

Colusión de normas sociolingüísticas en un hecho de variación sintáctica. Factores sociales en la selección de la variante *deber de* + infinitivo (vs. *deber*) en un corpus peninsular*

1. Introducción

La alternancia entre las perífrasis modales *deber/deber de* + infinitivo representa un fenómeno característico de variación sintáctica con una larga tradición en español.¹ Al mismo tiempo se trata de uno de esos hechos lingüísticos en los que las prescripciones normativas hallan escaso eco en el uso cotidiano, al menos en la lengua oral. No en vano la tradición gramatical lleva tiempo disponiendo que la citada alternancia posee un decisivo valor funcional al permitir la distinción entre contenidos de modalidad deóntica y epistémica. De este modo, por ejemplo, en (1) se diría que el hablante sigue estos dictados académicos al emplear la perífrasis sin preposición para argumentar a favor de la obligación moral de todo castellonense de colocar su provincia entre las mejores de España (modalidad deóntica). Y complementariamente, en (2) se sirve de la variante prepositiva, acorde de nuevo con la norma, para conjeturar acerca de la extraordinaria experiencia que debe de suponer para cualquiera persona la sensación de haber estado en peligro de muerte (modalidad epistémica).

(1) I: ... nosotros somos de Castellón, pues *debemos procurar* que Castellón sea una gran provincia (MCSCS 314)

(2) I: pues sí una

E: ¡qué dices!

I: una vez lo soñé , lo soñé// e:- no, no, no me he sentido nunca [en peligro de muerte] ¡pero *debe de ser* una experiencia! (MCSCS 314)

En la práctica, sin embargo, numerosos autores han llamado la atención acerca de la escasa huella que estas recomendaciones normativas encuentran en los hablantes de español (Yllera 1980; Gómez Manzano 1992; Samper et al. 1998, Troya 1998, Fernández de Castro 1999, Gómez Torrego 1999, Galué et al. 2004, García Fernández et al. 2006; Gómez Molina 2009, Blas Arroyo en prensa). Así, y por continuar con los ejemplos anteriores, extraídos de una de las entrevistas del *Macrocorpus sociolingüístico de Castellón y sus comarcas* (MCSCS), el mismo hablante que en (1) utilizaba la perífrasis *deber* + infinitivo de acuerdo con la norma para la expresión de contenidos deónticos, acude en el cotexto inmediatamente anterior a la otra variante para dar cuenta de valores de obligación similares, como puede verse ahora –(3)– en la reconstrucción completa de ese fragmento. Y si en (2) empleaba la perífrasis *deber de* conforme a los valores canónicos de la probabilidad y la conjetura, poco más adelante lo hace con la variante contraria para expresar un valor modal idéntico, como se aprecia en (4):

(3) I: pues sí, porque esto es como todas las cosas, cada uno *debe de cuidar* lo suyo/ nosotros somos de Castellón, pues *debemos procurar* que Castellón sea una gran provincia (MCSCS 314)

(4) I: ... hay mucha gente que está diciendo que le ocurre cosas un poco raras, pues al final tienes que pensar que alguna de ellas *debe ser* verdad también... (MCSCS 314)

Pese a la escasa correspondencia entre la mencionada oposición funcional y el uso real, demostrada en diversos trabajos empíricos a uno y otro lado del Atlántico (Valádez 1969, Luna 1980, Gómez Manzano 1992, Samper et al. 1998, Troya 1998, Fairclough

2000, Fernández Ulloa 2000, Galué et al. 2004, García Fernández et al. 2006; Gómez Molina 2009), la bibliografía al uso continúa haciendo de ella una piedra angular para enfocar este hecho de variación. Sin embargo, poco es lo que sabemos acerca de la posible incidencia de otros factores lingüísticos, estilísticos y sociales que pueden intervenir en la distribución de estas variantes y en la estimulación o freno a lo que parece un imparable proceso de cambio lingüístico, en el que una de las perífrasis (*deber* + infinitivo) gana claramente la partida a la otra (*deber de* + infinitivo). En un estudio previo a partir de los principios y métodos de la sociolingüística variacionista, he podido confirmar este hecho en las comarcas castellonenses que sirven también como base a la presente investigación, si bien en ellas la variante prepositiva encuentra todavía algunos factores estructurales y comunicativos que alientan su empleo (Blas Arroyo, en prensa). Entre estos factores no figura la mencionada oposición funcional, pero sí otros hechos de modalización que actúan en sentidos opuestos: por un lado, aquellos que estimulan la aparición de la perífrasis, como ocurre con la presencia en el cotexto inmediato de recursos de intensificación, o con las interacciones verbales donde impera un mayor grado de espontaneidad; por otro, y en sentido opuesto, aquellos contextos en los que el hablante muestra menor compromiso con la verdad de sus enunciados, bien sea por el empleo de recursos de atenuación en el cotexto previo, bien por el empleo de ciertas modalidades oracionales como la interrogativa. Se trata de conclusiones que, en otro orden de cosas, demuestran la hipótesis funcional formulada a propósito de otros hechos de variación sintáctica en los que se halla implicada la preposición *de*, como el (de)queísmo, según la cual esta partícula favorecería un mayor distanciamiento por parte del hablante (García 1986).

El presente estudio pretende completar ese análisis, analizando la distribución social que en estas mismas comunidades de habla presenta este hecho de variación. En concreto, y tomando como punto de partida de nuevo el grado de resistencia ofrecido en el discurso oral por la variante prepositiva, se persiguen los siguientes objetivos: a) analizar qué grupos y subgrupos sociales se caracterizan por un mayor empleo de dicha perífrasis y en qué contextos; b) comprobar si la distribución sociolectal advertida en la comunidad objeto de estudio ofrece pistas sobre el proceso de cambio lingüístico en el que se hallan inmersas las dos variantes; c) estudiar el modo en que los factores lingüísticos y estilísticos seleccionados como significativos en el estudio anterior actúan sobre esos grupos sociales; y d) y aun partiendo de la base de que la reseñada oposición modal epistémico-deóntica no es significativa para el conjunto de la muestra, examinar si las prescripciones normativas pueden ejercer todavía algún papel en la actuación lingüística de ciertos grupos.

Del estudio empírico llevado a cabo a partir de los materiales del *Macrocorpus sociolingüístico de Castellón y sus comarcas* (MCSCS) se deduce una compleja distribución sociolectal, en la que la variante prepositiva aparece impulsada tanto por grupos caracterizados por un menor prestigio social como por otros en los que las prescripciones normativas parecen inducir a un mayor grado de inseguridad, y eventualmente, a la ultracorrección. Todo ello como consecuencia de la tensión entre la norma común, favorable claramente a la variante no prepositiva, y una norma prescriptiva cuya esencia funcional presenta demasiada complejidad como para ser aplicada correctamente por el común de los hablantes, incluso por aquellos que resultan más sensibles al atractivo de esta última.

De los fundamentos de todo ello se ocupa el extenso apartado 4, donde se examinan con detalle los datos aportados por el análisis empírico. Antes, sin embargo, se ofrecen unos breves apuntes sobre el perfil sincrónico y diacrónico de este hecho de variación sintáctica, así como sobre su distribución dialectal en el mundo hispánico (§ 2), sección

a la que seguirá el detalle de los principales aspectos metodológicos de la investigación. Entre ellos ocupa un lugar destacado el análisis estadístico multivariante (Goldvarb X), que permite examinar tanto el grado de significación de los factores estudiados como la posible interacción entre estos, con el objeto de evaluar potenciales diferencias en el seno de algunos grupos sociales. El artículo se cierra, finalmente, con la síntesis de los principales resultados obtenidos, y la discusión acerca de los patrones de variación y cambio más relevantes que se derivan del estudio.

2. Apuntes diacrónicos y sincrónicos sobre un fenómeno de variación sintáctica

Desde hace siglos las dos construcciones verbales objeto de estudio se utilizan en español para la expresión de contenidos modales, tanto de naturaleza deóntica, relacionados con la obligación y/o la necesidad, como de orden epistémico, modalidad asociada a los contenidos de probabilidad y la conjetura (Yllera 1980, Hernández 1984, Alcina y Blecua 1989, Díez de Revenga/Igualada 1989-90, Alarcos 1994, Fernández de Castro 1999, García Fernández et al. 2006). Por ello, autores como Yllera (1980) concluyen que la distinción propugnada desde instituciones académicas resulta artificial ya que “no existe diferencia entre el sentido de *dever* y el de *dever de* durante la Edad Media e incluso posteriormente ambos pueden expresar obligación o probabilidad” (Yllera 1980:140). De este modo, si en la Edad Media el significado principal de *dever* + infinitivo era indicar “obligación, necesidad o conveniencia”, ya en el siglo XV era también relativamente frecuente su empleo para expresar suposición o conjetura. Al mismo tiempo, la mencionada autora encuentra testimonios tan tempranos como en el siglo XIII acerca de la modalidad deóntica expresada mediante la perífrasis *deber de*. En relación con esta última, incluso un autor partidario de mantener la oposición por su virtualidad funcional, como Cuervo (1893/1994: 808) observaba que “arraigado el uso de *deber de* en el sentido de *haber de* para indicar necesidad, se extendió la analogía a los casos en que éste denota inducción, conjetura, duda”.

