



Algunas contribuciones de los corpus de la red PRESEEA al estudio del español en contacto con el catalán

Some contributions of the PRESEEA network corpora to the study of Spanish in contact with Catalan

Andrés Enrique-Arias
Universitat de les Illes Balears

Abstract

In this article we illustrate the possibilities of using the corpora of the PRESEEA Network to study the effects of Catalan contact on Spanish. This is achieved by contrasting data from Alcalá de Henares, Valencia and Palma de Mallorca among other communities. The phenomena considered are the variation in the use of the morphological futures cantaré versus the periphrastic one voy a cantar both meaning 'I will sing', and the distribution of the obligation periphrases haber de and tener que. The contrastive analysis reveals some effects of the influence of Catalan in bilingual communities: in the case of the futures, there is greater retention of the morphological future cantaré; as for the obligation periphrases, the influence of Catalan results in a lower degree of differentiation of the functions of deontic and epistemic modality in the distribution of haber de and tener que.

Keywords: variation, change, Spanish in contact with Catalan, future tense forms, deontic periphrases.

Resumen

En este artículo ilustramos las posibilidades del empleo de los corpus de la Red PRESEEA para estudiar los efectos del contacto con el catalán en el castellano a partir de estudios que contrastan datos de Alcalá de Henares, Valencia y Palma de Mallorca entre otras comunidades. Los fenómenos considerados son la variación en el uso de los futuros (cantaré frente a voy a cantar) y en la distribución de las perífrasis de obligación haber de y tener que. El análisis contrastivo revela algunos efectos de la influencia del catalán en las comunidades bilingües: en el caso de los futuros, se da una mayor retención del futuro morfológico cantaré; respecto de las perífrasis de obligación, la influencia del catalán se traduce en un menor grado de diferenciación de las funciones de modalidad deontica y epistémica en la distribución de haber de y tener que.

Palabras clave: variación, cambio, español en contacto con el catalán, expresión del futuro, perífrasis deonticas.

Citar como: Enrique-Arias, Andrés (2024). Algunas contribuciones de los corpus de la red PRESEEA al estudio del español en contacto con el catalán. *Normas*, 14(1), 88-100, doi: 10.7203/Normas.v14i1.28649.

1. Introducción¹

Un aspecto de la variación dialectal y sociolingüística del español peninsular que ha recibido una atención relativamente menor es el estudio del castellano hablado en territorios de habla gallega, vasca y catalana. Las variedades de español habladas en estas situaciones de contacto solo empezaron a constituir un asunto de interés teórico en las últimas décadas del siglo XX.

¹Esta investigación se enmarca en el proyecto PID2020-116863GB-I00 financiado por MCIN/ AEI /10.13039/501100011033/ FEDER "Una manera de hacer Europa"; asimismo la investigación ha sido parcialmente financiada por el CAIB a través de la Direcció General de Política Universitària i Recerca con fondos de la Ley de Estancias Turísticas ITS2017-006 en el marco del PDR2020/51.

A pesar de este comienzo tardío, contamos ya a día de hoy con un buen número de trabajos de corte sociolingüístico que analizan aspectos del castellano producido en comunidades bilingües peninsulares. Esta nueva ola de estudios superó por fin la idea de que el castellano hablado en Cataluña, Baleares o el País Vasco, por mencionar unos ejemplos, son anomalías; bien al contrario, existe ya un consenso de que nos encontramos ante variedades del español como puedan ser las habladas en Lima, Sevilla o Buenos Aires, con la única particularidad de que, en las áreas bilingües de España, un amplio repertorio de rasgos de interferencia de la lengua en contacto ha pasado a formar parte de la variedad nativa castellana de la zona (Moreno 2020: 35-37).

En las páginas que siguen me propongo hacer una serie de reflexiones metodológicas sobre el estudio de las variedades de contacto del español a partir del análisis de datos de Palma (Islas Baleares, España), una comunidad de habla que se caracteriza precisamente por presentar una situación de bilingüismo de sociedad de castellano y catalán desde hace varias generaciones. Para el estudio de la lengua producida en esta comunidad, el corpus PRESEEA de Palma (www.corpusmallorca.es/preseea), publicado en línea desde 2018, es un recurso especialmente valioso. Por un lado, en este corpus los hablantes han sido clasificados de acuerdo con su nivel de bilingüismo a partir de un cuestionario sobre el aprendizaje, uso y conocimiento del catalán y el castellano, distinguiendo entre dominantes en catalán, en castellano y bilingües equilibrados; por consiguiente, se puede medir el efecto de la influencia del catalán en el castellano de acuerdo con la lengua dominante. Por otro lado, la existencia de otros corpus comparables de la red PRESEEA permite contrastar la variedad de español de la comunidad bilingüe de Palma con la de otras comunidades en las que el catalán es desconocido o en las que se producen situaciones de bilingüismo diferentes a las de Palma.

