

La subjetivización del *de que* en el español de Colombia

The subjectification of *de que* in Colombian Spanish

Matías Guzmán Naranjo

Westfälische Wilhelms-Universität Münster

m_guzm01@uni-muenster.de

Resumen

En este trabajo analizo el fenómeno conocido como dequeísmo en español, y en particular las diferencias semánticas entre las oraciones canónicas con *que* y las oraciones dequeístas. Reviso estudios de corpus previos sobre dequeísmo, e intento confirmar las predicciones de estos con un corpus independiente de español oral colombiano. Finalmente, con una regresión logística pruebo nuevos posibles parámetros que podrían influenciar la elección de una u otra construcción. El resultado es que las conclusiones y predicciones de la mayoría de análisis previos no son replicables con el corpus usado en este estudio, y que el dequeísmo parece estar influenciado por el contexto discursivo (situaciones reales o no reales), y la subjetividad expresada en la oración.

Palabras clave

Lingüística de corpus, dequeísmo, subjetivización

Abstract

This paper deals with the phenomenon known as *dequeísmo* in Spanish, in particular with the semantic differences between canonical sentences with *que* and dequeísta sentences. I analyze previous corpus studies of dequeísmo, test their predictions with an independent corpus of spoken Colombian Spanish, and finally carry out a logistic regression to test new possible parameters that might influence speaker's choice. The result is that most previous accounts of dequeísmo are not consistent with the corpus used for this study, and that dequeísmo seems to be influenced by the discourse context (real or non real situations), and speaker's subjectivity.

Keywords

Corpus linguistics, dequeísmo, subjectification

1 Introducción

El fenómeno conocido como dequeísmo en español consiste en la alternación entre las formas *que* y *de que* para introducir oraciones subordinadas, oraciones relativas, y en algunos

casos para introducir oraciones principales como respuesta a una pregunta. El dequeísmo, a pesar de ser considerado como un error gramatical por la mayoría de prescriptivistas (Avila, 2003; Silva Villar and Gutiérrez Rexach, 2012), está extremadamente extendido por el mundo hispanohablante y puede encontrarse en la mayoría de dialectos de América y España, como lo evidencian la multitud de estudios realizados sobre el tema en diferentes regiones (Battini, 1949; Rabanales, 1974; Quilis Sanz, 1986; Prieto, 1995; Serrano, 1998; Schwenter, 1999; Del Moral, 2008, y referencias mencionadas en estos).

En estudios recientes algunos autores (Schwenter, 1999; Cornillie y Delbecque, 2008; Del Moral, 2008) han propuesto que el dequeísmo debe analizarse como un producto de un proceso de gramaticalización y subjetivización. Según estos estudios, el uso de *de que* no es producto de factores pragmáticos como cuando los hablantes intentan sonar importantes, o cuando intentan evitar errores (ver Bentivoglio, 1975; Suárez, 2009, para algunos ejemplos de esta perspectiva pragmática), sino que es un elemento con una función gramatical específica. De acuerdo con esta posición, *de que* no es una forma sinónima con *que*, y expresa diferencias semánticas o semántico-pragmáticas. No obstante, no ha sido posible hasta el momento establecer la naturaleza exacta de esta diferencia, y hay opiniones encontradas al respecto. García (1986) y Martínez-Sequeira (2000) ven el dequeísmo como un mecanismo que usa el hablante para crear distancia icónica entre manifestación gramatical del sujeto y lo dicho, y así distanciarse él mismo de la afirmación. Schwenter (1999) analiza el dequeísmo como un marcador evidencial que expresa información de segunda mano. Del Moral (2008) por su parte, toma una posición diferente y propone que *de que* no es un marcador evidencial ni una forma de distanciamiento, sino una forma de expresar información más subjetiva para el hablante, o que el hablante intenta mostrar que el sujeto de la oración tiene una posición subjetiva

frente a la información o la acción que se le atribuye. La hipótesis de que *de que* es un elemento gramaticalizado (de ahora en adelante H1) es supremamente interesante porque si es acertada, es posible explicar con ella por qué los hablantes son capaces de alternar entre las dos formas en un mismo texto y en una misma oración, y por qué *de que*, aun siendo más larga, es usada junto con *que*.

A pesar de que esta explicación para el desarrollo del dequeísmo funciona en la teoría, solo ha habido unos cuantos estudios empíricos (García, 1986; Martínez-Sequeira, 2000; Del Moral, 2008) para determinar si las oraciones con *de que* sí llevan alguno de las cargas semánticas mencionadas en el párrafo anterior, y entre estos estudios, Suárez (2009) no encontró resultados positivos que apoyaran H1. En el presente trabajo intentaré evaluar algunas de las diferentes propuestas de análisis del dequeísmo con un experimento en forma de encuesta y un análisis cuantitativo de un corpus oral para de español colombiano.

La sección 1. explica de forma más detallada los aspectos sintácticos del dequeísmo, y los tipos de construcciones en los que se encuentra. En la sección 2. reviso algunos de los estudios de corpus más relevantes realizados hasta la fecha sobre el dequeísmo. La sección 3. explica la metodología para ambos experimentos; y en la sección 4. presento los resultados obtenidos. En la sección 5. evalúo los resultados de ambos experimentos y las implicaciones de estos. Finalmente, la sección 6. presenta algunas consideraciones finales e implicaciones del presente estudio, así como algunas recomendaciones para futuros experimentos sobre el dequeísmo.

2 Aspectos sintácticos del dequeísmo

La primera fuente que hace uso de la palabra dequeísmo en Google Books es Estudios filológicos y lingüísticos publicado en 1974 (Rosenblat, 1974); y una de las primeras fuentes que menciona el fenómeno (sin darle el nombre de dequeísmo) es Bantini (1949). Otros trabajos pioneros sobre el dequeísmo fueron llevados a cabo por Bentivoglio (1975; 1980), quien realizó estudios sobre este fenómeno en Venezuela y Chile¹. No obstante, las oraciones dequeístas son bastante antiguas en la historia del español, en parte evidenciado por su

¹ Bentivoglio tomó un acercamiento un poco diferente al dequeísmo, ella propone que el uso de oraciones dequeístas se da porque los hablantes lo ven como un uso de las clases altas, no por razones funcionales. Si bien es posible que este factor juegue un papel en el uso del *de que*, este artículo no explora esta posibilidad.

distribución geográfica, y en parte por documentos históricos que demuestran su uso desde el siglo XIV (para algunos ejemplos ver Del Moral, 2008). A pesar de esto, el dequeísmo suele ser asociado en varios países con dialectos de las clases sociales más bajas (Rabanales, 1974; Prieto, 1995; Serrano, 1998).