Ahora bien, como ha destacado recientemente Balasch (2008), sorprende la escasa atención que se ha prestado a los patrones evolutivos de esa aparente «confusión» entre las dos perífrasis, y más concretamente al grado en que diversos factores han podido influir históricamente en la distribución de las dos variantes con los valores modales reseñados. Mediante la aplicación de una metodología variacionista a textos antiguos esta autora ha comprobado, por ejemplo, como aparte de algunas preferencias estilísticas significativas en algunos casos, entre los siglos XVII y XIX la variante *deber* + infinitivo no solo se extiende considerablemente a expensas de la variante con preposición, sino que –y lo que es más relevante– reemplaza a esta última primero en los contextos deónticos y luego en los epistémicos. Este proceso conduce, según Balasch (2008: 118), a que una misma forma (*deber*) pase a desempeñar dos funciones en un mismo canal de gramaticalización, que va desde la marcación de modalidad deóntica a la marcación de modalidad epistémica (obligación > probabilidad), en un sentido similar al postulado por Bybee et al. (1994: 199-202).

En la actualidad compiten actitudes diferentes hacia este fenómeno de variación sintáctica. Por un lado, se encuentran aquellos gramáticos que, de una forma más o menos explícita, apuestan por mantener la distinción funcional entre las dos perífrasis, cuyo origen preceptivo hay que remontar al *Diccionario de Autoridades*. En tiempos mucho más recientes, el *Esbozo* (1973) académico reconocía la confusión entre las dos construcciones tanto en el uso corriente actual como en épocas previas, pero seguía apostando por conservar la oposición. Una recomendación que, como cabía esperar, ha

sido asumida por gramáticas e instituciones normativizadoras diversas, así como por los modernos manuales de estilo españoles –El País 1990: 208; ABC 1993: 98; Telemadrid 1993: 189; TVE (Mendieta 1993: 75)– que condenan la no diferenciación. Representante destacado de esta actitud es Gómez Torrego (1989: 188-189), quien señala que esta confusión se halla en parte relacionada con el hecho de que “pocos hablantes conocen la diferencia entre estas dos construcciones perifrásticas, por lo que suelen emplearlas indistintamente [...]. Hay que hacer un esfuerzo para mantener esta diferencia, pues así se evitará incurrir en posibles ambigüedades”. Con todo, un poco más adelante matiza que “en la lengua oral (¡nunca en la escrita!) se puede permitir la omisión de la preposición *de* en los casos de posibilidad. No se justifica, por el contrario, intercalar *de* en los casos de obligación ambigüedades” (en parecido sentido, véanse R. Seco 1930/1967: 172-173 y M. Seco 1986: 136; asimismo, el lector interesado puede encontrar en Samper et al. (1998-1999) y Fernández de Castro (1999) testimonios muy ilustrativos acerca de las condenas hacia la confusión que, desde el siglo XIX en adelante, se han propalado desde diversos ámbitos académicos y normativos). Más recientemente, en la *Nueva Gramática de la lengua española*, las Academias de la lengua española (2009) navegan entre dos aguas en relación con este tema, al complementar las indicaciones gramaticales –“suele analizarse “*deber* + infinitivo” como modal radical y “*deber de* + infinitivo” como epistémico” (pág. 2143) – y preceptivas –“para expresar obligación se recomienda la variante sin preposición” (pág. 2144) –, con el reconocimiento de la notable variación existente en el mundo hispánico –“existen [...] numerosas excepciones a esta regla general” (pág. 2143) –, y eso que las ejemplificaciones parten en todos los casos de textos escritos y no orales.

Ahora bien, incluso quienes abogan por mantener la distinción como un recurso funcional óptimo para deshacer ambigüedades, reconocen la frecuente confusión que tiene lugar a diario en la lengua oral. En este sentido, por ejemplo, los autores del *Diccionario de perífrasis* (García Fernández et al. 2006) recuerdan que “la mayor parte de los hablantes utilizan ambas construcciones indistintamente, aunque algunos de ellos tiendan a utilizar (*deber* + infinitivo) como epistémica, pero no (*deber de* + infinitivo) como deóntica”. Estas impresiones han sido puestas a prueba empíricamente en los últimos años por diversas investigaciones en las que se recogen las tendencias que a este respecto operan en diversas comunidades de habla del mundo hispánico. Más aún, autoras como Gómez Manzano (1992: 162) señalan a partir de esos datos –en su caso extraídos del habla culta de Madrid– que “el análisis de estos casos *no nos ha permitido descubrir ninguna razón o norma para saber cuándo es más frecuente el uso con preposición o sin ella*. Hay un claro predominio de los casos sin preposición, *pero podemos hablar de variación libre en cuanto al empleo de una u otra*, ya sea para expresar la obligatoriedad ya sea para expresar la hipótesis o probabilidad” (la cursiva es nuestra). Opinión que, implícitamente, viene a refrendar Fernández de Castro (1999: 175) desde otra óptica cuando, de forma retórica, se plantea el siguiente interrogante: “¿tiene sentido hablar de 'confusión' entre dos cosas que en realidad nunca han sido distintas?”.

Como puede observarse en el gráfico 1, tampoco en las comunidades de habla castellanenses las dos perífrasis aparecen con la misma intensidad. Como era previsible, la forma sin preposición se usa en muchas más ocasiones (77.7%) que la perífrasis correspondiente con *de* (22.3%) (p-valor: 0,000).

INSERT GRÁFICO 1

Con todo, y por lo que a la variante prepositiva se refiere, si se extrapolan los resultados obtenidos en otras investigaciones –no siempre abarcadoras de todo el

espectro social, sino tan solo uno de los sociolectos, como sucede con los estudios basados en la norma culta– Castellón aparecería en una posición más conservadora, frente a la clara tendencia a la elisión que, con alguna excepción aislada –el caso de México, San José de Costa Rica o Houston–, se advierte en las comunidades latinoamericanas. Como puede comprobarse en la tabla 1, las cifras castellanenses están más próximas a las de otras poblaciones españolas, como Bermeo, Madrid, Valencia o Las Palmas de Gran Canaria.

INSERT TABLA 1

Por otro lado, los resultados disponibles permiten confirmar que la expresión se utiliza tanto para los usos considerados como canónicos por la normativa –la expresión de la probabilidad–, como para los no canónicos, cifrados en los contenidos modales de obligación. En Castellón, estos empleos se reparten casi en la misma proporción, con una ligera ventaja para la expresión de la modalidad epistémica (54.3%) que para la deóntica (45.7%). No se cumplen así los vaticinios que realizan algunos gramáticos, que dan por sentado el empleo no canónico de *deber* para la probabilidad –al tiempo que se muestran más benévolos–, pero no así los de *deber de* para la obligación. Así, García Fernández et al. (2006: 113) indican en su *Diccionario de perífrasis* que “...la mayor parte de los hablantes utilizan ambas construcciones indistintamente, aunque algunos de ellos tiendan a utilizar (deber + infinitivo) como epistémica, *pero no (deber de + infinitivo) como deóntica* (la cursiva es nuestra). Los hechos desmienten también la opinión de M. Seco (1986: 135), quien sostiene que "por el hecho de que la suposición firme se suele expresar en forma de obligación, es más frecuente el uso, no académico, de 'deber + infinitivo' en el sentido de probabilidad, que el de 'deber de + infinitivo' en el sentido de obligación".

Por otro lado, el estudio confirma también que la pretendida alternancia entre las dos variantes cifrada en la oposición entre las modalidades epistémica y deóntica apenas se sostiene empíricamente. En el conjunto de la comunidad analizada, los hablantes utilizan indistintamente ambas perífrasis para la expresión de la probabilidad y la obligación o necesidad. Ahora bien, hay otros factores lingüísticos y estilísticos relacionados con la modalización que ayudan a explicar la variabilidad sintáctica. Algunos como el grado de espontaneidad o el énfasis expresivo, actúan a favor de la variante prepositiva. Tanto en las interacciones verbales más espontáneas como en aquellos contextos en los que el hablante introduce signos de intensificación para resaltar sus juicios, los índices de presencia de la preposición son significativamente más elevados. Otros factores, por el contrario, como la atenuación o la modalidad oracional interrogativa lo hacen en contra. En el presente estudio se comprobará cómo se distribuyen estos condicionantes por todo el espectro social.

3. Metodología

La muestra se halla integrada por 151 hablantes, incluidos en 126 entrevistas semidirigidas correspondientes al *Macrocorpus sociolingüístico de Castellón y sus comarcas* (MCSCS), un amplio corpus de habla oral representativo de todo el espectro social en las comunidades de habla castellanenses. Situada en el extremo norte de la Comunidad Valenciana, Castellón es la provincia menos poblada de esta comunidad autónoma española. Sus 600.000 habitantes se concentran sobre todo en el área metropolitana de Castellón de la Plana, capital de la provincia. Por el contrario, las zonas del interior se hallan mucho menos densamente pobladas. Desde el punto de vista sociolingüístico, hay que destacar también la presencia de dos áreas claramente delimitadas: mientras que las áreas costeras son valencianohablantes desde los tiempos

de la Reconquista, y por tanto, bilingües –tradicionalmente, con el español como la lengua A de una situación diglósica– en algunas comarcas del interior nunca se ha hablado otra lengua que el castellano, dado que fueron repobladas mayoritariamente por aragoneses.