El objetivo de este artículo es, por tanto, explorar cómo se pueden estudiar los efectos del contacto de lenguas—castellano y catalán en este caso—mediante el análisis contrastivo de corpus de la red PRESEEA. Más concretamente me centraré en dos fenómenos variables a partir de investigaciones recientes que han empleado datos de los corpus de la Red. El primer estudio concierne la variación en el uso del futuro sintético *cantaré* frente al perifrástico *voy a cantar* empleados para referencia temporal futura. En esta investigación, realizada en colaboración con Beatriz Méndez Guerrero, se han considerado los corpus PRESEEA de Palma y de Alcalá de Henares (Enrique-Arias y Méndez Guerrero 2020). La segunda investigación es un estudio preliminar que estoy desarrollando con Malte Rosemeyer sobre la distribución de las perífrasis de obligación *haber de* y *tener que* en los corpus PRESEEA de Palma y de Valencia (Rosemeyer y Enrique-Arias 2023).

2. Corpus y metodología

Como ya se ha mencionado más arriba, el objetivo de esta investigación es explorar las posibilidades metodológicas de la comparación de corpus de la red PRESEEA compilados en zonas monolingües y en situaciones con diferentes grados de bilingüismo para estudiar los efectos del contacto lingüístico. Una de las ventajas de la Red es la posibilidad de contrastar datos de corpus comparables, pues en todos ellos se han utilizado criterios idénticos para la selección de los informantes mediante un muestreo basado en cuotas de sexo, edad y nivel de educación, siguiendo la metodología diseñada para el proyecto PRESEEA (<https://preseea.uah.es/metodologia>). Además, debido a lo previsto en la metodología del

proyecto, las entrevistas tienen la misma duración aproximada, de unos 45 minutos, y contienen los mismos módulos temáticos, lo cual contribuye a la comparabilidad de las muestras de cada ciudad. Un aspecto especialmente útil a efectos de la investigación sobre la expresión del futuro es que algunos de los temas de la entrevista facilitaban la referencia a diferentes grados de distancia temporal en el futuro, como puedan ser planes para el fin de semana, para vacaciones, proyectos de vida o predicciones sobre el clima en el contexto del calentamiento global.

Asimismo, y de acuerdo con el objetivo de comparar comunidades con diferentes grados de intensidad de bilingüismo con el catalán, hemos considerado cuatro corpus de entrevistas sociolingüísticas con datos de Alcalá de Henares, de Valencia, de la ciudad de Palma y de la Mallorca rural. Los datos de Alcalá sirven como corpus de control al tratarse de una comunidad castellana monolingüe en la que no se habla catalán. Por lo que respecta a las otras comunidades, los datos sociodemográficos más recientes nos permiten distinguir diferentes grados de intensidad del contacto con el catalán (cf. Generalitat Valenciana 2021 para Valencia y Melià y Vanrell 2018 para las Baleares). Así pues, el porcentaje de la población que declara que puede hablar catalán bien o muy bien en la Comunidad Valenciana es un 50,6% frente a un 80,5% en las Islas Baleares. Por lo que respecta a la densidad de hablantes dominantes en catalán, el porcentaje de la población que utiliza esa lengua de forma predominante en su día a día es un 18,9% en la ciudad de Valencia, un 27,9% en la ciudad de Palma y un 45,5% en la Mallorca rural.

Para el estudio de las formas de futuro utilizamos, para representar el castellano monolingüe, el corpus PRESEEA de Alcalá de Henares, que contiene 54 entrevistas que en su mayoría fueron recogidas en 1998, si bien algunas datan de 1991. Junto a este corpus empleamos como muestra de castellano en contacto con el catalán el Corpus PRESEEA Palma, que reúne 54 entrevistas sociolingüísticas realizadas entre 2007 y 2010. Además, como ya se ha señalado, en este corpus los hablantes se clasificaron de acuerdo con su nivel de bilingüismo, distinguiendo entre dominantes en catalán, en castellano y bilingües equilibrados.

Para el estudio de las perífrasis de obligación consideramos, además del corpus de Palma, muestras de la ciudad de Valencia y de la Mallorca rural. Para el corpus de Valencia, que sirve para representar la comunidad con menor densidad de hablantes de catalán, se ha tenido en cuenta la muestra de 18 entrevistas del corpus PRESEEA de esa ciudad publicadas en la web de la Red. Las entrevistas fueron recogidas entre 1996 y 2006. Para los datos de la comunidad de Palma, en la que se daría una presencia de catalán relativamente más intensa, hemos recurrido al ya descrito corpus PRESEEA de Palma. Y, en tercer lugar, hemos utilizado como corpus de control para estudiar el habla producida en un contexto de mayor densidad de hablantes dominantes en catalán las entrevistas del Corpus Oral y Sonoro del Español Rural (COSER) (www.corpusrural.es). A diferencia de las entrevistas de la Red PRESEEA, que representan la lengua de zonas urbanas, el Corpus COSER contiene entrevistas recogidas en localidades rurales entre personas de avanzada edad dedicadas al sector primario, con pocos años de escolarización y escasa movilidad fuera de su lugar de origen. En el caso de Mallorca, el corpus consta de 15 entrevistas a personas mayores (86,5 años de media), hombres y mujeres, dominantes en catalán. La recogida de datos se produjo en mayo de 2017. Por sus características, no se trata de un corpus directamente comparable con los de la red PRESEEA; en este estudio incluimos un corpus rural para contrastar los resultados de las áreas urbanas, en las que existe la presencia de individuos procedentes de la inmigración de

zonas castellanohablantes, con los de individuos con un perfil que presenta un mayor grado de transferencia de rasgos del catalán.

