2002, pp. 732–733, 2006, p. 246).

No obstante, si hubiéramos jerarquizado los factores según el rango probabilístico, hubiéramos llegado a conclusiones muy diferentes. Más concretamente, debido a la distribución más desequilibrada de los datos cubanos y sanjuaneros, el método basado en el rango le atribuiría más importancia al *priming* de la comprensión a la producción lingüística. Por la misma razón, la comparación en base al rango nos llevaría a considerar la variable polaridad de la oración como igual de importante que la posición más natural del referente del SN en la cadena de acciones. Esto, a su vez, no sería congruente con la bibliografía tipológica, que clasifica la definitud como un rasgo menos importante de los sujetos prototípicos (Langacker, 1991, p. 312), ni con el hecho de que la variante pluralizada ya parece haber invadido el terreno conceptual indefinido no específico en La Habana y Santo Domingo. Los modelos de bosque aleatorio, en cambio, corrigen estas limitaciones y dejan ver que la polaridad de la oración es la restricción de menor peso en San Juan.

En lo que se refiere a la jerarquización de los niveles de las variables lingüísticas, la Tabla 2 muestra que todos los factores presentan el mismo orden, con la salvedad del *priming* de la comprensión a la producción. Tocante a esta variable, parece que el habanero y el capitaleño son más propensos que el sanjuanero a emplear una variante pluralizada cuando acaban de procesar una variante impersonal. De todas formas, la tendencia general apunta en la misma dirección: las variantes pluralizadas favorecen las pluralizadas, mientras que las impersonales impulsan las impersonales.

**Tabla 5.** Restricciones lingüísticas en La Habana, Santo Domingo y San Juan, jerarquizadas según su importancia relativa para la precisión de los modelos de bosque aleatorio y el rango probabilístico

	<i>La Habana</i>		<i>Santo Domingo</i>		<i>San Juan</i>	
	<i>Peso relativo</i>	<i>Rango</i>	<i>Peso relativo</i>	<i>Rango</i>	<i>Peso relativo</i>	<i>Rango</i>
Tiempo verbal	0.0998	52	0.1197	52	0.1146	62
<i>Priming</i> de la producción a la producción	0.0441	26	0.0653	31	0.0378	26
Posición típica del referente del SN en la cadena de acciones	0.0067	12	0.0255	22	0.0212	14
<i>Priming</i> de la comprensión a la producción	0.00007	15	0.0164	21	0.0044	22
Polaridad de la oración	No signifi- cante	No signifi- cante	No signifi- cante	No signifi- cante	0.0012	14

### 5.3 Factores sociales

La aplicación del Rbrul muestra que la edad, el logro académico y el estilo de habla no condicionan la variación en ninguna de las tres comunidades. En cuanto a la edad, la hipótesis principal ya anticipaba este resultado al describir la variación como un lento cambio lingüístico desde abajo. De la misma manera, la hipótesis 7 ya implicaba que el fenómeno podría ser insensible a la formalidad del encuentro. En cambio, los resultados obtenidos para el nivel de instrucción no corroboran la hipótesis 8. De todas formas, los hallazgos de este estudio corroboran los de otras investigaciones anteriores que han mostrado que la pluralización de *haber* es un rasgo prominente del habla culta de las capitales antillanas (DeMello, 1991; Vaquero, 1978, pp. 135–140). Además, como el efecto de la educación formal suele explicarse en términos de una mayor sensibilidad por parte de los hablantes cultos hacia la formalidad del encuentro y un mejor dominio de los estilos formales (Labov, 1972, p. 138), la ausencia de una dimensión estilística parece explicar por qué el nivel de instrucción no resulta repercutir en la variación.