Los informantes del corpus se hallan distribuidos en diferentes grupos representativos de otros tantos factores sociológicos que conforman la estructura original del MCSCS, y que, en mayor o menor medida, se han revelado significativos en otros estudios de variación anteriores (Blas Arroyo 2007, 2008). Se trata del *sexo* – hombres (52%),² mujeres (48%)–, la *edad* (jóvenes (35%), adultos (32%), mayores (33%)– y el *nivel de instrucción* (primarios (20%), secundarios (44%), superiores (36%). En una fase postestratificacional, a los anteriores se añaden el *estatus social* – medido a partir de la combinación de factores parciales como el nivel de instrucción, la profesión, el tipo de residencia y el nivel de renta–, y que da lugar a tres estratos – diferentes –bajo (24%), medio-bajo (46%), medio-alto (30%)–; y el *lugar de residencia* –Castellón de la Plana (54%), Provincia (46%). Por último, la particular configuración lingüística de la población analizada, con la presencia habitual de dos lenguas en el repertorio lingüístico de la comunidad de habla (Blas Arroyo 2009), justifica el interés por un factor adicional como es la *lengua dominante* de los hablantes –castellano (58%), valenciano (42%)–. Junto a estas variables sociológicas cada ocurrencia del presente fenómeno de variación fue codificada también de acuerdo con diversos factores lingüísticos y comunicativos, de cuya incidencia he dado cuenta en el estudio referido anteriormente (Blas Arroyo, en prensa). Entre los más sobresalientes, por su significación estadística en el conjunto de la muestra (véase tabla 2), destacan diversos recursos expresivos, como el énfasis y la atenuación, así como la modalidad oracional (interrogativa) y el grado de espontaneidad de las interacciones verbales. Asimismo, se analizaron otros parámetros lingüísticos y estilísticos cuyo poder explicativo fue desechado, sin embargo, por el análisis de regresión, de ahí que sus datos se hayan omitidos en las tablas que se incluyen en el presente trabajo. Este es el caso del contexto fono-sintáctico, el número de sílabas del grupo verbal, la persona y número del verbo auxiliar, la sintaxis del sujeto (omitido/explicito), el grado de animacidad del sujeto (humanos/no humanos), el tipo de cláusula (subordinadas/otras), el tiempo y modo verbales, la forma del verbo auxiliado (simple/compuesto), el modo de acción del verbo auxiliado (estativos/dinámicos), el grado de adyacencia entre verbos auxiliar y auxiliado, la diátesis (activa/pasiva), el tipo de secuencia textual (argumentativas/otras), las fases de la interacción (iniciales/finales) (para más detalles sobre estos factores, véase Blas Arroyo, en prensa).

La extracción de las variantes ofrecidas por el MCSCS mediante el programa de concordancias WordSmith 4.0 supuso la obtención inicial de 491 casos, de los que 79 tuvieron que ser eliminados del análisis por diversas razones (v. gr. ocurrencias de la variable en las que el infinitivo finalmente no aparece, como consecuencia de titubeos, rectificaciones o interrupciones; posible “contaminación” de la variante empleada en la pregunta previa por el entrevistador; etc.; para más detalles sobre estas exclusiones, véase Blas Arroyo, en prensa). Las 412 ocurrencias finales de la variable se analizaron mediante el auxilio del programa de regresión logística Goldvarb X, con el que se obtienen los pesos probabilísticos de aquellos factores seleccionados como significativos, así como su rango explicativo (Sankoff, Tagliamonte y Smith, 2005). Así, los pesos probabilísticos que aparecen en la tabla 2 indican la probabilidad –en una escala de 0 a 1– de que un factor concreto influya en la selección de una determinada variante. Los valores numéricos próximos a 1 favorecen la elección de la variante elegida (en este caso, la perífrasis con preposición), mientras que, en el extremo

opuesto, los más cercanos a 0 la desfavorecen. Por su parte, los situados en cifras intermedias, cercanas a 0.5, ejercen una influencia neutra. Por otro lado, el programa multivariante no solo nos indica qué grupos de factores son significativos, sino también la jerarquía entre estos. Esta se obtiene a través de la comparación entre los *rangos* respectivos, resultantes de calcular la diferencia entre las cifras mayor y menor de cada grupo de factores. Al mismo tiempo, el programa estadístico permite examinar posibles interacciones entre estos factores, con el fin de afinar en la interpretación de los resultados (Tagliamonte 2006).

Los datos globales de este análisis multivariante aparecen en la tabla 2, con la que se abre el apartado 4. A continuación se examinará con detalle el peso explicativo de los factores sociales seleccionados por el programa de regresión logística, así como los cruces más significativos con otras variables lingüísticas y sociales, al objeto de obtener una radiografía sociolingüística más ajustada. Por último, las eventuales diferencias obtenidas mediante estos cruces serán sometidas a la prueba de múltiples rangos de Fisher, apta para la comparación de medias pareadas. Por razones expositivas se ofrecerán únicamente los valores p. que indican la probabilidad de que las diferencias puedan ser o no debidas al azar (nivel de significación: $p.<0,05$).

4. Análisis de los resultados

Como se indicaba en el párrafo anterior, la tabla 2 muestra los resultados generales del análisis. Los factores significativos tienen asignado un peso probabilístico (*factor weight*), así como un rango (*range*), con el que se mide la relevancia explicativa de cada variable independiente incluida en el análisis: cuanto mayor es aquel, mayor es también la importancia predictiva de esta última. Los factores no seleccionados por el análisis multivariante (entre corchetes en la tabla) poseen, lógicamente, menor interés interpretativo, pero en algunos casos muestran tendencias marcadas en las distribuciones porcentuales correspondientes, cuya interpretación es necesario tener en cuenta. En unos casos para descartar que su falta de significación obedezca a la interacción o dependencia con otros factores; en otros para vislumbrar posibles tendencias distribucionales a las que, sin embargo, falta el aval estadístico por su menor representación muestral.

INSERT TABLA 2

4.1 Sexo

Como puede comprobarse en la tabla 2, las mujeres castellanenses (27.3%) superan a los hombres (17.8%) en la realización de la variante con preposición, diferencias que resultan significativas a la luz del análisis estadístico multivariante (pesos probabilísticos de .568 y .437, respectivamente).

Estudios previos en otras regiones del mundo hispánico muestran una notable variabilidad en relación con este mismo factor, aunque no siempre resulte fácil establecer comparaciones fiables, bien por las diferencias muestrales entre unas investigaciones y otras, tanto de naturaleza cuantitativa como cualitativa (buena parte de esos estudios abarcan tan solo al sociolecto alto), bien por la ausencia de análisis estadísticos de naturaleza inferencial, lo que impide comprobar la significación de las diferencias generolectales observadas. Con todo, en el habla culta de Madrid Gómez

Manzano (1992) obtiene diferencias del mismo tenor, si bien estas son mucho más abruptas que las advertidas por nosotros (mujeres: 72.7%; hombres 27.3%).³ Por el contrario, en otros casos las diferencias generolectales son inversas a las observadas aquí. Así, Gómez Molina (2009) revela un mayor uso de la perífrasis prepositiva entre los hombres del sociolecto alto de Valencia (54.4%) que entre las mujeres (45.6%). Por su parte, Galué et al. (2004) señalan porcentajes de retención similares a los castellanenses entre los hombres de Caracas, pero a diferencia de aquellos, las mujeres de la capital venezolana omiten la preposición de forma absoluta. Por último, no faltan investigaciones en las que aprecian diferencias nulas (Troya 1998) o poco representativas (Samper et al. 1998) entre ambos grupos.

De regreso al corpus castellanense, las mujeres superan a los hombres no solo en términos absolutos, sino también en los cruces con numerosos factores lingüísticos de los considerados en el estudio. Por reducir el alcance de estos cruces a los factores estructurales y estilísticos seleccionados como significativos por el análisis multivariante, en la tabla 3 se advierte, por ejemplo, como el empleo de la variante prepositiva es más amplio entre las mujeres en aquellos contextos que favorecen la retención de la preposición. Como he mostrado en otro lugar (Blas Arroyo, en prensa), uno de esos contextos corresponde a aquellos enunciados en los que el hablante utiliza en el entorno inmediato a la variable marcas lingüísticas destinadas a intensificar expresivamente el contenido de sus ideas, argumentos, dudas, etc. Así ocurre, por ejemplo, en (5), donde la hablante que está en el uso de la palabra se sirve de mecanismos de intensificación como cuantificadores (*tanta televisión*), adverbios (*totalmente*) y oraciones consecutivas (*tantas cosas que no debería ver*) para reforzar la comparación entre las cosas que ocurrían en su niñez y las que rodean hoy a los niños. En esos contextos enfáticos, el empleo de la preposición aumenta significativamente (.697) en el conjunto de la muestra, pero como se puede ver en la tabla 3, lo hace todavía más entre las mujeres que entre los hombres (con todo, estas diferencias generolectales no resultan significativas según la prueba LSD de Fisher; p-valor: 0,701).