El dequeísmo no es un cambio en la forma en la que los verbos codifican los argumentos, de objeto directo a objeto preposicional, *de que* es una conjunción que sustituye a *que*. Los siguientes ejemplos (y todos los demás ejemplos presentados en este artículo) fueron tomados del corpus PRESEEA Medellín (Rátiva, 2007)². El corpus PRESEEA es un corpus del español oral de Medellín, Colombia; cuenta con alrededor de 800000 palabras, y fue recogido con entrevistas semidirigidas. Los ejemplos en (1) muestran que los hablantes en la segunda ocurrencia de *ser* usan como una cópula en la forma canónica sin una oración subordinada, pero introduce la subordinada anterior con *de que*:

- (1) la juventud no es **de que** tengo la piel lisa, la juventud es lo **que** uno respire.

Este primer ejemplo muestra que los hablantes no han cambiado el paradigma verbal, y que no necesariamente hacen uso de *de que*. Esto indica que *de que* no es analizado como una conjunción más una preposición, sino como un solo elemento.

Un ejemplo similar en (2) muestra aún mejor el hecho de que el dequeísmo es una elección entre dos construcciones diferentes, y que *de que* no ha remplazado realmente a *que*:

- (2) de todas maneras ellos no compartían, **de que** después de cuatro años, y mi forma de ser mía y todo eso, no compartían que de pronto las cosas ya no estuvieran funcionando como, como debían funcionar.

Como podemos ver en (2), el hablante puede usar *que* y *de que* en la misma construcción con el mismo verbo (*compartir (de) que...*), en el mismo contexto y en el mismo hilo discursivo. Esta característica del dequeísmo es extremadamente difícil de explicar, y constituye uno de los mayores desafíos al elaborar una teoría sobre este fenómeno. En este caso no parece haber ninguna diferencia

² Este corpus es de acceso libre y todos los datos y experimentos presentados en este artículo pueden ser replicados:
<http://comunicaciones.udea.edu.co/corpuslinguistico/>

semántica entre ambas oraciones, la actitud del hablante no parece haber cambiado de la primera a la segunda oración, y aun así el hablante decide alternar entre ambas. Casos como estos hacen que una teoría puramente semántica del dequeísmo sea poco probable.

De la misma manera, un hablante dequeísta puede elegir no alternar en una serie de oraciones y mantenerse consistente con una de las dos formas:

- (3) a. [el] problema del narcotráfico, el problema de las de los paras, el problema de la guerrilla, los enfrentamientos las pandillas, todas esas cosas, todo eso ha hecho **de que** realmente se acentúe más la división que hay en la ciudad [...] y eso ha hecho **de que** realmente pues se pierda como ese calor que uno tenía o sentía antes de una ciudad muy acogedora.
b. Para que la gente no diga pues **de que** es el beneficio **de que** es la familia.

Ambos ejemplos son interesantes. En (3a) el hablante usa en ambos casos *de que* para el mismo verbo pero en oraciones separadas, y en (3b) el hablante usa dos veces *de que* como cabezas de dos subordinadas de la misma oración principal.

En (4) vemos otro ejemplo de alternación entre *que* y *de que* por el mismo hablante, en este caso con verbos diferentes:

- (4) yo haría algo pero por los niños, por los de escasos recursos, por ellos sí haría, porque considero **de que**, espacio para mucha gente hay, me parece **que** falta mucho espacio para los niños, de escasos recursos.

Otro contexto sintáctico en el que *de que* puede substituir a *que* es cuando actúa como un pronombre relativo. Este caso ha sido poco discutido en la literatura, y parece ser que solo Del Moral (2008) lo menciona. No obstante, es especialmente relevante cuando se considera que *de que* puede ser un caso de gramaticalización. En (5) a-d se presentan ejemplos del uso pronominal de *de que*:

- (5) a. El metro [...] es **algo de** que va directo, para en la estación
b. me imaginaba de que sí yo seguía soltero iba a ser **una persona de que** iba a estar muy sola en mi casa.
c. Yo no cambio el Doce de Octubre por Bello, por Envigado, El Poblado, no, o sea,

es **un barrio de que** ha venido de menos a más.

d. por decir yo tengo **un amigo de que** tiene un niño.

Del Moral (2008) encontró construcciones similares en su corpus. Estos casos sugieren que *de que* está fuertemente asociado a *que* en el sistema de los hablantes dequeístas, y estos han extendido su uso a otras construcciones en las que *que* se usaría. Si este análisis es correcto, estos ejemplos representan fuerte evidencia en favor de H1.

Algunas construcciones con *que* aún no pueden substituirse por *de que* (al menos en el corpus no se encontraron ejemplos de estos tipos):

Oraciones comparativas:

- (6) a. más grande que él.
b. * más grande de que él.

como un pronombre relativo después de un artículo:

- (7) a. el que, la que, lo que.
b. * el/la/lo de que.

como un pronombre interrogativo:

- (8) a. ¿qué es eso?.
b. ¿de qué es eso?

En este caso, aunque el ejemplo (8b) es gramatical, el *de que* no es interpretado como un solo elemento, sino como dos, y con el significado de la preposición *de* aún presente.

Con estos ejemplos podemos ver que el dequeísmo es un fenómeno supremamente complejo, y que no es claro a simple vista por qué los hablantes a veces escogen *que* y a veces *de que*. Si en realidad hay una diferencia semántica entre ambas no es posible encontrarla mirando tan solo algunos ejemplos, es necesario estudiar muchos casos en conjunto.

3 Aspectos semántico-pragmáticos sobre el dequeísmo

En esta sección presentaré algunos estudios que han tratado de identificar las propiedades semántico-pragmáticas del dequeísmo, y los dividiré en tres hipótesis que intentan explicar el fenómeno.

