

# El lexema verbal y otras variables en la gramaticalización del futuro y el condicional en navarroaragonés antiguo

Antoine Primerano (Universiteit Gent/Humboldt-Universität zu Berlin)

 <https://orcid.org/0000-0003-0437-5646>

Miriam Bouzouita (Humboldt-Universität zu Berlin)

 <https://orcid.org/0000-0001-8196-620X>

**Résumé:** Cet article se penche sur la variation morphosyntaxique propre des langues ibéro-romanes médiévales entre les formes analytiques de futur et conditionnel avec un pronom mésoclitique (p. ex. *cantar lo é, cantar lo ía* ‘je le chanterai(s)'), d’une part, et les formes synthétiques de ces paradigmes avec un pronom postverbal (p. ex. *cantaré lo, cantarí lo* ‘je le chanterai(s)'), d’autre part. Étant donné que cette variation a été peu étudiée de manière systématique, il reste encore à déterminer quels sont les facteurs linguistiques qui la conditionnent et influencent la sélection d’une ou l’autre forme. Un travail préliminaire (Primerano/Bouzouita à paraître) examine, suivant une méthodologie bivariée, le possible rôle de trois variables morphosyntaxiques, en l’occurrence, la syncope, le temps verbal et la présence d’un verbe non fini/d’une construction périphrastique. L’influence du lexème verbal dans la variation, en revanche, n’a pas été analysée pour cause de la taille réduite du corpus employé dans cette étude. Par conséquent, l’article que nous présentons ici vise à étoffer le corpus de Primerano/Bouzouita (à paraître) et à mesurer quantitativement l’effet de cette variable, ainsi que son interaction avec d’autres variables considérées comme pertinentes pour la variation en question. De plus, nous analyserons de nouveau le rôle de ces autres variables au moyen de techniques quantitatives multivariées plus précises et nous nous interrogerons sur les éventuels effets de la fréquence d’emploi des différentes constructions. Notre étude se base sur des données extraites des textes navarro-aragonais des 13<sup>e</sup> et 14<sup>e</sup> siècles consultables dans une banque de données digitale relativement récente, le *Old Spanish Textual Archive* (OSTA). Les résultats de nos analyses suggèrent que la syncope et la construction périphrastique sont les facteurs les plus décisifs dans la variation. Le lexème verbal semble également jouer un rôle, qui pourrait bien être un épiphénomène de la régularité de la syncope pour certains verbes et de la fréquence d’usage des lexèmes.

**Palabras clave:** futuro, condicional, mesoclisís, posposición, lexema, síncopa, frecuencia, navarroaragonés

El presente trabajo<sup>1</sup> aborda desde un ángulo variacionista cuantitativo el papel del lexema verbal y diversas variables en una alternancia sintáctica del iberorromance medieval hasta ahora escasamente estudiada: la variación entre las formas analíticas de futuro y condicional, con pronombre mesoclítico (p. ej. *cantar lo é, cantar lo ía*), y las sintéticas de esos mismos paradigmas con pronombre pospuesto (p. ej. *cantaré lo, cantarí lo*). Este estudio, basado en 64 lexemas extraídos de un corpus navarroaragonés de los siglos XIII-XIV, emplea métodos cuantitativos para tratar de descubrir qué condicionantes morfosintácticos y léxicos entran en juego en la variación en cuestión, que está directamente relacionada con el grado de gramaticalización de estos paradigmas verbales. A través de estos análisis, el estudio, por un lado, revela los beneficios de combinar enfoques cuantitativos y cualitativos, a mayor y menor escala y, por otro, aborda las dificultades de aplicar métodos estadísticos a fenómenos de baja frecuencia, incluso en masas textuales extensas.

El artículo se organiza de la siguiente manera: tras una introducción general sobre la gramaticalización del futuro y el condicional en iberorromance (§1), nos detendremos en la variación entre las formas analíticas con pronombre mesoclítico y las sintéticas con pronombre pospuesto y sus posibles condicionantes, entre ellos el lexema verbal (§2). A continuación, se especificarán los objetivos, detallando el corpus y la metodología empleados en este estudio (§3), antes de pasar a los resultados de los análisis empíricos (§4). La descripción y discusión de resultados comienza por una visión de conjunto de los datos y del perfil variacional de los lexemas tratados (§4.1), sigue con la aplicación de modelos estadísticos, más específicamente, de regresión lineal generalizada de efectos mixtos (§4.2) y un árbol generalizado de efectos mixtos (§4.3) para medir el efecto del lexema y de otros condicionantes de la variación introducidos en §2, y finaliza con un tratamiento de los posibles efectos de frecuencia (§4.4). Cerraremos este estudio detallando las principales conclusiones (§5).