En la Tabla 1 a continuación se incluyen a modo de resumen las principales características de los cuatro corpus considerados:

Corpus	Fecha	Tipo de comunidad	Número de entrevistas	Metodología
Alcalá	1991-1998	Monolingüe castellano	54	PRESEEA
Valencia	1996-2006	Bilingüe	18	PRESEEA
Palma	2007-2010	Bilingüe	54	PRESEEA
Mallorca rural	2017	Bilingüe dominante en catalán	15	COSER

Tabla 1: Características de los corpus empleados

Si bien son claras las ventajas del estudio contrastivo de los tres corpus PRESEEA (son corpus con idéntica composición) también es cierto que la comparabilidad total es un ideal que difícilmente se llega a cumplir. La diferencia más evidente es la fecha de la compilación de las entrevistas, entre 1991 y 2010. De todos modos, la gran mayoría de las entrevistas han sido recogidas en los once años que van de 1996 a 2007, un intervalo de una década que no compromete la comparabilidad de las muestras. Por lo que atañe a las entrevistas del COSER, nuestra intención no es una muestra estratificada por edad, sexo y nivel socioeconómico comparable sino un corpus de control con hablantes en los que se dé el máximo posible de interferencia con el catalán.

3. Expresión del futuro

En el español actual encontramos dos formas de futuro: el futuro morfológico *cantaré* (FM) y el futuro perifrástico *voy a cantar* (FP). La forma más antigua, el FM, deriva de la perífrasis latina *infinitivo + HABERE* (AMARE HABEO > *amar-é*); la forma más nueva, el FP, es una innovación románica creada a partir de una perífrasis con *ir* que ha ido ganando terreno progresivamente desde al menos el siglo XVII (Aaron 2010, Melis 2006). En su mayor parte, la literatura sobre los futuros españoles se ha centrado en el problema de la diferencia de significado entre el FM y el FP cuando compiten en los mismos contextos de ocurrencia. Los dos factores lingüísticos que han resultado sistemáticamente más relevantes para explicar la variación entre las dos formas son la distancia temporal (el FP se ve favorecido para expresar acciones que tienen lugar en un intervalo de tiempo más cercano al momento de la enunciación, mientras que el MF se correlaciona con mayor distancia temporal del evento futuro) y la asertividad (el FP se da con mayor frecuencia cuando existe certeza epistémica por parte del hablante mientras que el FM revela incertidumbre epistémica). En cuanto a las variables extralingüísticas, se ha observado que la formalidad de la situación también influye en la elección de las formas de futuro, de modo que el FM es más predominante en los textos escritos y registros formales (Sedano 1994: 238).

Por lo que respecta a la frecuencia relativa de las dos formas, numerosos estudios han observado que el FM está siendo progresivamente desplazado por el FP hasta el punto de que en

variedades del español de América ha disminuido considerablemente o incluso ha desaparecido por completo (para un estado de la cuestión actualizado véase Enrique-Arias y Gomila Albal 2023). Sin embargo, en las zonas donde el español coexiste con el catalán este proceso de sustitución está menos avanzado. Esta ralentización de una tendencia de cambio que es prácticamente panhispánica se explicaría por la asimetría estructural que se da entre las dos lenguas: la tendencia histórica a remplazar el FM por el perifrástico con *ir* se ve contrarrestada en los bilingües en catalán por el hecho de que, mientras la variante tradicional tiene un paralelo en el futuro catalán *amaré*, la forma perifrástica innovadora no tiene un equivalente en catalán. De hecho, la perífrasis con *anar* ‘ir’ en catalán tiene un significado de pretérito indefinido (*vaig cantar* ‘canté’). Se trata por tanto de un caso en el que el contacto de lenguas tiene como efecto la inhibición de un cambio lingüístico (Enrique-Arias 2010: 104-105). Blas Arroyo (2007) ha demostrado este efecto inhibitor del contacto lingüístico en el empleo de los futuros en el español de Castellón. En su estudio, en el que analiza datos de hablantes bilingües, los individuos con mayor dominio del catalán presentan sistemáticamente un mayor porcentaje de uso del FM en comparación con los hablantes dominantes en castellano, es decir, el contacto de lenguas en esta situación promueve la ralentización de una tendencia de cambio del español general.