(5) bueno/ pues: es totalmente diferente/ las travesuras mías, o de mi juventud, o de mi niñez/ con compañía de los otros chicos a lo que se hace hoy en día// hoy en día la juventut como ve **tanta** televisión/ hay **tantas cosas que** no *deberia de ver*/ y: y- o sea es otro mundo **totalmente** diferente (MSCS 413)

Por el contrario, sí son significativas las diferencias relacionadas con el tenor de la entrevista, otro de los factores que favorece la retención de la variante prepositiva. Las muestras de habla que componen el *Macrocorpus sociolingüístico de Castellón y sus comarcas* presentan cierta diversidad, derivadas en el modo en que se llevaron a cabo las entrevistas semidirigidas. En algunos casos, los entrevistadores participaban activamente en el desarrollo de las conversaciones, por medio de interrupciones, señales de retroalimentación, aportando sus opiniones y/o experiencias personales en diversas fases del diálogo, cuando no, permitiendo la participación de más de un interlocutor, lo que generalmente facilitaba una considerable mayor distensión comunicativa. De las 126 entrevistas que componen el corpus, 46 (37%) pertenecen a esta categoría. En el extremo contrario, sin embargo, algunos entrevistadores se limitaban básicamente a seguir el guión de las preguntas que servían como base para la entrevista, y dejaban que fueran los entrevistados quienes llevaran el peso de la conversación. La representación de este tipo de entrevistas es ligeramente menor, con una cifra de 38 (30%). El resto del corpus (42; 53%) está compuesto por entrevistas en las que se aprecian signos intermedios entre los dos tipos extremos mencionados, por lo que no han sido tenidas en consideración para el análisis del presente factor.

Como se aprecia en la tabla 3, la incidencia de este factor estilístico es especialmente reveladora entre las mujeres. Mientras que las diferencias entre estas y los hombres se neutralizan en las entrevistas menos espontáneas (H: 18% / M: 17%; p-valor: 0,908), en las más espontáneas las mujeres triplican las ocurrencias de la preposición (38%) con respecto a las de los hombres (11%) (p-valor: 0,000).

INSERT TABLA 3

Ahora bien, un análisis más detenido del factor sexo permite comprobar como estas diferencias favorables a las mujeres tienen lugar preferentemente en algunos subgrupos sociales, quedando neutralizadas en otros. Así ocurre, por ejemplo, cuando se observa la interacción entre el sexo y el estatus social de los hablantes. En efecto, tal como puede advertirse en la tabla 4, las diferencias entre hombres y mujeres desaparecen en los extremos del espectro social, representados por la clase baja (H: 25% / M: 24%; p-valor: 0,983) y media-alta (H: 17% / M: 17%; p-valor: 1.000), respectivamente. Por el contrario, en la clase media-baja, las mujeres (34%) superan en más del doble de realizaciones de la variante prepositiva a los hombres (14%) (p-valor: 0,003).

En parecido sentido, cabe hablar de las diferencias generolectales en función del lugar de residencia de los informantes, más acusada en los contextos rurales (H: 13% / M: 26%; p-valor: 0,025), que en los urbanos (H: 22% / M: 28%; p-valor: 0,280).

INSERT TABLA 4

A la vista de estos resultados, cabe sospechar que ciertos subgrupos sociales puedan verse impelidos al empleo de la perífrasis prepositiva en mayor medida que otros, impulsados tal vez por una vaga conciencia lingüística asociada a normas dictadas desde instancias normativas. Así ocurre con grupos en los que la tradición sociolingüística ha advertido mayores dosis de inseguridad, como las mujeres de clase media-baja, o las residentes en enclaves rurales, donde los preceptos de la norma académica llegan, por lo general, más debilitados.⁴

Para poner a prueba esta hipótesis, se ha procedido al recuento por separado de las frecuencias de uso de la perífrasis prepositiva en dos contextos:

a) canónicos, esto es, los enunciados con valor modal epistémico según la norma del español, como en (6), donde el hablante realiza un cálculo mental acerca del número aproximado de alumnos con los que compartía clase durante su niñez; y

b) no canónicos, en los que la misma variante expresa contenidos deónticos, relacionados con la obligación, como sucede en (7), donde otro hablante argumenta acerca de la necesidad de que los dos miembros de una pareja cuiden de sus hijos, especialmente si ambos trabajan fuera de casa.

(6) ...y la época que: umm/ que nosotros estábamos allí *debían de haber* pues el doble de niños de los que cabían realmente en aquel colegio (MCSCS 18)

(7)... si tú tienes un hijo con otra persona y trabajan las dos personas, pienso que *deben de cuidar* los dos, la pareja, de su hijo ¿no? (MCSCS 129)

Los resultados de este análisis comparativo aparecen en el gráfico 2, donde puede observarse una apreciable diferencia entre hombres y mujeres en función de esos valores modales asociados a la perífrasis. Así, el empleo de la preposición se acentúa entre las mujeres (36% vs. 21% entre los hombres; p-valor: 0,001) cuando se expresan valores epistémicos, aquellos en los que la norma académica prescribe, justamente, dicha variante, mientras que, por el contrario, tales diferencias generolectales aparecen

prácticamente neutralizadas en los contextos deónticos (H: 20% / M: 21%; p-valor: 0.772)

INSERT GRÁFICO 2

De hecho, un reanálisis del factor sexo mediante Goldvarb X, considerando ahora cada uno de estos contextos modales por separado arroja resultados muy reveladores. Para la expresión de la probabilidad, el factor no solo continúa siendo seleccionado como significativo por el análisis multivariante, sino que, además, su rango explicativo prácticamente triplica (32) al que se obtiene en la muestra global (13) (ver tabla 5), dado que las mujeres favorecen ahora con mucha más claridad la variante prepositiva (.663) que los hombres (.345). Por el contrario, dicha significación estadística desaparece cuando se considera aisladamente la expresión de la necesidad u obligación, contextos en los que las diferencias generolectales prácticamente desaparecen (H: 21% / M: 20%).

INSERT TABLA 5

Por otro lado, el mayor seguimiento de la norma académica por parte de las mujeres, pero especialmente entre algunos subgrupos más preocupados con el prestigio de las variables lingüísticas, ayudaría a explicar también ciertas diferencias llamativas que se observan en el cruce con el nivel de instrucción de los hablantes. No en vano, y como se observa de nuevo en la tabla 4, la retención de la variante prepositiva tiene uno de sus locus principales entre las mujeres con estudios superiores, quienes triplican (32%) a los hombres del mismo nivel (11%) en los usos de la perífrasis *deber de* (p-valor: 0,001). Sin embargo, ello se produce fundamentalmente en los contextos de modalidad epistémica, de nuevo aquellos en los que la norma prescribe el empleo de dicha variante. Como se observa en el gráfico 3, las distancias entre hombres y mujeres en el grupo de hablantes con formación universitaria son muy abultadas (H: 13% / M: 44%; p-valor: 0,002), seguidas por las obtenidas entre los informantes de estudios secundarios (H: 6% / M: 29%; p-valor: 0,015). Por el contrario, en aquellos hablantes donde la presión normativa es menor –hablantes de estudios primarios– las diferencias porcentuales dejan de atenerse a dicho esquema (H: 56% / M: 40%), y pierden la significación estadística observada en los otros pares (p-valor: 0,481).

INSERT GRÁFICO 3

Ahora bien, llegados a este punto ¿cómo explicar que, aunque en menor medida, estas diferencias favorables a las mujeres de formación superior se mantengan incluso en los contextos no canónicos, esto es, los que expresan contenidos modales deónticos? En efecto, como se aprecia en el gráfico 4, es este de las mujeres universitarias el único subgrupo en el que tiene lugar dicha distribución sociolingüística (H: 8% / M: 21%; pese a ello, diferencias no significativas en la presente muestra; p-valor: 0,113), que se invierte (primarios: (H: 42% / M: 30%; p-valor: 0,382) o neutraliza (secundarios: (H: 20% / M: 17%; p-valor: 0,644) en el resto. A mi juicio, ello podría obedecer a la aplicación de mecanismos de ultracorrección, que llevarían a las mujeres de formación superior al empleo, por analogía, de la misma norma en contextos no canónicos, en un fenómeno sintáctico de variabilidad cuyos preceptos normativos, defensores de la oposición modal, resultan especialmente complejos (sobre esta cuestión, véase más adelante § 5).

INSERT GRÁFICO 4

En resumen, de los datos analizados en el presente apartado se colige que el uso de la preposición se ve favorecido sobre todo en algunos subgrupos femeninos especialmente afectados por el prestigio de las norma académica, lo que impulsa su utilización canónica en los contextos epistémicos, pero también, por analogía, en los no canónicos, relacionados con la expresión de la necesidad u obligación. Ahora bien, como se verá a continuación ello no sucede en todo el espectro social, advirtiéndose diferencias muy destacadas entre diferentes grupos generacionales.