En cambio, en las tres comunidades de habla, la pluralización de *haber* señala la clase social y el género del hablante. Más específicamente, la Tabla 2 muestra que en La Habana el uso frecuente de la variante pluralizada se asocia a la clase baja, mientras que en Santo Domingo, se conecta esta variante con la clase media. En San Juan, la clase social del hablante sólo alcanza significancia estadística cuando se considera conjuntamente con la edad del informante. O sea, para los hablantes mayores, la pluralización de *haber* señala la clase media, mientras que los hablantes más jóvenes correlacionan el fenómeno con individuos de clase baja. Tocante al género del hablante, la Tabla 2 informa que en La Habana y Santo Domingo los hablantes no usan la pluralización de *haber* para posicionarse sobre el trasfondo de roles genéricos y no se detectan interacciones significantes entre el género y la edad del hablante o cualquiera de las otras variables sociales. En San Juan, por el contrario, el uso frecuente de la variante pluralizada se asocia con el género femenino.

Estos resultados permiten evaluar si la variación constituye un cambio lingüístico desde abajo, como afirma la hipótesis principal, o más bien un cambio desde arriba o una variación estable, como arguyen Díaz-Campos (2003) y Quintanilla-Aguilar (2009). A este respecto, recordemos que el Rbrul indica que el fenómeno carece de una dimensión estilística y que esto es característico de los cambios desde abajo (Labov, 2001, cap. 3; Silva-Corvalán, 2001, pp. 248–249). Asimismo, aunque los resultados para el género y la clase social del hablante no dan cuenta de los patrones establecidos en las hipótesis 5–6, ello no invalida necesariamente la hipótesis de cambio lingüístico. Más bien, Labov (2001, p. 309) indica que durante las últimas etapas de los cambios desde bajo, tienden a desaparecer las diferencias

de género y puede desarrollarse una asociación lineal entre la clase social del hablante y su uso de la variante innovadora (Labov, 2001, p.309). Como éste es el patrón que observamos para las variables sociales, los resultados sugieren que en el Caribe insular, como en otros lugares de Latinoamérica (D' Aquino-Ruiz, 2008; Fontanella de Weinberg, 1992a), la pluralización de *haber* constituye un cambio lingüístico desde abajo avanzado que permite a los hablantes posicionarse en términos de clase social y, por lo menos en San Juan, de género.

## 6. Discusión y conclusiones

En este artículo investigamos la pluralización de *haber* presentacional a la luz de cinco factores sociales (la clase social, la edad, el género y el nivel de instrucción del hablante, así como el estilo de habla) y cinco parámetros lingüísticos (la posición más natural del referente del SN en la cadena de acciones, el tiempo verbal, la polaridad de la oración, el *priming* de la comprensión a la producción y el *priming* de la producción a la producción) en tres muestras recientes de hablantes cubanos, dominicanos y puertorriqueños, residentes en, respectivamente, la Provincia de Ciudad de La Habana, Gran Santo Domingo y el Área Metropolitana de San Juan. En términos generales, podemos afirmar que la variante más frecuente en español caribeño es la forma impersonal, aunque hay que añadir que la mayoría de las ocurrencias de este tipo se centran en la categoría temporal presente de indicativo, que apenas presenta variación. Sin este tiempo verbal, la variante pluralizada es, con mucho, la alternativa imperante en las tres variedades, lo que corrobora los resultados de otros estudios anteriores llevados a cabo en las Antillas (Brown & Rivas, 2012; Rivas & Brown, 2012, 2013) y América Latina (p.ej., Bentivoglio & Sedano, 1989; D'Aquino-Ruiz, 2004, 2008; Díaz-Campos, 2003; Quintanilla-Aguilar, 2009).

Volvamos a las hipótesis para algunas observaciones finales. En lo que se refiere a la primera hipótesis, los datos presentados en este estudio indican que sólo el boricua se deja guiar por la polaridad de la oración (es decir, la definitud/especificidad del SN) a la hora de elegir una de las dos variantes de la construcción presentacional con *haber*. Como Santo Domingo y La Habana también dan cuenta de tasas de pluralización un poco más elevadas, interpretamos estos resultados como indicios de que la variante pluralizada podría haber invadido el territorio conceptual indefinido no específico, lo cual indicaría un cambio en progreso. En cambio, la posición más natural del referente del SN en la cadena de acciones sí mostró ser una restricción importante para todos los hablantes de las tres variedades. Además, tal y como establece la H1, los SSNN cuyos referentes pueden ser imaginados fácilmente como cabezas de una cadena de acciones favorecen la variante pluralizada. Por lo tanto, cabe concluir que esta hipótesis queda corroborada.