García (1986) estudió el dequeísmo en Santiago, Caracas y Buenos Aires. Ella propuso que *de que* crea distancia entre el hablante y la oración

(Hipótesis de Distancia e Incertidumbre). Esta hipótesis también significa que cuando un hablante usa *de que*, éste se encuentra menos seguro de la afirmación. Para verificar su hipótesis, García realizó un estudio de corpus que encontró que *de que* se usa con mayor frecuencia con verbos en tercera persona y en formas impersonales que en primera persona, y que se usa más con conjugaciones no presentes. Estos datos son consistentes con su teoría porque indican que *de que* se usa cuando el hablante no está demasiado involucrado con la afirmación (cuando la oración no es en primera persona), o cuando ésta se encuentra más distante y es más incierta (en pasado o futuro).

De forma similar, Schwenter (1999) propuso que *de que* es un marcador evidencial (Hipótesis de Evidencialidad). Los datos del estudio de Schwenter también mostraron preferencia por la tercera persona con las oraciones dequeístas, lo cuál es interpretado como confirmación de la Hipótesis de evidencialidad.

Kanwit (2012), en un estudio aún no publicado, intentó corroborar las afirmaciones de Schwenter con el mismo corpus usado por aquel. Además de considerar las variables Tiempo y Persona, Kanwit también consideró el contexto discursivo. Para evaluar el contexto Kanwit propone tres niveles: cargado emocionalmente (temas posiblemente ofensivos), importante (temas de importancia para el hablante), y neutral. En su estudio encontró que *de que* se usa con más frecuencia en contextos cargados emocionalmente que en los demás contextos. Kanwit interpreta estos resultados como evidencia de que *de que* funciona como un marcador evidencial al hacer que la afirmación sea más débil, pero también podrían entenderse desde la Hipótesis de Distancia e Incertidumbre.

Del Moral (2008) realizó un estudio con un corpus histórico en el que analizó cómo *de que* ha evolucionado en los últimos siete siglos. Del Moral propone que el dequeísmo es un producto de subjetivización (Hipótesis de Subjetivización), y para verificar esta hipótesis comparó el número de oraciones dequeístas en primera persona y no primera persona en cada siglo. Del Moral encontró que la proporción de oraciones dequeístas en primera persona no solo es mayor que en no primera persona, sino que la proporción de éstas ha ido incrementando constantemente desde el siglo XIII. Del Moral también encontró que los verbos sicológicos eran más frecuentes con oraciones dequeístas, y concluye que esto es evidencia a favor de su teoría. La propuesta de Del Moral se enmarca en la teoría de gramaticalización (Bybee, 2003;

Bybee, Perkins, Pagliuca, 1994; Company C., 2004; Haspelmath, 2002, 2004 and 2006). Dentro de esta teoría, uno o varios elementos léxicos incrementan su frecuencia, sufren erosión fonética y semántica, y empiezan a actuar como elementos gramaticales. En este caso *de* ha perdido su componente semántico y forma una sola palabra fonológica en *de que*, y actúa como una conjunción independiente.

A pesar de ser una propuesta con una sólida base teórica, el estudio de Del Moral tiene algunos problemas. El más evidente es que el corpus usado, Corpus del Español (Davis, 2002), no es consistente en el tipo de textos que recoge entre siglos. Los documentos más antiguos no son comparables ni en tipo ni en extensión a los más modernos, por lo que cualquier generalización resulta algo problemática. En segundo lugar, Del Moral no explica qué contó exactamente, ni qué algoritmo (o búsqueda) usó para extraer las oraciones dequeístas. Tampoco explica cómo distinguió entre el *de que* canónico y el dequeísta.

Es importante dejar en claro que la Hipótesis de Subjetivización es totalmente opuesta a la Hipótesis de Distancia e Incertidumbre. En la primera *de que* indica que el hablante está mucho más involucrado con la afirmación que en las oraciones con *que*, mientras que la segunda se marca lo contrario, que el hablante se quiere distanciar de lo dicho.

Finalmente, un estudio por Suárez (2009) afirma que no hay evidencia positiva para ninguna de estas tres hipótesis (Suárez las llama Hipótesis Funcionalistas, p. 164), es decir, H1. Suárez no encontró evidencia de que haya diferencias significativas en el uso de presente-no presente, ni entre primera persona-no primera persona en las oraciones dequeístas. Los resultados de su estudio se presentan a continuación:

Primera persona			No primera persona		
N	T	%	N	T	%
20	934	2,10%	13	875	1,50%
Presente			No presente		
N	T	%	N	T	%
30	1584	1.9%	3	221	1,40%

Tabla I: Datos para Persona y Tiempo en oraciones dequeístas (Suárez, 2009, p. 164)

Si se comparan de esta forma, Suárez tendría razón y las diferencias entre las proporciones no serían significativas ($p < 0.001$). No obstante, no es claro cuáles son los totales reales en este caso.

Suárez no explica si contó el número total de oraciones, o el número total de oraciones con *de que* producidas por todos los hablantes, o solo las producidas por los hablantes dequeístas. La forma apropiada habría sido solo contar las oraciones de los hablantes dequeístas, pero sin los datos del estudio no es posible verificar cuál es el caso.

En resumen hay tres hipótesis sobre la función y significado de *de que*: La Hipótesis de Distancia e Incertidumbre, la Hipótesis de Evidencialidad y la Hipótesis de Subjetivización. Estudios diferentes con corpus diferentes han producido resultados que apoyan o una u otra hipótesis, y en general son inconsistentes los unos con los otros. Además, tanto la Hipótesis de Distancia e Incertidumbre como la Hipótesis de Evidencialidad hacen predicciones iguales sobre la distribución de los datos, lo que dificulta mucho poder distinguirlas empíricamente. En lo que sigue propongo una forma tanto diferente para evaluar la validez de estas tres hipótesis.

4 Metodología

En esta sección presento los métodos usados para verificar las tres hipótesis descritas en la sección anterior, y doy una descripción simple de las técnicas estadísticas empleadas.