Para dar un mayor apoyo empírico a la hipótesis del efecto ralentizador del contacto con el catalán en la extensión del FP, sería necesario contrastar datos de corpus comparables de español producidos en contextos monolingües con los producidos en situaciones de bilingüismo con el catalán. Con este objetivo en mente, Enrique-Arias y Méndez Guerrero (2018) aprovecharon la disponibilidad de los corpus de la red PRESEEA para comparar las frecuencias de las formas de tiempo futuro en el entorno bilingüe catalán-castellano de Palma en contraste con la ciudad de Alcalá de Henares donde no se habla catalán. Por tratarse de corpus con una composición semejante en lo que respecta al perfil sociolingüístico de los hablantes y la metodología de las entrevistas, no es extraño que los ejemplos de futuros encontrados en los dos corpus y sus contextos de ocurrencia sean comparables. Por ejemplo, con el fin de comprobar la relevancia de la distancia temporal en la distribución de las formas de futuro en nuestros datos, codificamos este factor para cada ocurrencia en los dos corpus según cuatro categorías: distancia temporal inmediata, intermedia, máxima e indefinida. Como se ilustra en los ejemplos a continuación, es posible encontrar ocurrencias de las cuatro categorías consideradas tanto en el corpus de Alcalá como en el de Palma, lo cual confirma su comparabilidad. Los eventos inmediatos fueron aquellos que iban a tener lugar inmediatamente o dentro del mismo día, como en el ejemplo (1). La categoría intermedia comprendía acciones que se esperaba que tuvieran lugar en los siete días (es decir, una semana) siguientes al momento de la enunciación, como en (2). Un evento que iba a tener lugar en un momento futuro más distante que una semana se codificó como distancia máxima, como en (3). Finalmente, cualquier evento cuyo momento de realización no pudiera determinarse, como en (4), se codificó como indefinido:

- (1) a. *en casa de mis padres, están comiendo allí; ahora iré a recogerlos* [Alcalá]
 b. *¿me vas a preguntar por qué?* [Palma]
- (2) a. *estoy estudiando pero bueno, este martes empezaré a trabajar* [Alcalá]
 b. *bueno, los va a cumplir el día 1* [Palma]
- (3) a. *yo no sé si podré tener vacaciones en Semana Santa todavía* [Palma]

b. *me parece que no va a ser acabar la carrera y encontrar trabajo inmediatamente* [Alcalá]

(4) a. *y si sale la opción de poder ser bombero, me voy a a hacer bombero* [Palma]

b. *me voy a morir antes de todo lo que tengo que hacer* [Alcalá]

La distribución global de los futuros, resumida en la Tabla 2, demuestra que el FM se conserva de manera considerablemente más robusta en Palma (53,6%) en comparación con Alcalá (23,2%).

	Palma	Alcalá
FM	53,6% (448)	23,2% (149)
FP	46,4% (388)	76,8% (493)
Total	100% (836)	100% (642)

X^2 : 137.94; <.000

Tabla 2. Distribución de formas de futuro en Palma y Alcalá

Los resultados de la comparación de Palma y Alcalá están en línea con lo observado en otros estudios de comunidades del español peninsular (para la fecha y procedencia de los datos véase Enrique-Arias y Gomila Albal 2023: 180-83). La frecuencia del FM en Alcalá (23,2%) es muy cercana a la de otras comunidades monolingües de la Península, como el 24,8% que registra Osborne (2008) en Puente Genil (Córdoba). Al mismo tiempo la frecuencia registrada en Palma (53,6%) es prácticamente igual a la de otros estudios hechos en áreas catalanohablantes, como es el caso del corpus PRESEEA de Valencia (55,4%; cf. Buzón García 2013), o el 55,4% de Castellón (Blas Arroyo 2008). Extraordinariamente alto es, sin duda, el porcentaje de 85.4% que alcanzan los resultados de los alumnos de 6º curso (11- 12 años de edad) investigados en las localidades de la provincia de Barcelona Mataró y Manlleu (Illamola 2013).

La influencia del catalán en la retención del FM se ve confirmada también por la distribución de futuros en los hablantes de Palma clasificados según la lengua dominante (ver Tabla 3). Los bilingües dominantes en castellano son los que menos usan el FM (42,2 %); no obstante el porcentaje es muy superior al registrado en Alcalá. Ello indica que, aun siendo castellanohablantes, el hecho de vivir en una comunidad catalanohablante influye en la frecuencia de uso del FM. Por su parte, y como era de esperar, los dominantes en catalán (61 %) son los que más emplean el FM, mientras que los bilingües equilibrados se sitúan en un punto intermedio (55,6 %).

	Castellano	Ambas	Catalán	TOTAL
FM	42,2% (106)	55,6% (154)	61,0% (188)	53,6% (448)
FP	57,8% (145)	44,4% (123)	39,0% (120)	46,4% (388)
TOTAL	100% (251)	100% (277)	100% (308)	100% (836)

$x^2=20.341$; $p < .000$

Tabla 3: Distribución de las formas de futuro en Palma de acuerdo con lengua dominante

Como hemos explicado anteriormente, el declive del FM y su reemplazo por el FP es una tendencia histórica que se remonta al siglo XVII. Si esta tendencia de cambio aún continúa en las dos comunidades consideradas, entonces es lógico que la distribución de las dos variantes sea sensible a variables sociolingüísticas como la edad, el género o el nivel de instrucción. Los resultados muestran que, si bien esto es cierto para los datos de Palma, ninguno de estos factores alcanzó significatividad estadística en el grupo de Alcalá.