Tocante a las hipótesis 2a–c, tanto en La Habana como en Santo Domingo y San Juan, la distribución de las formas temporales de *haber* en un corpus del siglo XVI resultó ser la restricción más potente en la variación. Sin embargo, es difícil explicar estos resultados en el marco de un análisis centrado en diferencias de productividad de la morfología verbal (p.ej., Rivas & Brown, 2013, pp. 113–114; Waltereit & Detges, 2008, p. 27). O sea, sin recurrir a dos construcciones de estructura argumental, construcciones parcialmente llenadas léxicamente y sus fuerzas de disuasión estadística divergentes, es difícil explicar por qué el pretérito desfavorece la variante pluralizada, mientras que todos los demás tiempos verbales otros que el presente de indicativo la propician. De la misma manera, si afirmáramos que el efecto del tiempo verbal se debe al mayor contraste fónico entre el singular y el plural de las formas pretéritas (p.ej., Bentivoglio & Sedano, 2011, p. 174; Hernández-Díaz, 2006, p. 1151) antes que a la disuasión estadística, no pudiéramos explicar por qué las conceptualizaciones aspectual o modalmente complejas –que se codifican mediante perífrasis verbales que dan cuenta del mismo ‘incremento silábico’ (Bentivoglio & Sedano, 2011, p. 174)– estimulan la construcción pluralizada. En este sentido, los resultados corroboran las hipótesis 2a–c, lo cual, a su vez, respalda la hipótesis principal.

Igualmente, la hipótesis 3 queda confirmada, pues hemos mostrado que la variación está sujeta a efectos de *priming*. Estos resultados sugieren que el hablante no repite las formas verbales específicas y las estructuras argumentales estipuladas por las mismas, sino que retoma la construcción de estructura argumental más abstracta. Por lo tanto, estos resultados también abogan a favor de la hipótesis principal a la vez que constituyen un argumento fuerte en contra de estudios que tratan el fenómeno como un reanálisis del SN de la construcción presentacional con *haber* como sujeto (Brown & Rivas, 2012; Rivas & Brown, 2012) o del verbo como un núcleo intransitivo (Montes de Oca, 1994; Waltereit & Detges, 2008).

Jerarquizando los factores lingüísticos según su contribución a la precisión de modelos de bosque aleatorio, hemos visto que, para las tres variedades, no solamente se observan las mismas tendencias al interior de las variables independientes, sino que dichas variables también tienen el mismo peso relativo. Estos resultados tienen algunas implicaciones. Primero, indican que las tres variedades están pasando por el mismo cambio lingüístico, como afirma la hipótesis 9. Segundo, tal y como proponen las hipótesis 1–3 y la hipótesis principal, los hallazgos apoyan que la variación constituye una competencia entre dos variantes de la construcción presentacional con *haber* que se ve constreñida por los mismos tres factores cognitivos, a saber, la preferencia por la codificación no marcada (véase el efecto de la polaridad de la oración y la posición más natural del referente del SN en la cadena de acciones), la disuasión estadística (véase el efecto del tiempo verbal) y el *priming* estructural. Asimismo, contraponiendo la jerarquización obtenida



mediante los modelos de bosque aleatorio con el orden establecido en base a los rangos probabilísticos, hemos mostrado que el primer método produce resultados más confiables, lo cual constituye un hallazgo metodológico importante para futuros estudios en el campo de la sociolingüística comparada.