## 1. Introducción: la gramaticalización del futuro y el condicional iberorrománicos

Es bien sabido que los paradigmas morfológicos de futuro y condicional (en adelante FC) en iberorromance (p. ej. esp. *cantaré, cantarí*) son el producto de la gramati-

1 Esta investigación ha sido posible por la financiación del *Fonds voor Wetenschappelijk Onderzoek* (FWO) Vlaanderen a través de una beca predoctoral (investigación fundamental, número de proyecto 11I0820N). Queremos dar las gracias a Mar Garachana, Álvaro Octavio de Toledo y Huerta y Javier Rodríguez Molina por sus valiosas sugerencias respecto a algunas decisiones metodológicas. Se nos impone también agradecer a Ludovic De Cuypere y Matti Marttinen Larsson, quienes nos ofrecieron amablemente su ayuda con los análisis estadísticos. Agradecemos asimismo a los/las evaluadores/as anónimos/as, que contribuyeron a mejorar este trabajo. Cualquier error que permanezca es, naturalmente, responsabilidad nuestra.

calización de la perífrasis latina [infinitivo + HABERE] (p. ej. CANTARE HABEO, CANTARE HABEBAM, Fleischman 1982, Nocentini 2001, Company 2006, inter alia). En el periodo medieval existe una variación entre, por un lado, las llamadas formas sintéticas (en adelante FCS), que pueden aparecer sin pronombre personal de objeto (1a-b), con un pronombre antepuesto (2a-b) o pospuesto (3a-b), y, por otro, las formas analíticas (en adelante FCA), que exhiben un pronombre mesoclítico intercalado entre el antiguo infinitivo y el morfema derivado de HABERE (4a-b)<sup>2</sup> (Company/Medina 1999, Company 2006, Bouzouita 2011, 2012, 2016, inter alia).

- (1) a. Si non, *dexaremos* Burgos (*Cid*, 1438, apud Company 2006: 351)
- b. et qualquier de nós que allá vaya, *costraría* la villa (*Lucanor*, 116, apud Company 2006: 388)
- (2) a. .ij. míos fijos *te dexaré* (*Fazienda*, 55, apud Bouzouita 2011: 115)
- b. mucho *me plazería* que sopiéssedes lo que conteçió (*Lucanor*, 254, apud Company 2006: 394)
- (3) a. e *tornaré los* a sus logares (E6, *Jeremías*, 23:3, apud Bouzouita 2016: 270)
- b. Mi orden & mis manos *querría las* ondrar (*Cid*, 2373, apud Graham 2018: 204)
- (4) a. *Enprenarte as* e avras fijo (*Fazienda*, 207, apud Bouzouita 2011: 111)
- b. Quanta riqueza tiene *aver la yemos* nos (*Cid*, 2663, apud Company 2006: 351)

Esta variación sincrónica demuestra que la gramaticalización de la perífrasis latina hacia una forma sintética, como la que encontramos en español contemporáneo, está menos avanzada en la Edad Media en comparación con la actualidad. Efectivamente, se atestigua en las variantes analíticas, por la presencia de un pronombre de objeto entre los dos componentes de la forma verbal, una falta de univervación, definida como «the union of two syntagmatically adjacent word forms into one» (Lehmann 2020: 206). El periodo de variación entre estas dos variantes se concluye en los siglos XVI-XVII en español, con la generalización de las formas sintéticas y la desaparición de las analíticas (Castillo Lluch 2002, Company 2006, Octavio de Toledo y Huerta 2015, inter alia).

Como se ha demostrado (Lema/Rivero 1991, Castillo Lluch 1996, 2002, 2004, Bouzouita 2011, 2012, 2013, 2016, Bouzouita/Sentí 2022, inter alia), los FCS con anteposición (cf. 2a-b) y los FCA con mesoclisís (cf. 4a-b) entran en una distribución complementaria que se vincula directamente con los principios que rigen la colocación

2 En los ejemplos se marca en cursiva la forma de FC, el pronombre átono de objeto en cursiva y negrita, mientras que otros constituyentes relevantes aparecen subrayados.

del pronombre personal átono de objeto frente al verbo y, por tanto, viene condicionada por el entorno sintáctico-pragmático del grupo verbal, entendido como el constituyente que lo precede (contra Company 2006).

En concreto, los FCS con pronombre antepuesto se documentan en los entornos de anteposición del pronombre con otros tiempos verbales, que se caracterizan por la presencia de algún elemento con una carga pragmática fuerte delante del verbo. Uno de ellos es el adverbio negativo *no(n)*, ilustrado en (5a-b), pero también pueden ser pronombres interrogativos, objetos no correferenciales con el pronombre átono, ciertos tipos de adverbio o de oración subordinada, y otros más (Granberg 1988, 1999, Castillo Lluch 1996, 2004, Bouzouita 2008a, 2008b). Los FCA, en cambio, aparecen en los mismos contextos sintáctico-pragmáticos que los pronombres posverbiales con otros tiempos (Castillo Lluch 1996, Bouzouita 2011, 2012, Bouzouita/Sentí 2022). Estos contextos son, por lo general, pragmáticamente no marcados, como la conjunción adversativa *mas* en (6a-b), por ejemplo, cuya función es más bien de organización del discurso<sup>3</sup>. En vista de ello, los FCA – así como las otras formas verbales con pronombres pospuestos – constituyen una construcción pragmáticamente no marcada y los FCS con pronombres antepuestos – al igual que los otros tiempos con anteposición del pronombre – surgen principalmente en entornos enfáticos o pragmáticamente cargados. Aparte de los FCS con anteposición y los FCA con mesoclisís, existen también FCS con pronombres pospuestos (en adelante FCS-P; cf. los ejemplos (3a-b)), cuya distribución e importancia serán el tema del siguiente apartado.