	Palma	Alcalá
Edad		
18-55 años	50.8% (286/563)	24.6% (103/418)
56 años o más	59.3% (162/273)	20.5% (46/224)
Palma ($\chi^2=5,050$; $p < .024$); Alcalá ($\chi^2= 1,380$; $p < .501$)		
Género		
Mujeres	49.5% (199/402)	22.4% (58/258)
Hombres	57.4% (249/434)	23.6% (91/384)
Palma ($\chi^2=5,198$; $p < .023$); Alcalá ($\chi^2=0,128$; $p < .720$)		
Nivel educativo		
Primaria	50.4% (136/270)	27.3% (50/183)
Secundaria	49.6% (129/260)	24.8% (60/241)
Post-secundaria	59.8% (183/306)	17.8% (39/218)
Palma ($\chi^2=7,527$; $p < .023$); Alcalá ($\chi^2=5,583$; $p < .061$)		

Tabla 4. Distribución del FM (frente al FP) en Palma y Alcalá según variables sociolingüísticas

La distribución de las formas de futuro en los datos de Palma parece apuntar a la existencia de un proceso de cambio por el que, tal como se ha atestiguado en muchas otras áreas de habla hispana, el FP está ganando terreno a expensas del FM. El hecho de que los hablantes más jóvenes y de mediana edad utilicen el FP con mayor frecuencia en comparación con los mayores sería indicativo de ese cambio. Asimismo, la mayor frecuencia del FM entre los hablantes con estudios superiores confirmaría que esta es la variante conservadora, ya que esta forma es la tradicional y además está asociada a registros formales. Si en efecto se estuviera produciendo un cambio, el uso relativamente más alto de la forma innovadora por parte de las mujeres sería evidencia de que, aunque los hablantes con mayores niveles de instrucción favorecen el FM como variante más tradicional, no hay connotaciones negativas asociadas a la variante innovadora, ya que estaríamos ante un proceso liderado por uno de los grupos sensibles al

prestigio sociolingüístico asociado a las variantes. Por otro lado, parece que el uso del FP es bastante estable en Alcalá, pues ninguna de las variables sociolingüísticas consideradas en nuestro estudio alcanzó significatividad estadística. Aun así, aunque la diferencia sea pequeña, en esta comunidad los jóvenes utilizan más el FP, lo cual apunta a que la tendencia de cambio está de algún modo presente.

En conclusión, el análisis contrastivo de dos corpus comparables nos permite demostrar la influencia del contacto con el catalán en la retención del FM, tendencia avalada por los índices relativamente altos de frecuencia de esta variable en otras comunidades bilingües de castellano y catalán según los estudios mencionados más arriba.

4. Perífrasis deónticas y epistémicas

Otro fenómeno variable del español es el empleo de las perífrasis deónticas y epistémicas *haber de* y *tener que*. Del mismo modo que sucedía en el caso de las formas de futuro, hay una acusada asimetría entre el castellano y el catalán en lo que refiere a la frecuencia relativa de estas estructuras. La Tabla 5 a continuación muestra la distribución de *haber de* y *tener que* en los datos de español peninsular del CORPES XXI (Corpus del Español del Siglo XXI) junto a la de *tenir de* y *haver de* en el CTILC (Corpus textual informatitzat de la llengua catalana).

CORPUS	Construcción	Tokens	Frecuencia relativa
Español peninsular CORPES	<i>tener que</i> INF	5358	85,5 %
	<i>haber de</i> INF	908	14,5 %
Catalán CTILC	<i>tenir de</i> INF	4245	3,1 %
	<i>haver de</i> INF	131159	96,2 %

Tabla 5: Distribución de perífrasis con ‘haber de’ y ‘tener que’ más infinitivo en corpus de español peninsular y catalán

Los datos de corpus resumidos en la Tabla 5 no dejan duda de que la construcción con *haver de* más infinitivo en catalán, que alcanza el 96,2%, es mucho más frecuente que la construcción con *tenir de*, con 3,1%, mientras que en castellano ocurre precisamente lo contrario: la perífrasis con *tener que* es ampliamente mayoritaria (85,5%) al tiempo que la de *haber de* es empleada solamente en el 14,5% de los casos. Es importante señalar que existe además en el catalán de Mallorca y Valencia la perífrasis *tenir que* como préstamo del castellano no admitido por la norma. Aunque esta perífrasis apenas tiene presencia en la lengua formal escrita y por ello no se refleja en el corpus sin duda su existencia en el registro oral puede haber influido en la distribución de las perífrasis en el castellano de los territorios bilingües. En cualquier caso, la distribución asimétrica de *haver de*, muy frecuente en catalán, frente a *haber de*, mucho menos usada en castellano, nos lleva a una predicción lógica: la influencia del catalán facilitará que en el castellano hablado por los bilingües catalanohablantes la perífrasis con *haber de* frente a la de *tener que* sea más frecuente que en el castellano monolingüe. Este efecto ha sido estudiado por Blas Arroyo (2016: 88-89), quien presenta resultados del análisis de un corpus

de lengua escrita formado por cartas privadas. Conforme a lo esperable, las cartas producidas en las regiones del ámbito lingüístico catalán exhiben un mayor uso de la perífrasis común a ambas lenguas, *haber de*, con un 32% en los datos de la segunda mitad del siglo XX, mientras que las cartas de zonas monolingües presentan un porcentaje mucho menor, del 9%. Otros estudios hacen observaciones semejantes, como es el caso de Martínez Díaz (2002); Sinner (2004); Enrique-Arias (2010: 106).