4.1 Encuesta

Para probar la Hipótesis de Evidencialidad se realizó una encuesta con hablantes de español de Colombia. En la encuesta a los participantes se les presentaron 20 oraciones en las que una supuesta persona estaba diciendo algo sobre alguien más, y ellos debían marcar de 1 a 5 (1 muy inseguro, y 5 muy seguro) qué tan seguro estaba la supuesta persona de información expresada en la oración. La mitad de las oraciones eran distractores, de las 10 oraciones restantes, 5 eran oraciones canónicas con *que* y 5 eran oraciones con *de que*; todos los verbos en todas las oraciones eran diferentes. Para evitar cualquier efecto producido por la elección particular de las oraciones, la muestra de hablantes se dividió en dos y a la mitad se le asignó la encuesta A, y a la otra mitad la encuesta B. Las oraciones dequeístas en A aparecían como canónicas en B, y las oraciones canónicas en A aparecían como dequeístas en B; los distractores no se alteraron entre A y B. Además se recogieron los datos Edad, Sexo, y Estrato (o Clase Social). La mayoría de los participantes que respondió la encuesta resultó estar entre los rangos de edad 20-35 años, por lo que esta variable no se consideró en el análisis estadístico.

Después de que los informantes concluyeron la encuesta, se les preguntó de forma informal y sin que tuvieran que anotar la respuesta, si ellos consideraban que las oraciones dequeístas expresaban duda sobre lo dicho. Se escogió hacer esta parte de forma informal porque en las pruebas los hablantes comunicaron muy poco cuando se les pidió que escribieran su respuesta, pero intentaron dar explicaciones más detalladas cuando se les pidió que simplemente dieran su opinión de forma oral. Esta decisión hace que los resultados de la pregunta complementaria no puedan ser analizados estadísticamente, y solo presentaré las impresiones generales de los hablantes.

Para analizar los datos comparé si en general, las oraciones dequeístas habían sido marcadas como más inciertas (tendiendo a 1) que las oraciones canónicas. También comprobé si había diferencia entre las clases sociales, y entre hombres y mujeres. La estadística escogida fue el test Chi cuadrado para proporciones.

4.2 Estudio de corpus

Para probar la Hipótesis de Distancia e Incertidumbre y la Hipótesis de Subjetivización se condujo un estudio de corpus con el corpus PRESEEA. Para procesar el corpus se utilizó la librería FreeLing con su API de Java (Padró y Stanilovsky, 2012).

Se extrajeron todas las oraciones dequeístas (56 en total), y se etiquetaron manualmente para las variables Sexo (del hablante), Tiempo (presente-no presente), Modo (indicativo, no indicativo), Número (plural, singular), Polaridad (si la oración era positiva o negativa), Pronominal (uso de pronombre personal o pro-drop), Número de Palabras (número de palabras que aparecen después de *de que* en la oración), Contexto (condicional o real), Tipo de Verbo (sicológico, percepción, decir, u otros marcados como 'do'), Persona (primera persona, no primera persona). Para la variable Contexto, las oraciones que expresaban hechos reales o de los que el hablante parecía estar seguro se marcaron como 'real', mientras que las oraciones que hacía relación a situaciones o eventos posibles, de los que el hablante no estaba seguro se marcaron como 'condicional'. La razón para esta variable es que las afirmaciones sobre eventos condicionales son más subjetivas que las oraciones sobre situaciones reales porque requieren que el hablante tome posición frente a ellas. La motivación para la variable PRONOMINAL es similar, asumo que si se hace uso explícito del pronombre, el hablante está haciendo énfasis en la persona que está haciendo la

afirmación y es por tanto más subjetiva³. Adicionalmente, 57 oraciones con uso de *de que* en forma canónica se extrajeron aleatoriamente del corpus de los textos de los hablantes dequeístas (los hablantes no dequeístas se ignoraron para esta parte del estudio), y se etiquetaron bajo los mismo criterios. Esto permite tener un punto de comparación sólido, algo que la mayoría de estudios mencionados en la sección anterior no presentan.

Primero que todo, se evalúan las propuestas de que ciertas categorías gramaticales como tiempo, modo, o persona ocurren con mayor o menor frecuencia en las oraciones dequeístas. Para esto se emplea el test Chi-cuadrado para proporciones.

Para evaluar la Hipótesis de Subjetivización se construyó un modelo con el método de regresión logística para variables binarias (ver Sheather (2009) para la explicación matemática de esta técnica). En una regresión logística la variable dependiente es categorial y binaria, a un valor se le asigna 1 y al otro valor se le asigna 0. El modelo intenta predecir a partir de las variables independientes la probabilidad de que la variable dependiente sea 1 o 0. Con este método es posible evaluar el efecto total de cada variable, y asignar un valor de qué tan bien una variable puede predecir la variable dependiente. Los resultados se resumen en una tabla de coeficientes. Finalmente, para ver qué tan bien logra predecir el Modelo datos nuevos se realiza una validación bootstrap que selecciona aleatoriamente un subconjunto de los datos, crea el modelo con el resto de los datos, y luego evalúa el nuevo modelo contra los datos no considerados; este proceso se repite 200 veces. Los resultados se presentan en una tabla de coeficientes y en una gráfica.

El propósito de este análisis es determinar qué predictores (variables independientes) muestran una fuerte correlación con la variable dependiente, en este caso si la oración es dequeísta o no. La ventaja de este método sobre otros es que permite analizar varias variables al mismo tiempo y las posibles interacciones entre estas.

5 Resultados

5.1 Resultados de la encuesta

Un total de 118 hablantes respondió la encuesta, 60 la encuesta A y 58 la encuesta B. La mayoría de

los hablantes, en especial los que se encontraban entre las clases sociales 3, 4 y 5, afirmaron que todas las oraciones eran más o menos iguales, pero que las oraciones dequeístas eran un poco raras. Para la mayoría, la elección de un nivel de certeza dependió del verbo usado y no de si la oración era o no dequeísta. Cuando se les preguntó después de que habían contestado la encuesta, si las oraciones con *de que* expresaban duda, todos contestaron negativamente. Como se mencionó en la sección anterior, esta respuesta no se anotó, por lo que no es de relevancia estadística, pero sí hace dudar que la Hipótesis de Evidencialidad sea acertada.