4.2 Edad

El gráfico 5 muestra los porcentajes de realización de la variante prepositiva por edades. Como señalaba anteriormente (véase § 3), la muestra aparecía inicialmente dividida en tres grupos, representativos de otros tantos cortes generacionales en el MCSCS: 1) menores de 30 años; 2) entre 31 y 54 años; y 3) mayores de 55 años. En el gráfico se observa un claro perfil de distribución lineal, en el que la realización de la variante prepositiva desciende progresivamente en paralelo a la edad de los hablantes: cuanto menor es la edad de estos, menores son también las frecuencias de realización de la perífrasis *deber de*; y como contrapartida, los porcentajes de realización de esta variante aumentan en las generaciones adultas. Pese al tenor idéntico de estas distribuciones, las diferencias muestrales existentes entre algunos grupos etarios – especialmente parcas son las ocurrencias de la variable en el tercer grupo de edad– hacen que el análisis de regresión logística no las haya seleccionado inicialmente como significativas, aunque sí las que oponen al grupo más desfavorecedor de la variante, el de los jóvenes menores de 30 años (.379), del resto de la pirámide genolectal, hablantes que, por razones expostivas, quedan etiquetados bajo el rótulo de *adultos* (.568) (véase tabla 2). Estos datos, que darían cuenta en tiempo aparente de la existencia de un cambio lingüístico en marcha, son acordes con la tendencia a la progresiva desaparición de la variante prepositiva en beneficio de la perífrasis sin preposición observada en la evolución del español (Balasch 2008).

INSERT GRÁFICO 5

La relevancia de este factor hace que se neutralicen o atenúen las diferencias observadas a partir de otros factores sociales. Así ocurre, por ejemplo, con el sexo, cuyos datos se analizaban en el apartado anterior. A este respecto es significativo (véase la tabla 4) que las frecuencias favorables a las mujeres en el uso de la variante prepositiva se atenúen considerablemente entre las jóvenes, cuyo porcentaje apenas supera levemente (20%) al de los hombres de su misma edad (16%) (p-valor: 0,147). Por el contrario, entre los adultos esas diferencias son mucho más abultadas y apuntan en el sentido reseñado más arriba (H: 19% / M: 33%; p-valor: 0,003).

Un efecto similar se aprecia en el cruce entre la edad y el lugar de residencia de los hablantes. En efecto, el gráfico 6 permite comprobar como las diferencias genolectales se disparan entre los hablantes capitalinos (Jóvenes: 18%; Adultos: 28%; p-valor: 0,004), pero desaparecen por completo entre los residentes en las zonas rurales (Jóvenes: 19%; Adultos: 19%; p-valor: 0,948), menos expuestos a la presión de las normas prescriptivas.

INSERT GRÁFICO 6

Los únicos incrementos en el uso de la preposición entre los jóvenes tienen lugar como consecuencia de la interacción con algunos factores lingüísticos. La oposición modal epistémico-deóntico aparece entre ellos. Aunque, como se ha señalado anteriormente, el factor no es significativo para explicar la variabilidad en el conjunto de la muestra, sí ofrece aquí y allá diferencias que podrían justificarse a partir del mayor o menor contacto de los grupos sociales con la norma. De este modo, puede ser revelador el hecho de que los porcentajes de realización de la variante prepositiva no difieran entre los contextos epistémicos (25%) y deónticos (25%) (p-valor: 0,110) en los hablantes adultos, pero sí lo hagan entre los más jóvenes. En efecto, entre estos últimos se aprecia un incremento en los usos canónicos, aquellos que se asocian normativamente con la expresión de la probabilidad (24%), mientras que descende en los empleos no canónicos (obligación) (15%) (p-valor: 0,045). El contacto habitual de una mayoría significativa de estos hablantes jóvenes con el sistema educativo, difusor de la norma, quizá ayude a explicar estas diferencias.

Con todo, mucha mayor relevancia que la oposición modal posee el énfasis expresivo para explicar las diferencias relacionadas con la edad. En efecto, como revela la tabla 6, los contextos enfáticos representan uno de los estímulos principales para el incremento de la preposición, especialmente entre los hablantes jóvenes (54%) que en estos casos llegan a hacer incluso un uso superior al de la variante no prepositiva (46%). Este incremento es todavía más llamativo si se consideran las proporciones alcanzadas por esos mismos hablantes en los contextos no enfáticos, que se sitúan entre las más bajas de la muestra (11%). Por el contrario, entre los adultos las distancias entre unos contextos y otros son mucho más tenues (enfáticos: 37%/ no enfáticos: 22%), lo que revela un menor peso explicativo de este factor entre los miembros de este segundo grupo etario.

INSERT TABLA 6

Como contrapartida, entre los jóvenes es también más acentuada la intensidad de aquellos factores que menos favorecen la variante prepositiva, como ocurre con la atenuación. Veíamos en otro lugar como, con independencia de la modalidad epistémica o deóntica que transmitan en sus enunciados, los hablantes pueden valerse en el discurso de diversos recursos lingüísticos destinados a mitigar su compromiso con la verdad de aquellos o, en los casos extremos, añadir matices de duda, incertidumbre, inseguridad, etc. (Blas Arroyo, en prensa). Esto último es, justamente, lo que puede apreciarse en el ejemplo siguiente, donde la hablante que está en el uso de la palabra expresa sus dudas acerca del efecto que en ella pueda tener en el futuro su traslado como maestra a un colegio grande de Castellón, desde su actual destino en una pequeña escuela rural.

(8) bueno, pero pienso que también es una cosa que: a la larga *debe debe de cansar*, pero no, no lo sé (MCSCS 210)

La presencia en el contexto inmediato a la variante de marcadores discursivos (*bueno*), predicados doxásticos (*pienso que...*), alargamientos vocálicos (*que:*), repeticiones (*debe debe*) y rutinas conversacionales (*no lo sé*) se dan la mano en este fragmento para completar ese sentido de incertidumbre que la hablante desea transmitir.

Sin embargo, el efecto atenuador de estos recursos no es privativo de los contenidos de la modalidad epistémica, como demuestra este otro ejemplo, en el que, mediante el auxilio del comprobativo *¿no?*, otro hablante mitiga su juicio acerca de la necesidad de que la gente llene su tiempo de una manera provechosa, aunque le toque mucho dinero en la lotería.

(9) ...pero, *deberías llenarlo* con otra cosa ¿no?/ el tiempo, el tiempo ... (MCSCS 293)

En un estudio anterior (Blas Arroyo, en prensa) he tenido la ocasión de comprobar como estos recursos de atenuación poseen un efecto claramente desfavorecedor en la selección de la perífrasis con preposición (.381; 14%). Pues bien, en este punto cabe señalar que dicho efecto es especialmente intenso entre los jóvenes, quienes arrojan una de las frecuencias más bajas de todo el espectro social (6%), a considerable distancia de los hablantes adultos (26%), cuyos índices no solo se sitúan por encima de la media, sino también a escasa distancia de los contextos no atenuados (29%).

4.3. Nivel de instrucción

Más aun que el sexo o la edad, el presente factor aparece como uno de los más significativos de todos los considerados en la investigación (rango 28) y encabeza los de naturaleza sociológica analizados en el presente trabajo. Como puede apreciarse en el gráfico 7, en el conjunto de la muestra se aprecia una disminución clara en el empleo de la variante prepositiva conforme aumenta el nivel de instrucción de los hablantes. Con todo, dentro de ese patrón común se aprecian dos tipos de distribución lineal. Por un lado, una estratificación abrupta entre el grupo que más favorece la variante (.713), el de estudios primarios, con frecuencias generales que se sitúan claramente por encima de la media global, y el resto de los grupos. Entre estos últimos, hay también una diferencia, de intensidad mucho menor en todo caso, que sitúa a los hablantes de formación superior como los más desfavorecidos de todos (.432), seguidos a corta distancia por los hablantes de estudios secundarios (.460).

INSERT GRÁFICO 7

La robustez del factor se demuestra no solo en los datos generales, sino también en el cruce con los factores lingüísticos seleccionados como significativos, donde las diferencias se atienen al esquema distribucional reseñado. De este modo, y como puede apreciarse en la tabla 7 el énfasis alienta considerablemente la elección de la perífrasis con preposición, pero de manera muy destacada –67%– entre los hablantes de estudios primarios, seguidos a notable distancia por los de estudios secundarios –39%– y superiores –33%– (p-valor: 0,053). Por el contrario, los recursos de atenuación en el contexto circundante a la variante rebajan los índices de la variante en todo el espectro social, pero de nuevo con una intensidad variable (superiores: 9%; secundarios: 10%; primarios: 30%; p-valor: 0,007).