Con el objetivo de investigar la influencia del contacto con el catalán en la distribución de las perífrasis de obligación, en Rosemeyer y Enrique-Arias (2023) examinamos datos del castellano producido en tres comunidades bilingües que, como ya se ha expuesto, presentan diferentes grados de intensidad en lo que respecta al conocimiento y uso del catalán: Valencia, Palma y Mallorca rural. En los dos primeros casos utilizamos los datos de los corpus PRESEEA respectivos, y para Mallorca rural los del COSER. Los resultados se resumen en la Tabla 6 a continuación.

Dialecto	<i>Tener que</i>	<i>Haber de</i>	Frecuencia relativa de <i>haber de</i>
Valencia	235	3	1.3 %
Mallorca	822	14	1.7 %
Mallorca rural	380	18	4.5 %

Tabla 6: Distribución de perífrasis verbales en los tres corpus ($n = 3991$)

Los datos confirman que, entre los hablantes con mayor grado de uso del catalán (en este caso los de la Mallorca rural) la prevalencia en el uso de la perífrasis con *haber de* es mayor que en las localidades en las que se usa más el castellano. Además, en los datos de Mallorca, para los que tenemos información de la edad de los hablantes, el uso de *haber de* es más frecuente entre los participantes del COSER, con hablantes de más edad, en comparación con los del PRESEEA, que son en conjunto más jóvenes. Ante todo, no obstante, hay que tener en cuenta que los porcentajes de uso de *haber de* son en general muy bajos, lo cual apunta a la casi desaparición de esta perífrasis en el castellano oral, incluso en áreas rurales con una alta densidad de catalanohablantes. El avance de *tener que* a expensas de *haber de* no solamente afecta al castellano: en el catalán coloquial de Mallorca algunos hablantes utilizan el calco del castellano *tenir que*, si bien no tenemos datos cuantitativos que nos indiquen la distribución de esta perífrasis en el catalán de Mallorca.

Otro aspecto relevante que cabe investigar es el de los diferentes valores de las perífrasis con *haber de* y *tener que*. Por un lado, estas perífrasis pueden expresar significados deónticos de necesidad u obligación, que se considera el significado básico y original. De este significado han derivado otros de tipo epistémico, es decir, de probabilidad o futuridad, tal como se representa en la siguiente evolución adaptada de van der Auwera y Plungian (1998: 96):

Necesidad > Obligación > Probabilidad/Futuridad

En efecto, un mismo enunciado con una perífrasis modal de obligación puede dar lugar a lecturas deónticas o epistémicas dependiendo del contexto. Consideremos, por ejemplo, la oración *Juan ha de / tiene que estar en el edificio*. Si el contexto es que Juan tiene una reunión obligatoria en ese edificio, es decir, que tiene una obligación de estar en ese lugar, se

trata de una interpretación deóntica. Sin embargo, si el contexto es que se ha visto el coche de Juan aparcado delante del edificio y, por tanto, cabe suponer que Juan se encuentra en él, no hay una obligación sino solo una posibilidad y, por lo tanto, se trata de una interpretación epistémica. Al respecto de este contraste de significados hay una asimetría entre el castellano y el catalán: se ha observado que en castellano *tener que* se usa más en contextos deónticos y *haber de* en los de tipo epistémico, mientras que en catalán *haver de* agrupa ambos significados de manera más intercambiable puesto que la perífrasis con *tenir* está “ausente en la gramática del catalán, salvo en empleos que obedecen a un claro calco sintáctico del castellano” (Blas Arroyo 2006: 85)².

En definitiva, nos interesa también evaluar cómo las particularidades del catalán respecto de la modalidad de *haver de*, y no solo la diferencia de frecuencia, tiene un efecto en su distribución. El problema es que en la mayoría de los casos la diferencia entre los valores deónticos y epistémicos no es tan nítida, y muchos enunciados pueden combinar los dos. Conscientes de esta situación hemos optado por hacer una codificación de la modalidad mediante *proxies*, es decir, a partir de ciertos rasgos lingüísticos que son indicativos de la probabilidad de que estas oraciones expresen lecturas modales epistémicas. En particular, y siguiendo a Blas Arroyo (2016: 92-93), nos servimos de tres factores que suponen un aumento de probabilidad de la interpretación epistémica. Los factores en cuestión son a) la presencia de un verbo estativo en el infinitivo de la perífrasis; b) que el verbo auxiliar (*tener* o *haber*) esté en 3ª persona singular; c) que el verbo auxiliar esté en un tiempo con aspecto imperfectivo. A la hora de clasificar los ejemplos de nuestros datos de acuerdo con la probabilidad de una interpretación epistémica damos un punto por cada uno de estos factores, de modo que la probabilidad puede ir de 0 a 3 como se ilustra en los siguientes ejemplos de codificación:

[Probabilidad = 0]

se marcharon a la Argentina, Cuba, y después **tuvieron que volver**

[Probabilidad = 1]

si no tienen papeles no les dan trabajo (.) entonces **tienen que comer**

[Probabilidad = 2]

entonces entre todos la cosa **tiene que salir bien**

[Probabilidad = 3]

en aquella época **tenía que ser** un poco distinto de ahora

Los resultados de la codificación de la probabilidad de modalidad epistémica aparecen resumidos en el Gráfico 1:

²Una tercera perífrasis que puede expresar valores epistémicos y deónticos tanto en castellano como en catalán es la de *deber / deure* con infinitivo. Un estudio más amplio de la expresión de estos valores en el castellano de los bilingües catalanohablantes debería tener en cuenta estas perífrasis.