Una primera mirada a los resultados parece indicar que la encuesta no apoya la hipótesis propuesta, y que no hay diferencia en seguridad entre las oraciones canónicas y las dequeístas. Si los datos se analizan como un todo, tenemos dos distribuciones idénticas para *que* y *de que*:

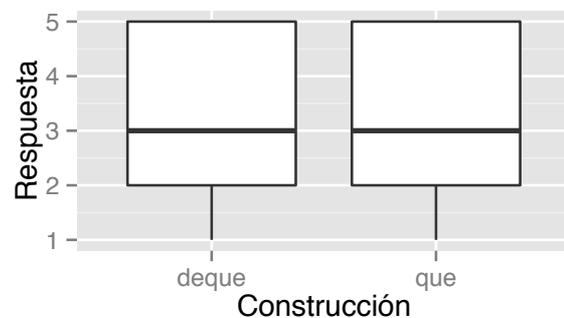


Figura 1: Diagrama de cajas de la distribución de respuestas para ambos cuestionarios.

Como podemos observar en la Figura 1., no parece haber diferencia en la distribución de respuestas para las oraciones canónicas y dequeístas. Ambas distribuciones tienen una media⁴ de 3, no hay valores extremos, y ambas distribuciones tienden hacia el límite superior, lo cual quiere decir que la mayoría de las respuestas estuvieron más cerca de 5 que de 1. Un histograma representando las distribuciones por respuesta se presenta en la Figura 2:

3 Hay que notar que autores como Company C. (2004) ven esta variable al revés, y afirman que las oraciones más subjetivas no tienen un sujeto agentivo. No habiendo evidencia empírica de que este es el caso, rechazo la propuesta de Company C.

4 Como los datos son ordinales, el promedio no puede usarse como una medida de tendencia central. La diferencia entre 1 y 2 no es necesariamente la misma que la diferencia entre 3 y 4.

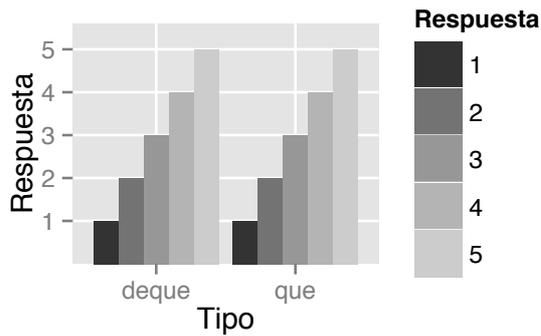


Figura 2: Histograma de la distribución de respuestas para ambos cuestionarios.

De la Figura 2. podemos concluir que no parece haber diferencias entre ambas construcciones. Si los datos se dividen según los diferentes factores considerados en el estudio las proporciones cambian ligeramente. En la Figura 3. podemos ver la frecuencia de que cada respuesta por construcción según el estrato de los hablantes.

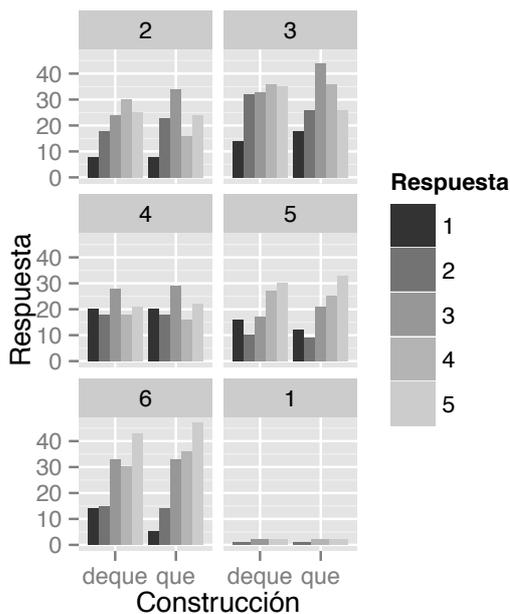


Figura 3: Histograma de la distribución de respuestas para ambos cuestionarios dividido por estrato.

La Figura 3. muestra que para las clases sociales (estratos) 4, 5 y 6 las distribuciones son casi idénticas, pero en los estratos 2 y 3 sí se observa algo de variación entre las oraciones con *que* y *de que*. Estos son los dos estratos en los que se podría esperar mayor variación por lo comentado anteriormente, que el *dequeísmo* es asociado

principalmente con las clases sociales más bajas. No obstante, si se comparan las distribuciones para cada construcción por estrato con el test Chi-cuadrado de proporciones, obtenemos que no hay diferencias estadísticamente relevantes entre éstas: para el estrato 2: $X=20$, $df=16$, $p=0.2202$; para el estrato 3: $X=15$, $df=12$, $p=0.2414$.

Una gráfica similar en la Figura 4. muestra las proporciones para hombres y mujeres.

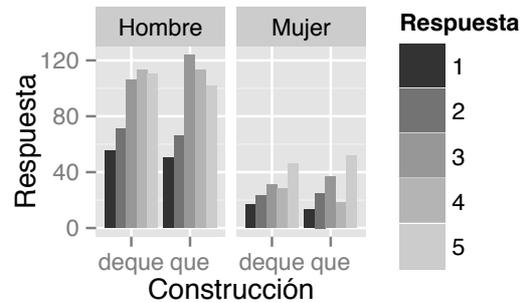


Figura 4: Histograma de la distribución de respuestas para ambos cuestionarios dividido por sexo.

Ambas distribuciones son bastante similares, ambas tienen una media de 3, y el test Chi-cuadrado de proporciones no revela diferencias estadísticamente relevantes (Para mujeres: $X=20$, $df=16$, $p=0.2202$; para hombres: $X=20$, $df=16$, $p=0.2202$).

Las distribuciones para cada clase social en ambos cuestionarios se presentan en la Figura 5.