En lo que se refiere a las variables sociales, mostramos que la variación se correlaciona con la clase social del hablante, aunque en San Juan esta última variable sólo alcanza significancia estadística al considerarla junto con la edad. En esa ciudad, la variación también señala el género del hablante. Aunque los resultados obtenidos para estos dos factores sociales no corroboran las hipótesis 5–6, siguiendo a Labov (2001, p. 309), los interpretamos como indicios de que, en el Caribe insular, la pluralización de *haber* constituye un cambio avanzado desde abajo.

En conclusión, los datos aducidos en este artículo confirman la hipótesis principal: en las capitales de las Antillas Mayores, la pluralización de *haber* presentacional constituye un cambio lingüístico avanzado durante el cual la variante pluralizada (<SAdv *haber* Sujeto>) está reemplazando el formato impersonal (<SAdv *haber* Objeto>). Asimismo, esta investigación revela que la variación se ve condicionada por los mismos tres factores cognitivos que restringen cualquier expresión lingüística, a saber, la preferencia generalizada por la codificación no marcada, el *priming* estructural y la disuasión estadística. Además, hemos visto que las alternancias le permiten al hablante ubicarse socialmente. Finalmente, cabe destacar que el enfoque comparado contribuye a la solidez de nuestras conclusiones y que no hubiera sido posible llegar a las mismas sin el apoyo de un marco teórico que nos permite modelar la variación y los conocimientos de los hablantes acerca de su distribución social como facetas centrales de la competencia lingüística.

## Referencias

- Alba, O. (2004). *¿Cómo hablamos los dominicanos?: Un enfoque sociolingüístico*. Santo Domingo: Grupo León Jimenes.
- Baayen, H. R. (2014). Multivariate statistics. In R. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 337–372). Cambridge: Cambridge University Press.
- Bentivoglio, P., & Sedano, M. (1989). *Haber: ¿Un verbo impersonal? Un estudio sobre el español de Caracas*. En G. de Granda (Ed.), *Estudios sobre el español de América y lingüística afroamericana. Ponencias presentadas en el 45 Congreso internacional de americanistas* (pp. 59–81). Bogotá: Instituto Caro y Cuervo.
- Bentivoglio, P., & Sedano, M. (2011). Morphosyntactic variation in Spanish-speaking Latin America. En M. Díaz-Campos (Ed.), *The handbook of Hispanic sociolinguistics* (pp. 123–147). Oxford: Blackwell.
- Blas-Arroyo, J. L. (1999). *Lenguas en contacto: Consecuencias lingüísticas del bilingüismo social de las comunidades de habla del este peninsular*. Frankfurt am Main/Madrid: Vervuert/Iberoamericana.