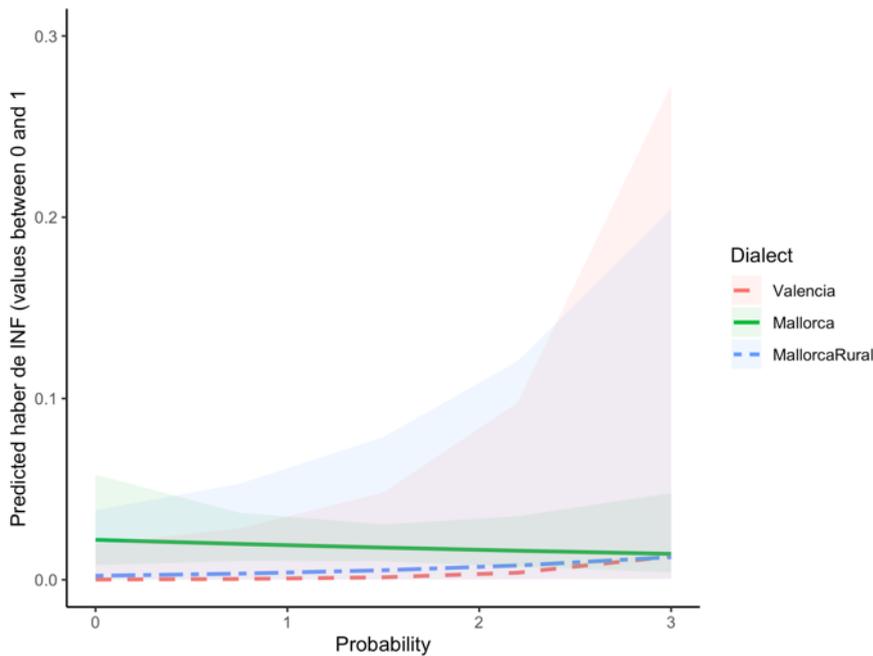


Gráfico 1. Probabilidad de uso de *haber de* según *proxies* de valor epistémico.

El Gráfico 1 muestra en el eje vertical la probabilidad de uso de *haber de* frente a *tener que* y en el eje horizontal los valores de posibilidad de interpretación epistémica que van de 0 a 3. De acuerdo con los resultados, la distribución de *haber de* y *tener que* se rige en menor medida por la diferencia entre modalidad deóntica y epistémica en el español de Mallorca que en el español de Valencia, es decir, el grado de bilingüismo predice la distribución de las variables no solo en términos de frecuencia (recordemos que en Valencia el porcentaje de uso de *haber de* era el más bajo), sino también en lo que se refiere a las funciones discursivas de las perífrasis.

5. Conclusiones

La principal conclusión que se puede sacar de los datos y análisis expuestos en las páginas precedentes es de tipo metodológico. En efecto, la comparación de corpus de la red PRE-SEEA compilados en zonas monolingües y en situaciones con diferentes grados de bilingüismo ofrece grandes posibilidades para estudiar los efectos del contacto lingüístico en los procesos de variación y cambio. La principal ventaja es la de poder contrastar datos de corpus comparables, pues en todos los corpus de la red PRESEEA se han utilizado criterios idénticos para la selección de los informantes y se han seguido los mismos módulos temáticos en el desarrollo de las entrevistas. De este modo se ha podido demostrar cómo la influencia del catalán y el grado de bilingüismo predice la mayor frecuencia de variantes recesivas en el español monolingüe, como es el caso del futuro morfológico *cantaré* o de la perífrasis *haber de* (Enrique-Arias 2010). Asimismo, como hemos visto en el caso de las perífrasis de obligación, la influencia de la lengua en contacto condiciona además el significado y función de las variantes y los patrones de competición entre ellas. En concreto, se ha demostrado cómo, en el caso del castellano producido por hablantes dominantes en catalán en Mallorca, la distribución de

haber de y *tener que* se rige en menor medida por la diferencia entre modalidad deóntica y epistémica de lo que se ha señalado para el español monolingüe.

Por otro lado, tanto en el caso de las formas de futuro como en el de las perífrasis con *haber de*, el comportamiento de los jóvenes en Mallorca exhibe menos influencia del catalán y mayor semejanza con el castellano de los monolingües en comparación con los mayores. Esta distribución apuntaría cierta tendencia de cambio por el cual se está dando un proceso de convergencia con el castellano monolingüe. Esta tendencia está en línea con otros procesos de convergencia en el castellano de Mallorca facilitados por la extensión de la escolarización, los medios de comunicación en castellano y la inmigración masiva de castellanohablantes a las Islas Baleares (cf. Enrique-Arias 2020, 2021, 2022).