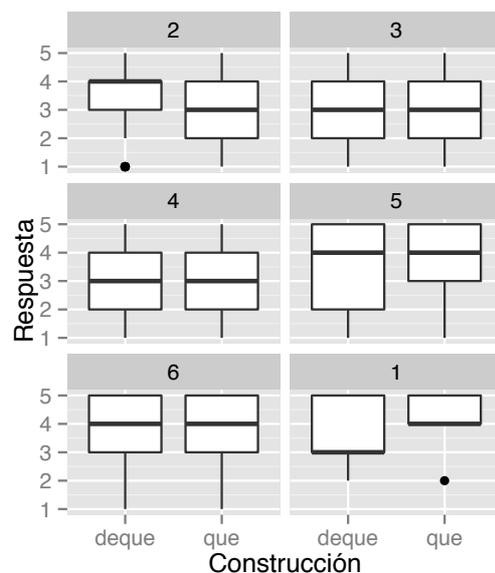


Figura 5: Diagramas de cajas de la distribución de respuestas para ambos cuestionarios dividido por estrato.

De las Figura 5. podemos concluir que la variación entre ambas construcciones es mínima y no parece estar presente en ninguna clase social.

5.2 Resultados del estudio de corpus

Para el análisis de corpus las proporciones de las 10 variables etiquetadas se presentan en la Figura 6.

En las Figura 6. podemos ver que la mayoría de las variables muestran proporciones casi idénticas para ambas construcciones. Las únicas variables que parecen ser diferentes son Contexto (condicional o real), Tipo de Verbo (percepción, psicológico, decir, otros) y Pronominal (uso de pronombre personal o pro-drop). Para intentar verificar las propuestas presentadas en la sección 3. se realizó un test Chi-cuadrado de proporciones para cada variable presentada como significativa en las Hipótesis de Evidencialidad e Hipótesis de Distancia e Incertidumbre. Para Modo los resultados son que la diferencia en proporciones no es estadísticamente significativa ($X=1.472$, $df=1$, $p=0.225$); para Tiempo los resultados son que la diferencia en proporciones no es estadísticamente significativa ($X=3.0473$, $df=1$, $p=0.08105$); para Número los resultados son que la diferencia en proporciones no es estadísticamente significativa ($X=0.0012$, $df=1$, $p=0.9729$). Ninguna de las predicciones de las Hipótesis de Evidencialidad e Hipótesis de Distancia e Incertidumbre se pudo confirmar con el corpus analizado.

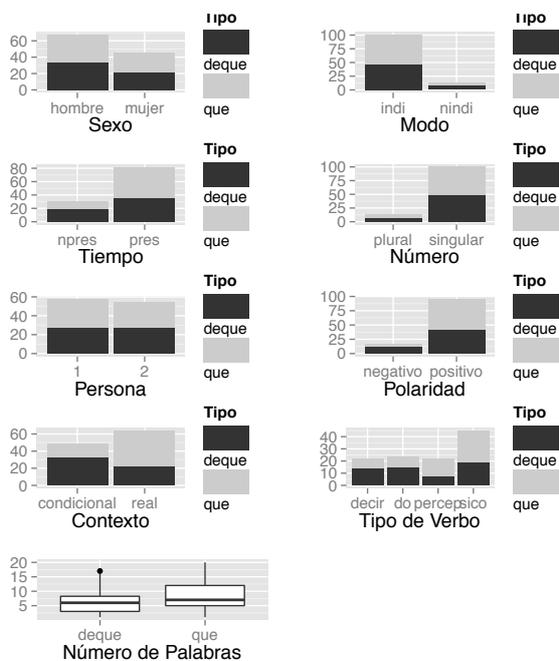


Figura 6: Distribución de las categorías gramaticales analizadas para el corpus PRESEEA.

A continuación se presentan los resultados del análisis logístico sobre los datos recolectados, así como la validación bootstrap del modelo. Las implicaciones de estos resultados serán discutidas en la siguiente sección.

La Tabla II presenta los índices del modelo adaptado a los datos. Para el modelo solo las variables significantes ($p < 0.05$) se mantuvieron, estas fueron: Contexto * Pronominal, y en cierta medida Tipo de Verbo.

La validación bootstrap para el modelo se presenta en la Tabla III. La validación retuvo 3 factores en 178 experimentos, 2 en 18, y solamente 1 en 4 experimentos.

En las Tablas II y III podemos ver que las variables Contexto, Ponominal y Tipo de Verbo todas fueron clasificadas como predictores significativos de la variable dependiente por el modelo de regresión logística. Ninguna de las otras variables tuvo relevancia como predictor de la variable dependiente.

Modelo			
	Coefficient	S.E.	Wald ZPr(Z)
Intercept	-2.3921	0.7982	-3.00 0.0027*
Contexto=real	3.04113	0.7342	4.14 0.0001***
Pronominal=sí	-0.1923	0.6746	-0.29 0.7756
TipoVerbo=percepción	322.876	0.7813	1.64 0.1018
TipoVerbo=psicológico	706.777	0.7586	3.14 0.0017**
TipoVerbo=decir	0.5152	0.7896	0.65 0.6233
Contexto * Pronominal	-2.4685	0.9736	-2.54 0.0112*

Probabilidad del modelo e índices de discriminación	
C	0.813
Pseudo R2	0.384
L.R.	38.40
D.f.	6
Pr (chi2)	0.0001***
Num. obs	0.113
Dyx	0.626
Brier	0.175

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $p < 0.1$

Tabla II: Coeficientes de la regresión logística.

Factores	origen	training	test	optimism	corrected
Dxy	0.6259	0.6625	0.5890	0.0735	0.5524
R2	0.3842	0.4341	0.3382	0.0959	0.2883
Intercept	0.0000	0.0000	-0.0012	0.0012	-0.0012
Slope	1.0000	1.0000	0.7803	0.2197	0.7803
Brier	0.1746	0.1672	0.1896	-0.0224	0.1970

Tabla III: Coeficientes de la validación bootstrap para la regresión logística.

La estadística D de Somers (Dyx en este caso) mide el efecto de las variables independientes sobre

la variable dependiente. Dyx puede tomar valores de -1 a 1, -1 indica una perfecta correlación negativa, mientras que 1 indica una perfecta correlación positiva. Un valor para Dyx de 0.626 es índice de un efecto moderado.

El índice de discriminación general C (Overall Discrimination Index) es una concordancia entre la probabilidad predicha y la respuesta real. Un valor C de 0.5 marca una predicción totalmente aleatoria, y un valor de 1 una capacidad de predicción perfecta. Un C de 0.813 indica una buena capacidad predictiva (0.8 indica una capacidad predictiva muy alta).