- Brown, E., & Rivas, J. (2012). Grammatical relation probability: how usage patterns shape analogy. *Language Variation and Change*, 24(3), 317–341. DOI: 10.1017/S0954394512000154
- Bybee, J. (2003). Cognitive processes in grammaticalization. En M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language* (Vol. 2) (pp. 145–167). Londres: Laurence Erlbaum.
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711–733. DOI: 10.1353/lan.2006.0186
- Claes, J. (2014). A cognitive construction grammar approach to the pluralization of presentational *haber* and its social significance in Puerto Rican Spanish. *Language Variation and Change*, 26(2), 219–246. DOI:10.1017/S095439459451400052.
- Company-Company, C. (2003). La gramaticalización en la historia del español. *Medievalia*, 35, 3–61.
- Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology* (2nd edition). Chicago: University of Chicago Press.
- Croft, W. (2003). *Typology and universals*. Cambridge: Cambridge University Press.
- D'Aquino Ruiz, G. (2004). *Haber* impersonal en el habla de Caracas. *Análisis sociolingüístico. Boletín de lingüística*, 21, 3–26.
- D'Aquino Ruiz, G. (2008). El cambio lingüístico de *haber* impersonal. *Núcleo*, 20(25), 103–124.
- De Granda, G. (1994). *Español de América, español de África y hablas criollas hispánicas: Cambios, contactos y contextos*. Madrid: Gredos.
- DeMello, G. (1991). Pluralización del verbo *haber* impersonal en el español hablado culto de once ciudades. *Thesaurus*, 46, 445–471.
- Díaz-Campos, M. (2003). The pluralization of *haber* in Venezuelan Spanish: A sociolinguistic change in real time. *IU Working Papers in Linguistics*, 3(5), 1–13.
- Domínguez-Hernández, M. A. (2007). Palabras preliminares. En M. A. Domínguez-Hernández (Ed.), *La lengua en Cuba: Estudios* (pp. 7–26). Santiago de Compostela: Universidade de Santiago de Compostela.
- Fontanella de Weinberg, M. B. (1992a). Variación sincrónica y diacrónica de de las construcciones con *haber* en el español americano. *Boletín de filología de la universidad de Chile*, 33, 35–46.
- Fontanella de Weinberg, M. B. (1992b). *El español de América*. Madrid: Mapfre.
- Freites-Barros, F. (2008). Más sobre la pluralización de *haber* impersonal impersonal en Venezuela: El estado Táchira. *Lingua americana*, 12(22), 36–57.
- Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.
- Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, A. (2009). The nature of generalization in language. *Cognitive Linguistics*, 20(1), 93–127.
- Goldberg, A. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, 22(1), 131–153. DOI: 10.1515/cogl.2011.006
- Hernández-Díaz, A. (2003). La semántica locativa de las construcciones existenciales con *haber*: Un ejemplo de gramaticalización. *Medievalia*, 35, 135–154.
- Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 3(1), 359–383. DOI: 10.1111/j.1749-818X.2008.00108.x
- Kany, C. E. (1951). *Sintaxis hispanoamericana*. Madrid: Gredos.

- Keenan, E. L. (1976). Towards a universal definition of 'subject'. En C. N. Li (Ed.), *Subject and topic* (pp. 303–333). Nueva York: Academic Press.
- Labov, W. (1972). *Sociolinguistic patterns*. Filadelfia: University of Pennsylvania Press.
- Labov, W. (1994). *Principles of linguistic change. Volume 1: Internal factors*. Oxford: Blackwell.
- Labov, W. (2001). *Principles of linguistic change. Volume 2: Social factors*. Oxford: Blackwell.
- Labov, W. (2010). *Principles of linguistic change. Volume 3: Cognitive and cultural factors*. Oxford: Blackwell. DOI: 10.1002/9781444327496
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press. DOI: 10.7208/chicago/9780226471013.001.0001
- Langacker, R. W. (1987). *Foundations of cognitive grammar. Volume 1: Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, R. W. (1991). *Foundations of cognitive grammar. Volume 2: Descriptive application*. Stanford: Stanford University Press.
- Lavandera, B. (1984). *Variación y significado*. Buenos Aires: Hachette.
- Lyons, J. (1967). A note on possessive, existential and locative sentences. *Foundations of language*, 3, 390–396.
- Montes de Oca, M. (1994). La concordancia con *haber* impersonal. *Anuario de Letras*, 32, 7–35.
- Moreno-Fernández, F. (2003). *Metodología del proyecto para el estudio sociolingüístico del español de España y América (PRESEEA)*. Recuperado el 9 de abril del 2008 de: [www.linguas.net/portalpresea](http://www.linguas.net/portalpresea).
- Navarro-Correa, M. (1992). Valoración social de algunas formas verbales en el habla de Valencia. *Lingüística española actual*, 14(1), 97–106.
- Pérez-Martín, A. M. (2007). Pluralización de *había* en el habla de El Hierro: datos cuantitativos. *Revista de Filología de la Universidad de La Laguna*, 25, 505–513.
- Prince, E. F. (1992). The ZPG letter: Subjects, definiteness, and information-status. En W. C. Mann & S. A. Thompson (Eds.), *Discourse description: Diverse analyses of a fund-raising text* (pp. 295–325). Amsterdam/Filadelfia: John Benjamins. DOI: 10.1075/pbns.16.12pri
- Quintanilla-Aguilar, J. R. (2009). *La (des)pluralización del verbo 'haber' existencial en el español salvadoreño: ¿Un cambio en progreso?*. Tesis doctoral. Miami: University of Florida.
- Real Academia Española. (2008-). *Corpus diacrónico del español (CORDE)*. Recuperado el 10 de junio del 2010 de: <http://corpus.rae.es/cordenet.html>.
- Real Academia Española, & Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- Rivas, J., & Brown, E. (2012). Stage-level and individual-level distinction in morphological variation: An example with variable *haber* agreement. *Borealis: An International Journal of Hispanic Linguistics*, 1(2), 73–90. DOI: 10.7557/1.1.2.2340
- Rivas, J., & Brown, E. (2013). Concordancia variable con *haber* en español puertorriqueño. *Boletín de lingüística*, 24(37-38), 102–118.
- Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español*. Washington: Georgetown University Press.
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23, 193–229. DOI: 10.1016/S0271-5309(03)00013-2
- Tagliamonte, S. (2002). Comparative sociolinguistics. En J. K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 729–763). Oxford: Blackwell.
- Tagliamonte, S. (2006). *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511801624