INSERT TABLA 7

Pese a ello, otros datos apuntan a la existencia de algunas diferencias interesantes en el seno de estos grupos. Uno de esos datos tiene que ver de nuevo con la oposición modal epistémico-deóntica. Los resultados de la tabla 7 permiten comprobar la existencia de diferentes patrones de distribución en función del tipo de modalidad expresada. Así, en la modalidad deóntica, los porcentajes de empleo de la preposición se atienen al perfil lineal esperado, con los empleos no canónicos de *deber de* más acusados entre los hablantes de estudios primarios (35%), seguidos a distancia por los hablantes de estudios secundarios (19%), y más todavía por los informantes de formación universitaria (13%) (p-valor: 0,002). En los usos de la modalidad epistémica, sin embargo, se observa un patrón curvilíneo, con los hablantes de estudios primarios en cabeza de nuevo (46%), pero seguidos esta vez por los de estudios universitarios (25%)

(p-valor: 0,179). Como se vio anteriormente (véase § 4), una parte considerable de responsabilidad en el comportamiento de estos hablantes de formación superior estriba en la actuación lingüística de las mujeres, que afectadas probablemente por el recuerdo de la norma académica disparan –por ultracorrección– los usos no canónicos de la preposición en proporciones significativamente más elevadas que los hombres del mismo grupo.

Con todo, no es este el único cruce que revela la importancia de estos factores sociolingüísticos, en los que se imbrican tanto el prestigio de las normas preceptivas como una vaga conciencia lingüística acerca de la relevancia social de estas, con especial intensidad en algunos grupos sociales. La tabulación cruzada entre el nivel de instrucción y el estatus social de los hablantes apunta en esa misma dirección. De este modo podría explicarse por qué son sobre todo los hablantes de estudios primarios pertenecientes a la clase media-baja quienes sobresalen por encima de todos los demás en el empleo de la preposición. En efecto, nada menos que un 55% de todas las ocurrencias de la variante corresponden a hablantes en los que se conjugan ambos factores sociológicos, mientras que estos porcentajes son mucho más reducidos en otros subgrupos. La aplicación de nuevo de mecanismos de ultracorrección podría intervenir decisivamente en la actuación lingüística de estos hablantes, impelidos por una norma que “conocen” de una forma todavía más vaga y difusa que otros sectores afectados por similares patrones de inseguridad lingüística, como se vio anteriormente a propósito de las mujeres adultas de clase media-baja, o incluso, de las que poseen formación universitaria (véase más arriba y anteriormente en § 4.2).

4.4 Otros factores sociales

Frente a los reseñados hasta el momento, los factores que se consideran en el presente apartado no aparecen seleccionados como significativos por el análisis estadístico. Sin embargo, las diferencias frecuenciales observadas en su seno ayudan a interpretar algunas divergencias destacadas dentro de otros factores. Este es el caso, por ejemplo, del *lugar de residencia* de los hablantes, que para la presente investigación se ha dividido en dos grupos. Por un lado, el integrado por aquellos informantes que han nacido y/o residido toda o gran parte de su vida en la ciudad de Castellón de la Plana. Por otro, los residentes en las poblaciones de las comarcas valencianohablantes de la provincia de Castellón, de raigambre históricamente rural.⁵ Desde un punto de vista variacionista es de prever, por ejemplo, que los cambios lingüísticos cobren una mayor intensidad en las poblaciones más importantes, y que desde ellas irradien en mayor o menor medida a las poblaciones vecinas, cuyo comportamiento se antoja inicialmente más conservador. En la práctica, algunas investigaciones llevadas a cabo a partir de los materiales del Corpus de Castellón han permitido advertir ya desenlaces sociolingüísticos de esta naturaleza. Así, en el análisis acerca de la expresión variable del futuro verbal en esta región (Blas Arroyo 2008), pudo comprobarse como eran precisamente los hablantes que habían crecido en las comarcas rurales de Castellón quienes se situaban en cabeza del fenómeno de convergencia gramatical que lleva a la notable retención del futuro morfológico (*cantaré*), en claro contraste con las tendencias de cambio más favorables a otras variantes alternativas (v. gr. futuro perifrástico) en buena parte del mundo hispánico (y en el presente caso también, entre los informantes de la capital). Y de forma parecida, la restitución del segmento fonológico /-d-/ en las palabras terminadas en *-ado* (*cantado*, *demasiado*) -un aparente proceso de cambio lingüístico en marcha con un especial protagonismo en tierras

castellonenses, especialmente entre los más jóvenes- cobra particular intensidad entre los hablantes capitalinos (Blas Arroyo 2007).

Estas diferencias se aprecian también en la presente investigación, si bien de forma mucho más limitada (véase tabla 2), de ahí que Goldvarb X no las seleccione como significativas. Por ello, más interesantes son aquí las interacciones entre este factor y algunos de los considerados anteriormente. A este respecto, recuérdese, por ejemplo, como las diferencias generolectales se disparaban en los enclaves rurales, lo que poníamos en relación con la probable mayor presión normativa e inseguridad sentidas por las mujeres que viven en un entorno generalmente más aislado de los focos de irradiación de la norma prescriptiva, como sucede con los pueblos (tabla 4). Del mismo modo, en el cruce con la edad se apreciaba como las diferencias entre los dos tipos de población se neutralizan entre los hablantes menos favorecidos de la variante prepositiva (los jóvenes), mientras que se incrementan considerablemente entre los adultos residentes en la capital, tal vez impelidos por similares presiones normativas a las ya comentadas anteriormente (véase gráfico 6).

Del mismo tenor son, por último, algunas diferencias arrojadas por el factor *estatus social*, para cuya configuración en el presente estudio se han tenido en cuenta diversos parámetros sociales (nivel de instrucción, prestigio de la profesión, tipo de residencia y nivel de ingresos).⁶ Aunque el análisis multivariante tampoco lo selecciona como significativo, las distancias frecuenciales obtenidas señalan la existencia de normas sociolingüísticas complejas en todo el espectro social. A diferencia del nivel de instrucción, cuyos resultados se analizaban anteriormente (véase § 4.3), el patrón distribucional obtenido ahora es curvilíneo, con la clase media-baja como principal avanzadilla en el empleo de la perífrasis prepositiva (26%), seguida de cerca por la clase baja (24%) y con la clase media-alta como principal responsable del descenso en su empleo (17%). Con todo, se ha visto como estas diferencias favorables a la clase media-baja se producían sobre todo en algunos subgrupos sociales caracterizados por una mayor inseguridad lingüística, en fenómenos de variación como el presente en los que la presión de la norma se halla latente, pero cuyo conocimiento objetivo está lejos de cumplirse en la mayoría de los casos. Así ocurría, por ejemplo, entre las mujeres (véase § 4.1) o, en mayor medida, incluso, entre los hablantes de este mismo nivel social con menores niveles de formación académica (véase § 4.3).

INSERT GRÁFICO 8

5. Conclusiones

En las comunidades de habla castellonenses la variación entre las perífrasis modales *deber/deber de* + infinitivo ofrece una notable complejidad sociolectal, ya que da cuenta de la existencia de normas sociolingüísticas en apariencia contradictorias, pero que de consuno actúan en una misma dirección. En el conjunto de la sociedad, el uso de la variante sin preposición (*deber* + infinitivo) se impone claramente, y es especialmente destacada entre grupos sociales muy diversos, como sucede con los hombres en general, pero especialmente entre los que proceden de un enclave rural, los jóvenes de cualquier sexo, residencia y condición social; pero incluso también, entre amplios sectores de las clases más elevadas y con mayores niveles de instrucción. Al mismo tiempo, y por lo que al factor generacional se refiere, la nítida distribución alcanzada entre los diferentes grupos etarios confirma la continuidad de un viejo cambio lingüístico que, favorable a la perífrasis en incidencia directa, se viene gestando desde

antiguo en la historia del español, pese a que en las comunidades españolas parece no haber alcanzado el estadio extremo en el que se sitúan otros lugares del mundo hispánico.

En el desarrollo de este cambio en marcha diversos factores lingüísticos, estilísticos y sociales actúan todavía como un freno para su culminación y, consiguientemente, como un estímulo para la selección de la perífrasis prepositiva (*deber de* + infinitivo). En relación con los factores sociales, que han constituido el objeto de atención del presente trabajo, los datos empíricos de la investigación muestran una distribución sociolectal compleja, ya que en ella se dan cita grupos sociales aparentemente enfrentados en lo que a la recepción de las normas sociolingüísticas de prestigio se refiere. Y es que, por un lado, entre los grupos más favorecedores de la retención se sitúan agregados sociales tales como las clases bajas, los informantes de estudios primarios, o los hablantes de edad más avanzada, en los que tradicionalmente se ha advertido un menor condicionamiento de la actuación lingüística en función de esas normas de prestigio. Con todo, algunos cruces entre estos grupos revelan la más que hipotética relevancia de factores de inseguridad lingüística, conducentes a la aplicación de mecanismos de ultracorrección. Especialmente llamativo es, a este respecto, el grado de empleo de la variante prepositiva entre los hablantes de estudios primarios y entre los de clase media-baja –y más aún, entre quienes combinan ambos factores–, cuyos índices figuran entre los pocos de toda la muestra en los que dicha construcción supera a la variante en incidencia directa.