El Brier Score es un índice de qué tan exacta es una predicción probabilística. Puede tomar valores de 0 (el mejor) hasta 1 (el peor). En este caso 0.175 es indicativo de que el modelo tiene una capacidad predictiva alta (para explicaciones más detalladas de estos índices ver: Newson, 2006; Baayen, 2008; Gries, 2009).

En la Tabla III vemos que el optimismo para Dyx y Brier es relativamente alto (por encima de 0.02), y que los valores corregidos son un poco peores que en el modelo original. Esto mide qué tan optimista es el modelo, entre menos optimismo, mejor puede predecir el modelo datos nuevos. Estos valores parecieran indicar que el modelo no es muy capaz de predecir nuevos datos; pero, por otro lado la validación retuvo las tres variables independientes la gran mayoría de las veces, y consideró Contexto como significativa en 199 experimentos. En general podemos tener confianza en el modelo. Además, el modelo logró predecir correctamente la construcción en un 75% de las veces. Esto significa que 25% o más de la variación está dada por estas tres variables (es posible predecir el 50% de forma aleatoria). Un 25% de la variación se debe a otros factores no capturados en el modelo.

Las Figuras 7. y 8. a continuación muestra los efectos parciales de cada variable independiente sobre la variable dependiente. En la Figura 8. podemos ver que la interacción entre Contexto y Pronominal es significativa, y que el Contexto tiene más impacto sobre la probabilidad de la construcción si la oración no es pronominal.

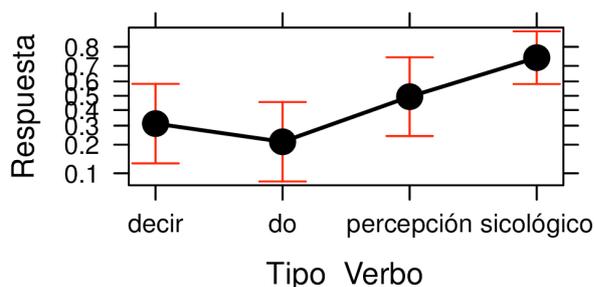


Figura 7: Efectos parciales de la variable Tipo de Verbo.

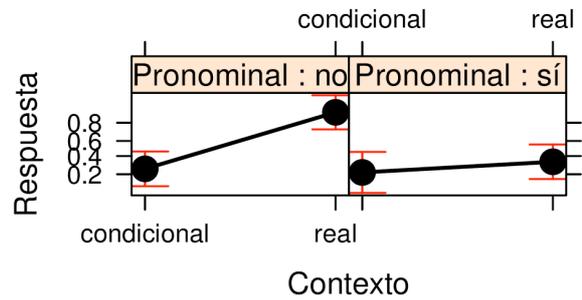


Figura 8: Efectos parciales de la interacción de variables Contexto * Pronominal.

6 Discusión

La encuesta realizada, tal como se describe en 5., no reveló ningún efecto estadísticamente positivo. La construcción *de que* no parece influenciar la forma en la que los hablantes ven la fuente de la información, o la seguridad que tiene el hablante sobre lo dicho. Por esta razón parece improbable que el *de queísmo* sea una forma de evidencialidad.

Los resultados del estudio de corpus son más reveladores. Aunque no se pudieron confirmar las predicciones iniciales propuestas por los autores de las Hipótesis de Distancia e Incertidumbre e Hipótesis de Subjetivización, el modelo demostró que hay una fuerte correlación entre la alternación entre *que* y *de que* y el contexto de la oración así como el uso de pronombres personales. Ambos factores apoyan la Hipótesis de Subjetivización. La semántica verbal parece jugar un papel en la elección de la construcción pero no es claro cómo puede ser interpretado, o si favorece alguna de las hipótesis, especialmente porque Del Moral (2008) predice que los verbos psicológicos deberían aparecer con mayor frecuencia con oraciones *de queístas*.

Situaciones irreales, preguntas hipotéticas y opiniones personales requieren un alto nivel de subjetividad por parte del hablante, y el sujeto de la oración debe estar más involucrado con la afirmación que si se reportan situaciones concretas y reales. No obstante, también se puede manifestar gran nivel de subjetividad por un evento real si el hablante así lo desea, es por esta razón que no existen pares de oraciones en las que *de que* no sea posible en lugar de *que* y viceversa. Esto también explica por qué los hablantes, en medio de una oración, pueden cambiar de *que* a *de que* o

viceversa con el propósito de aumentar o disminuir el nivel de subjetividad que quieren expresar.

Finalmente, el hecho de que el modelo no pudiera predecir más del 75% de las oraciones indica que aparte de los factores indicados en estos, hay otros motivos que juegan un papel importante en la alternación.

7 Consideraciones finales

Después de revisar los estudios más significativos sobre el dequeísmo, los resultados presentados en este artículo apoyan la hipótesis presentada por Del Moral (2008) de que el dequeísmo es un caso de gramaticalización de subjetividad, aunque desafían algunas de sus predicciones y su interpretación del fenómeno. Como Gutiérrez Rexach (2012), no se encontraron correlaciones sugeridas en estudios previos entre Tiempo, Modo o Persona y el uso de *de que*. Una razón para dudar de estos estudios es la falta de claridad metodológica y el hecho de que no se reportan datos totales. No obstante, hay que reconocer que el corpus usado para este trabajo es relativamente pequeño, y un corpus más grande produciría resultados más confiables.