- Tagliamonte, S., & Baayen, H. R. (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. DOI: 10.1017/S0954394512000129
- Vaquero, M. (1978). Enseñar español, pero ¿qué español? *Boletín de la academia puertorriqueña de la lengua española*, 6, 127–146.
- Vaquero, M. (1996). Antillas. En M. Alvar-López (Ed.), *Manual de dialectología hispánica: El español de América* (pp. 51–67). Barcelona: Ariel.
- Waltereit, R., & Detges, U. (2008). Syntactic change from within and from without syntax: A usage-based analysis. En U. Detges & R. Waltereit (Eds.), *The paradox of grammatical change: Perspectives from Romance* (pp. 13 – 30). Amsterdam: John Benjamins. DOI: 10.1075/cilt.293.02wal
- Ward, G., & Birner, B. (1995). Definiteness and the English existential. *Language*, 71(4), 722–742. DOI: 10.2307/415742
- Wolfram, W., & Schilling-Estes, N. (2003). Dialectology and linguistic diffusion. En B. D. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 713–735). Oxford: Blackwell. DOI: 10.1002/9780470756393.ch24

### Apéndice A. Distribución de las formas terciopersonales de *haber* en documentos americanos (1492–1600) (RAE, 2008–)

	<i>Había</i>		<i>Hubo</i>		<i>Habría</i>		<i>Habrá</i>		<i>Haya</i>		<i>Hubiera</i>		<i>Hay</i>		<i>Ha habido</i>		<i>Total</i>	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Verbo auxiliar	1806	52.4	38	6.5	6	10.5	41	13.4	202	32.3	110	50	0	0	0	0	2203	25.2
<i>Haber de</i> + infinitivo	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
<i>Haber que</i> + infinitivo	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
<i>Haber</i> ‘tener’	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
Construcción presentacional impersonal	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
<i>Ha</i> + tiempo	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

#### *Dirección del autor*

Jeroen Claes  
Bresstraat 14, apt. 2  
2018 Amberes  
Bélgica

Jeroen Claes@gmail.com

*Sobre el autor*

**Jeroen Claes** es estudiante de doctorado en la Universiteit Antwerpen (Amberes, Bélgica). Posee una maestría en español de la misma universidad, con especialización en el Caribe, México y España. Sus principales intereses son la sociolingüística de los dialectos caribeños y la manera en que fenómenos variables y los conocimientos de los hablantes acerca de sus distribuciones sociales pueden modelarse en teorías cognoscitivas del lenguaje. Actualmente, trabaja estos aspectos en su tesis doctoral, que se centra en la pluralización de *haber* presentacional en el español de las capitales antillanas desde las perspectivas combinadas de la sociolingüística comparada y la gramática de construcciones. Ha publicado en revistas de prestigio internacional, tales como *Cognitive Linguistics*, *Language Variation and Change*, *Revista Internacional de Lingüística Iberoamericana* y *Spanish in Context*.