En relación con hechos similares Galué et al. (2004: 74) han señalado que “el desconocimiento de la norma existe en todos los niveles socioeconómicos y son los hablantes de nivel bajo quienes, bajo la creencia de que usan la norma más prestigiosa y correcta, insertan la preposición *de* en donde no se requiere”. En las comarcas castellonenses ello es así especialmente entre hablantes situados en una posición particularmente inestable dentro de la escala social, y en los que diversos estudios sociolingüísticos han revelado mayores niveles de inseguridad lingüística a la hora de enfrentarse a la norma preceptiva. Por limitarnos al español, este es el caso, por ejemplo, de fenómenos como la variación morfológica *-ra/se* en las terminaciones del imperfecto de subjuntivo en épocas pretéritas de la lengua (Martínez 2001) o más recientemente el (de)queísmo en diversas comunidades de habla del mundo hispánico (Rabanales 1977, Bentivoglio 1980-1981, Boretti de Macchia 1989, Serrano 1998). En el presente estudio, este patrón distribucional se ha observado entre hablantes de estudios primarios y clases medias-bajas, pero también entre mujeres adultas de esa misma condición social, o entre las residentes en enclaves rurales, donde el poder irradiador de los preceptos académicos suele llegar, por lo general, más debilitado que a los urbanos.

Ahora bien, junto a estos –y quizá sea aquí donde se aprecien los resultados más inesperados– en el estudio se han encontrado también seguidores de esos mismos patrones de distribución sociolingüística en otros extremos de la pirámide social. Este es el caso, sobre todo, de las mujeres con estudios superiores, en las que se advierte un seguimiento significativamente mayor de los usos canónicos reservados por la tradición normativa a la perífrasis con preposición en contextos epistémicos; pero también, y aunque en menor medida, en los no canónicos (deónticos), lo que de nuevo haya que interpretar quizá en términos de ultracorrección.

A diferencia de los anteriores, este último perfil sociolingüístico difiere, pues, cualitativamente del que presentan otros fenómenos sintácticos, como el mencionado dequeísmo. Si en los estudios sobre este fenómeno son casi siempre las clases medias-bajas quienes sobresalen por el empleo de la variante dequeísta, en el presente caso se

obtienen también índices superiores a la media en grupos más elevados socioculturalmente. La diferencia reside probablemente en el hecho de que, frente al dequeísmo, cuya saliencia normativa es particularmente elevada incluso fuera de los círculos normativos y académicos –lo que explicaría en buena medida su limitado alcance, al menos en el contexto peninsular–, la tradicional oposición funcional entre las perífrasis con y sin proposición para la expresión de contenidos modales diferentes es mucho más vaga y compleja, y en el fondo, desconocida por la mayoría de hablantes. Los más sensibles al prestigio de las variantes lingüísticas son potencialmente conscientes de normas aprendidas en la escuela, pero en muchos casos han “olvidado” cuál es su alcance exacto (si es que alguna vez lo supieron con exactitud). Como señala Gómez Torrego (1989: 188): "pocos hablantes conocen la diferencia entre estas dos construcciones perifrásticas, por lo que suelen emplearlas indistintamente [...]. Hay que hacer un esfuerzo para mantener esta diferencia, pues así se evitará incurrir en posibles ambigüedades". Un esfuerzo, cabría añadir, que para la mayoría está lejos de sus posibilidades. A ello contribuye no solo la menor recurrencia de los fenómenos de variabilidad sintáctica en el discurso, menos proclives por ello que los fonológicos a poseer significación sociolectal, sino también la misma complejidad de la distinción normativa, dificultada por la permeabilidad de la frontera entre los valores modales. Y es que, como recuerda Fernández de Castro (1999: 181): “hay razones ‘internas’ por las que el español se negó siempre a aprovechar la oportunidad que le brindaban los gramáticos de popularizar *la tantas veces sutilísima distinción entre lo epistémico y lo dinámico* (la cursiva es nuestra)”. Al final, este “desconocimiento” de la norma académica permitiría explicar por qué los mismos hablantes que superan a otros en los usos canónicos de la variante prepositiva (contenidos epistémicos) lo hagan a menudo también en los no canónicos (contenidos deónticos).

Por otro lado, estos hechos permiten justificar el hecho de que la distinción modal epistémico-deóntica no sea seleccionada como significativa por el análisis estadístico multivariante. Su vago recuerdo en la conciencia lingüística de algunos hablantes no es suficiente para establecer patrones de diferenciación sociolingüística nítidos que (des)favorezcan una u otra variante. De ahí que dicho factor se vea claramente aventajado por otros de naturaleza lingüística y estilística, cuya incidencia a favor (énfasis expresivo, espontaneidad de las interacciones verbales) o en contra (recursos de atenuación, modalidad interrogativa) de la variante prepositiva resultan mucho más determinantes (Blas Arroyo, en prensa). Y ello tanto en el conjunto de la muestra como en el comportamiento de diversos grupos sociales. A este respecto, en estas páginas se ha visto como, incluso aquellos grupos menos favorecedores de la variante, como los jóvenes, ven incrementados considerablemente sus índices en los contextos enfáticos, mientras que, como contrapartida, casi desaparecen en los atenuados.

En otro orden de cosas, el estudio pone de manifiesto la notable variabilidad social, y las consiguientes dificultades para trasladar los resultados obtenidos en otros estudios sobre la norma culta o los sociolectos más elevados, al conjunto de la sociedad. Un resultado ya advertido al comparar los índices de otras investigaciones. Así, por ejemplo, el reducido porcentaje alcanzado por la perífrasis en incidencia directa (*deber* + infinitivo) para la expresión de la modalidad epistémica (15.9%) en las encuestas de la norma culta de Caracas (Samper et al. 1998-1999), difiere considerablemente del que obtienen Galué et al. (2004) en el conjunto de esa misma comunidad de habla (48%). En Castellón, sin embargo, aunque no se dispone de trabajos exclusivos sobre el sociolecto alto, los resultados coinciden tan solo parcialmente con los que arrojan otros estudios sobre comunidades de habla españolas. Si bien las cifras generales de retención de la variante prepositiva son comparables a las de esas investigaciones, no lo son, por el

contrario, otros resultados parciales relacionados con la distribución de las variables sociales analizadas.

A estas dificultades para establecer comparaciones fiables no son tampoco ajenos algunos condicionantes metodológicos, como las diferencias muestrales –a veces muy importantes– entre unas investigaciones y otras, o el empleo de técnicas de análisis estadístico diferentes. En la práctica, la mera lectura de porcentajes y grupos sociales aislados ofrece resultados incompletos y, en ocasiones, sencillamente engañosos, sobre todo cuando las muestras son reducidas. Por el contrario, una radiografía sociolingüística ajustada sobre cualquier fenómeno de variación lingüística implica el empleo de técnicas de análisis inferencial. Mediante el auxilio de programas multivariantes como Goldvarb X y otras pruebas de significación estadística el analista puede aproximarse no solo a la significación real de los factores considerados, sino también al posible hallazgo de diferencias importantes en el seno de algunos grupos, condicionadas por la interacción con otros factores.

Referencias

- Alarcos, Emilio. (1994). *Gramática de la lengua española*. Madrid: Espasa-Calpe. Print.
- Alcina, Juan y José Manuel Bleca. (1989). *Gramática española*. Barcelona: Ariel. Print.
- Balash, Sonia. (2008). “*Debe (de) ser*: evolución de la variación”. *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics*. Ed. Maurice Westmoreland y Juan Antonio Thomas. Somerville: Cascadilla. 109-119. Print.
- Bentivoglio, Paola. (1980-1). “El dequeísmo en Venezuela: ¿un caso de ultracorrección?”. *Boletín de Filología de la Universidad de Chile* 31: 705–19. Print.
- Blas Arroyo. (2005). *Sociolingüística del español. Desarrollos y perspectivas en el estudio de la lengua en contexto social*. Madrid: Cátedra. Print
- . (2007). “Competing norms in a bilingual community. Ethnolinguistic and social factors in the reversal of a change in progress in peninsular Spanish”. *Sociolinguistic Studies* 1.3: 37-68. Print.
- . (2008). “The variable expression of future tense in Peninsular Spanish: The present (and future) of inflectional forms in the Spanish spoken in a bilingual region”. *Language Variation and Change* 20.1: 85-126. Print.
- . (coord.). (2009). *Corpus sociolingüístico de Castellón de la Plana y su área metropolitana*. Castellón: Universitat Jaume I. Print.
- . (en prensa). “Deber (de) + infinitivo: ¿un caso de variación libre en español? Factores determinantes en un fenómeno de alternancia sintáctica”, *Revista de Filología Española* 90.2 (en prensa).
- Boretti de Macchia, Silvia. (1989). “(De)queísmo en el habla culta de Rosario”. *Anuario de Lingüística Hispánica* 5: 27–48. Print.
- Bybee, Joan, Revere Perkins y William Pagliuca. (1994). *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press. Print.
- Cuervo, Rufino José et al. (1886-1893). *Diccionario de construcción y régimen de la lengua castellana*. Bogotá: Instituto Caro y Cuervo (edición de 1994). Print.
- Díez de Revenga, Pilar y Dolores A. Igualada (1989-1990). “Construcciones modales en documentos notariales del siglo XIII”, *Glossae* 2: 119-132. Print.
- Fairclough, Marta. (2000). “Expresiones de modalidad en una situación de contacto: *deber (de) vs. tener que* en el español hablado en Houston”. *Southwest Journal of Linguistics* 19.2: 19-30. Print.
- Fernández de Castro, Francisco. (1999). *Las perífrasis verbales en el español actual*. Madrid: Gredos. Print.
- Fernández Ulloa, Teresa. (2000). “Perífrasis verbales en el castellano en el castellano de Bermeo (Vizcaya)”. *Revista Española de Lingüística* 30.1: 1-34.
- Galué, Daisy, Marcos Suárez y Betsabé Valor (2004). “¿Debe estar ahí... o debe de estar ahí...? Estudio de las perífrasis con el verbo *deber* en el español hablado en Caracas”. *Núcleo* (nº especial): 53-77. Print.
- García Fernández, Luis. (dir.). (2006). *Diccionario de perífrasis verbales*. Madrid: Gredos. Print.
- García, Erica. (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*. Ed. José Moreno de Alba. México: UNAM. 46-65. Print.
- Gómez Manzano, Pilar. (1992). *Perífrasis verbales con infinitivo (valores y usos en la lengua hablada)*. Madrid: UNED. Print.
- Gómez Molina, José Ramón. (2009). “Estudio sociolingüístico de la perífrasis *deber* + infinitivo y *deber*