La realización de estudios sistemáticos de un buen número de fenómenos de variación del español en contacto con el catalán nos permitirá conocer mejor los mecanismos lingüísticos y sociales que facilitan la convergencia de gramáticas en situaciones de bilingüismo. Como hemos ilustrado en estas páginas, los corpus de la Red PRESEEA son un elemento de gran valor para llevar a cabo esta tarea.

6. Referencias bibliográficas

Aaron, J. E. (2010). Pushing the envelope: Looking beyond the variable context. *Language Variation and Change*, 22(1), 1–36.

Blas Arroyo, J. L. (2007). El contacto de lenguas como factor de retención en procesos de variación y cambio lingüístico. Datos sobre el español en una comunidad bilingüe peninsular. *Spanish in Context*, 4(2), 263–289.

Blas Arroyo, J. L. (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.

Blas Arroyo, J. L. (2016). La relevancia del contacto de lenguas como factor condicionante en un proceso de cambio lingüístico en español: la perífrasis *haber de* + infinitivo. In C. de Benito Moreno & Á. S. Octavio de Toledo y Huerta (Eds.), *En torno a 'haber': Construcciones, usos y variación desde el latín hasta la actualidad* (pp. 79–110). Peter Lang.

Buzón García, J. M. (2013). La expresión de futuridad: Variación y cambio lingüísticos. In J. R. Gómez Molina (Ed.), *El español de Valencia. Estudio sociolingüístico* (pp. 145–182). Peter Lang.

Enrique-Arias, A. (2022). El castellano en Mallorca 60 años después: un estudio en tiempo real. *Revista Internacional de Lingüística Iberoamericana*, 20(39), 137–163.

Enrique-Arias, A. (2021). Vamos en Palma 'we are going to Palma. On the persistence (and demise) of contact features in the Spanish of Majorca. In W. Chappell & B. Drinka (Eds.), *Spanish Socio-historical Linguistics: Isolation and Contact* (pp. 186–203). John Benjamins.

Enrique-Arias, A. (2020). Agreement mismatches in the Spanish preterite of Catalan-dominant bilinguals in Majorca: a receding interlanguage phenomenon. In M. Schlechtweg

(Ed.), *The learnability of complex constructions. A cross-linguistic perspective* (pp. 219–237). De Gruyter.

Enrique-Arias, A. (2010). On language contact as an inhibitor of language change: the Spanish of Catalan bilinguals in Majorca. In A. Breitbarth et al. (Eds.), *Continuity and change in grammar* (pp. 97–118). John Benjamins.

Enrique-Arias, A., & Méndez Guerrero, B. (2020). On the effects of Catalan contact in the variable expression of Spanish future tense. A contrastive study of Alcalá de Henares, Madrid and Palma, Majorca. In L. Ortiz López et al. (Eds.), *Hispanic Contact Linguistics. Theoretical, methodological and empirical perspectives* (pp. 315–334). John Benjamins.

Generalitat Valenciana. (2021). *Conocimiento y uso social del valenciano. Encuesta 2021. Síntesis de resultados*. Retrieved from <https://ceice.gva.es/>

Illamola, C. (2013). La convergència lingüística: Possible retardador del procés de gramaticalització? *Treballs de Sociolingüística Catalana*, 23, 81–101.

Martínez Díaz, E. (2002). *Las perífrasis modales de obligación “tener que + infinitivo” y “haber de + infinitivo”: Variación e interferencia en el español de Barcelona* [Doctoral dissertation, Universitat de Barcelona]. Retrieved from <https://www.tdx.cat/handle/10803/1713>

Melià, J., & Vanrell, M. del M. (2018). *Enquesta d’usos lingüístics a les Illes Balears 2014. Anàlisi*. Govern de les Illes Balears/Generalitat de Catalunya/Universitat de les Illes Balears.

Melis, C. (2006). Verbos de movimiento. La formación de los futuros perifrásticos. In C. Company Company (Dir.), *Sintaxis histórica de la lengua española* (Vol. 2, Tomo 1, pp. 875–970). Fondo de Cultura Económica, Universidad Nacional Autónoma de México.

Moreno Fernández, F. (2020). *Variedades de la lengua española*. Routledge.

Osborne, S. D. (2008). *Variable Future Tense Expression in Andalusian Spanish* [Master’s thesis, University of Georgia, Athens]. Retrieved from https://getd.libs.uga.edu/pdfs/osborne_samantha_d_200808_ma.pdf

Rosemeyer, M., & Enrique-Arias, A. (2023). Deontic and epistemic verbal periphrases in Catalan-Spanish contact varieties. Paper presented at the 45th Annual Conference of the Deutsche Gesellschaft für Sprachwissenschaft – DGfS, University of Cologne, March 7–10, 2023.

Sedano, M. (1994). El futuro morfológico y la expresión ir a + infinitivo en el español. *Verba*, 21, 225–240.

Sinner, C. (2004). *El castellano de Cataluña. Estudio empírico de aspectos léxicos, morfosintácticos, pragmáticos y metalingüísticos*. Niemeyer.

van der Awera, J., & Plungian, V. (1998). Modality’s semantic map. *Linguistic Typology*, 2, 79–124.