Referencias

- Avila, Fernando. 2003. *Donde Va La Coma Where Is the Bed Going*. Editorial Norma.
- Baayen, R Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Battini, Berta Elena Vidal de. 1949. *El habla rural de San Luis*. Vol. 1. Universidad de Buenos Aires. Facultad de Filosofía y Letras.
- Bentivoglio, Paola. 1975. Queísmo y dequeísmo en el habla culta de Caracas. En: *Colloquium on Hispanic linguistics 1975*. Georgetown University: 1–18.
- Bentivoglio, Paola. 1980. El dequeísmo en Venezuela: ¿un caso de ultracorrección? En: *Homenaje a Ambrosio Rabanales*. En: *Boletín de Filología Santiago de Chile* 31: 705–719.
- Boye, Kasper and Peter Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88(1): 1–44.
- Bybee, Joan L. 2003. *Mechanisms of Change in Grammaticalization: The Role of Frequency*. Oxford: Oxford University Press.
- Bybee, Joan, Revere Perkins, and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.
- Company C., Concepción. 2004. Gramaticalización por subjetivización como prescindibilidad de la sintaxis. *Nueva Revista de Filología Hispánica* 52(2): 1–28.
- Cornillie, Bert and Nicole Delbecque. 2008. Speaker commitment: Back to the speaker. Evidence from Spanish alternations. *Belgian Journal of Linguistics* 22(1): 37–62. DOI: 10.1075/bjl.22.03cor. <http://benjamins.com/#catalog/journals/bjl.22.03cor/details> (consultado el 02/22/2013).
- Davis, Mark. (2002). *Corpus del Español: 100 million words, 1200s-1900s*. <http://www.corpusdelespanol.org> (consultado el 02/22/2013).
- Del Moral, Gabriel. 2008. Spanish dequeísmo a case study of subjectification. *Nueva Revista de Lenguas Extranjeras*. 10: 183–214. <http://bdigital.uncu.edu.ar/2643> (consultado el 02/22/2013).
- García, Erica C. 1986. El fenómeno (de) queísmo desde una perspectiva dinámica del uso comunicativo de la lengua. *Actas del II Congreso Internacional sobre el español de América*: 46–65.
- Geurts, Bart. 2000. Explaining grammaticalization (the standard way). *Linguistics* 38(4): 781–788.
- Gries, Stefan Th. 2009. *Statistics for linguistics with R: a practical introduction*. De Gruyter Mouton.
- Haspelmath, Martin. 2002. On Directionality in Language Change, with particular reference to unidirectionality in grammaticalization. En: *International Conference: New Reflections on Grammaticalization*.
- Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. *Typological Studies in Language* 59: 17–44.
- Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Cambridge Journal of Linguistics* 42(1): 25–70.
- Hopper, Paul J. and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Kanwit, Matthew. 2012. *Discourse Topic and (De)queísmo: A Variationist Study of the Spanish of Caracas*. <https://www.indiana.edu/~iulcwp/pdfs/12-Kanwit.pdf> (consultado el 02/22/2013).

- Lee, Binna. 2006. A Corpus-based Approach on the Development of Conjunction while. En: Lehmann, Christian. 2002. Thoughts on grammaticalization. Universität Erfurt, Philosophische Fakultät, Seminar für Sprachwissenschaft.
- Martínez-Sequeira, Ana Teresa. 2000. El dequeísmo en el español de Costa Rica: Un análisis semántico-pragmático. PhD thesis. University of Southern California.
- Newmeyer, Frederick J. 2000. Deconstructing grammaticalization. *Language Sciences* 23(2): 187–229.
- Newson, Roger. 2006. Confidence intervals for rank statistics: Somers' D and extensions. *Stata Journal* 6(3): 309.
- Norde, Muriel. 2002. The final stages of grammaticalization: affixhood and beyond. *Typological Studies in Language* 49: 45–66.
- Padró, Lluís and Evgeny Stanilovsky. May 2012. FreeLing 3.0: Towards Wider Multilingual-ity. *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*. ELRA. Istanbul, Turkey.
- Prieto, Luis. 1995. Análisis sociolingüístico del dequeísmo en el habla de Santiago de Chile. *Boletín de Filología*: 379–452.
- Quilis Sanz, María José. 1986. El dequeísmo en el habla de Madrid y en la telerradio difusión española. *Boletín de la Academia Puertorriqueña de la Lengua Española*: 14–139.
- Rabanales, Ambrosio. 1974. Queísmo y dequeísmo en el español de Chile. *Estudios filológicos y lingüísticos. Homenaje a Ángel Rosenblat*: 413–444.
- Rátiva, González. 2007. PRESEEA-Medellín-Co. Informe sobre el estado de la investigación. En: El español hablado en las comunidades hispánicas. Informe PRESEEA.
- Real Academia Española and Asociación de Academias de la Lengua Española. 2005a. Diccionario panhispánico de dudas. Santillana. <http://lema.rae.es/dpd/srv/search?id=vTr05If13D6tGOqCWV> (consultado el 02/22/2013).
- Real Academia Española and Asociación de Academias de la Lengua Española. (2005b). Diccionario panhispánico de dudas. Santillana. <http://lema.rae.es/dpd/srv/search?id=0WI01LaCjD655ud6n5> (consultado el 02/22/2013).
- Rosenblat, Angel. 1974. Estudios filológicos y lingüísticos: homenaje a Angel Rosenblat en sus 70 años. Caracas: Instituto Pedagógico.
- Schwenter, Scott A. 1999. Evidentiality in Spanish morphosyntax: a reanalysis of (de) queísmo. *Estudios de variación sintáctica. Iberoamericana*: 65–87.
- Serrano, María José. 1998. Estudio sociolingüístico de una variante sintáctica: el fenómeno dequeísmo en el español canario. *Hispania*: 392–405.
- Sheather, Simon. 2009. A modern Approach to Regression with R. Vol. 13. Springer.
- Silva Villar, Luis and Javier Gutiérrez Rexach. 2012. Predication, Complementation and the Grammar of Dequeísmo Structures. In: Current Formal Aspects of Spanish Syntax and Semantics. Ed. by Melvin González-Rivera and Sandro Sessarego. Cambridge: Cambridge Scholars Publishing: 8–41.
- Suárez, Manuel Almeida. 2009. Funcionalismo y antifuncionalismo en la teoría lingüística: a propósito del dequeísmo. In: XI Jornadas de Lingüística: Homenaje Al Profesor José Luis Guijarro Morales: Cádiz, 22 y 23 de Abril de 2008: 149.
- Traugott, Elizabeth Closs. 2003. From subjectification to intersubjectification. En: Motives for language change 124: 139.
- Traugott, Elizabeth Closs and Richard B. Dasher. 2001. Regularity in semantic change. Cambridge: Cambridge University Press.