- de + infinitivo en el español hablado”. *Lengua viva. Estudios ofrecidos a César Hernández Alonso*, Ed. Álvarez Tejedor, Antonio. Valladolid: Universidad de Valladolid. 867-880. Print.
- Gómez Torrego, Leonardo. (1988). *Perífrasis verbales. Sintaxis, semántica y estilística*. Madrid: Arco Libros. Print.
- . (1989). *Manual de español correcto. Vol. II*. Madrid: Arco Libros. Print.
- . (1999). “Los verbos auxiliares. Las perífrasis verbales de infinitivo”. *Gramática descriptiva de la lengua española*. Ignacio Bosque y Violeta Demonte (dir.). Madrid: Espasa-Calpe. 3223-3390. Print.
- Hernández Alonso, César. (1984). *Gramática funcional del español*. Madrid: Gredos. Print.
- Luna, Elizabeth. (1980). *Sintaxis de los verboides en el habla culta de la Ciudad de México*. México: UNAM. Print.
- Martínez, Glenn. (2001). “Política lingüística y contacto social en el español mexicano-tejano: la oposición -ra y -se en Tejas durante el siglo XIX”. *Hispania* 84. 1: 114-124. Print.
- Mendieta, Salvador. (1993). *Manual de estilo de TVE*. Barcelona: Labor. Print.
- Rabanales, Ambrosio. (1977). “Queísmo y dequeísmo en el español de Chile”. *Estudios sobre el español hablado en las principales ciudades de América*. Ed. Juan Manuel Lope Blanch. México: UNAM. 541-569. Print.
- Real Academia Española (1726-1739). *Diccionario de Autoridades*. Madrid: Gredos (Edición de 1990). Print.
- . (1973). *Esbozo de una nueva gramática de la española*. Madrid: Espasa-Calpe. Print.
- Real Academia Española y Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española*. Madrid: Espasa-Calpe. Print.
- Samper, José Antonio, Clara Eugenia Hernández y Magnolia Troya. (1998). *Macro-corpus de la norma lingüística culta de las principales ciudades del mundo hispánico*. Las Palmas de Gran Canaria: Universidad de las Palmas de Gran Canaria y ALFAL. Print.
- . (1998-1999). “El uso de deber + infinitivo y deber de + infinitivo en la norma lingüística culta de América y España”. *Boletín de Filología (Homenaje a Ambrosio Rabanales)* 37: 1059-1083. Print.
- Sankoff, David, Sali A. Tagliamonte y Eric Smith. (2005). *GOLDVARB X: A Multivariate Analysis Application for Macintosh and Windows*. Department of Linguistics: University of Toronto and Department of Mathematics: University of Ottawa. Web.
- Seco, Manuel. (1986). *Diccionario de dudas y dificultades de la lengua española*. Madrid: Calpe. Print.
- Seco, Rafael. (1967). *Manual de gramática española*. Madrid: Aguilar. 9ª edición (1ª ed. 1930).
- Serrano, María José. (1998). “Estudio sociolingüístico de una variante sintáctica: el fenómeno dequeísmo en el español canario”. *Hispania* 81.2: 392-405. Print.
- Silva-Corvalán, Carmen. (1995). “Contextual conditions for the interpretation of poder and deber”. *Modality in Grammar and Discourse*. Ed. Joan Bybee y Suzanne Fleishman. Amsterdam: John Benjamins. 68-105. Print.
- Tagliamonte, Sali A. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press. Print.
- Troya, Magnolia. (1998). *Perífrasis verbales de infinitivo en la norma lingüística culta de Las Palmas de Gran Canaria*. Madrid: Real Academia Española y Universidad Palmas de Gran Canaria (Anejo LVI del BRAE). Print.
- Valádez, Carmen. (1969). *Perífrasis de infinitivo en la norma lingüística culta de la ciudad de México*. México: UNAM. Print.
- Yllera, Alicia. (1980). *Sintaxis histórica del verbo español: Las perífrasis medievales*. Zaragoza: Universidad de Zaragoza. Print.

*El presente estudio forma parte de un proyecto de investigación financiado por el Ministerio de Ciencia e innovación español (referencia: FFI2010-15280).

¹ En el presente estudio no se plantea la tradicional polémica acerca del carácter perifrástico de las combinaciones con el verbo *deber* + infinitivo (sin preposición). No obstante, y al igual que otros lingüistas recientemente, considero que dicha combinación cumple escrupulosamente con los criterios sintácticos utilizados para la delimitación de las perífrasis de infinitivo: transformaciones (pasiva total e interrogativa con *qué*), imposibilidad de transformar la perífrasis en una construcción ecuacional o la conmutación del infinitivo por un elemento equivalente funcionalmente, dado que no se trata de un complemento del verbo auxiliar (para el desarrollo y ejemplificación de estos y otros argumentos en la misma línea, véanse Troya 1998; Gómez Torrego 1988, 1999; Gómez Manzano 1992, Fernández de Castro 1999, García Fernández et al. 2006).

² Entre paréntesis aparecen los porcentajes de ocurrencias totales de la variable correspondientes a cada grupo social. Aunque la muestra está integrada por cuotas similares de sexo, edad y nivel de instrucción, no siempre estas se corresponden con las cifras obtenidas. Ello es especialmente llamativo en el caso del

nivel de instrucción, donde los hablantes de estudios primarios han empleado considerablemente menos estas perífrasis de infinitivo –en favor de otras como *tener que* o *haber que/de*- que los de otros niveles educativos.

³ Ahora bien, estas cifras dan cuenta exclusivamente de las proporciones alcanzadas en términos absolutos por la variante prepositiva, sin relación con la otra variante y contrariamente, pues, a lo que se ha hecho aquí. Aplicado este mismo criterio al presente caso, veríamos como en Castellón las mujeres superan a los hombres por 58% a 42%.

⁴ En España, el poder irradiador de la norma, tanto en el sentido académico como en el de las nuevas variantes de prestigio, ha circulado tradicionalmente desde los enclaves urbanos a los rurales, si bien esta dirección ha podido experimentar algunos cambios en los últimos tiempos como consecuencia de la globalización.

⁵ Pese a los notables cambios demográficos y urbanísticos –catastróficos, a menudo- experimentados en las últimas décadas, el carácter básicamente no urbano de los pueblos que rodean Castellón de la Plana se sigue apreciando en buena medida, incluso entre los sectores más jóvenes. Los lazos afectivos de estos habitantes con sus poblaciones de origen son especialmente intensos, lo que se demuestra en multitud de comportamientos individuales y colectivos: desde el regreso al pueblo a la menor oportunidad que se presenta, a la particular intensidad con que se viven las -ubicuas- fiestas patronales o de cualquier otro tipo, pasando -desde el punto de vista sociolingüístico-, por la considerable mayor presencia del valenciano en los hábitos comunicativos cotidianos de sus hablantes que entre los vecinos de la capital.

⁶ Con todo, no siempre ha sido posible contar con este último factor, por los recelos de algunos informantes a mostrar explícitamente su estatus socioeconómico. Estas dificultades, junto a otras de diferente entidad, hacían verdaderamente complicada la tarea de clasificar a algunos individuos en una determinada categoría social, por lo que, en estos casos, se decidió no codificarlos en función del presente factor. De este modo, las 412 ocurrencias de la variable en el conjunto de la muestra, quedan reducidas a 387 en la distribución a partir del estatus social.