- (5) a. non *me oyestes?* (*Fazienda*, 55, apud Bouzouita 2011: 115)  
 b. NoI *farán mal* (*Fazienda*, 164, apud Bouzouita 2011: 115)
- (6) a. mas *dixome* (*Fazienda*, 207, apud Bouzouita 2011: 112)  
 b. mas *abatirlas as* (*Fazienda*, 77, apud Bouzouita 2011: 112)

## 2. La variación entre formas analíticas y sintéticas con pronombre pospuesto

Los FCS-P del iberorromance medieval (cf. 3a-b) se documentan en exactamente los mismos entornos sintáctico-pragmáticos que los FCA, es decir, los entornos de posposición pronominal con otros tiempos verbales (Bouzouita 2011, 2013, 2016). Desafortunadamente,

3 Para un tratamiento más profundo de la variación en la colocación de los pronombres átonos de objeto en español antiguo, véanse Granberg (1988, 1999), Castillo Lluch (1996, 2002, 2004), Nieuwenhuijsen (2002, 2006) y Bouzouita (2008a, 2008b, 2011). Para más información sobre la estructura informativa del español antiguo, véanse Batllori (2021) y Bouzouita/Sitaridou (2023).

tunadamente, han sido desatendidos por la mayor parte de los estudios sobre la variación entre FCS y FCA hasta ahora, con la notable excepción de Eberenz (1991), Bouzouita (2011, 2013, 2016, 2017) y Graham (2018), por citar las contribuciones más significativas. La razón de este vacío bibliográfico reside seguramente en la escasez de los datos castellanos medievales disponibles, ya que los FCS-P apenas aparecen en los textos castellanos anteriores al siglo XV. A pesar de ello, estas construcciones aparentemente marginales revisten un gran interés, ya que exhiben la unverbación de sus partes constitutivas y, por tanto, representan un mayor grado de coalescencia<sup>4</sup> entre los dos componentes de la perífrasis de origen en comparación con los FCA, si bien ambas estructuras se documentan en los mismos entornos sintáctico-pragmáticos.

En otras palabras, en contextos oracionales idénticos compiten dos estructuras, una de las cuales exhibe un mayor grado de fusión que la otra. Este mayor avance en la coalescencia puede ser indicativo de un proceso de gramaticalización más desarrollado (cf. Hopper/Traugott 2003, Haspelmath 2011). Esta interpretación viene corroborada por los hechos diacrónicos, ya que, en castellano a partir de los siglos XVI-XVII, los FCS-P ganan en frecuencia hasta volverse más numerosos que los FCA, conforme al avance de la gramaticalización de los futuros y condicionales. Además, algunos trabajos recientes (Bouzouita/Sentí 2022, Sentí/Bouzouita 2022, Torres-Latorre 2023, Torres-Latorre/Sentí 2023, Primerano/Bouzouita en prensa, Primerano et al. en prensa) han demostrado que, en los siglos XIII-XIV, estos FCS-P son más comunes en las variedades iberorrománicas orientales, como el navarroaragonés y el catalán, que en castellano, por lo que la gramaticalización parece estar más avanzada en aquel periodo al este de la península ibérica, en comparación con el centro(-oeste) (cf. también Bouzouita 2016: 294-95)<sup>5</sup>.

Debido a la escasa frecuencia de los FCS-P antes del siglo XV en castellano, la variedad en la que indudablemente se han centrado los estudios de morfosintaxis histórica del español (cf. Fernández-Ordóñez 2006, 2011, Pons Rodríguez 2006, Pato 2018), se conoce aún en la actualidad relativamente poco sobre las variables que motivan la variación entre estas estructuras y las analíticas. Una notable excepción son los trabajos versados sobre la Biblia escorialense I-i-6 (en adelante E6) de mediados del siglo XIII, ya que esta obra ofrece un número de FCS-P significativamente más elevado que otros textos castellanos de la misma época. A modo de ilustración, en los libros *Isaías* y *Jeremías* los FCS-P representan un 44,1% (174/395) de las ocurrencias en los entornos sintáctico-pragmáticos relevantes frente a un 55,9% (221/395) de FCA (Bouzouita

4 Seguimos aquí la descripción de Lehmann (2020), para quien la coalescencia se refiere al incremento gradual de las relaciones de dependencia entre varios elementos (cf. Haspelmath 2011) y la unverbación es una de las fases en este proceso más amplio.

5 De hecho, la veintena de ejemplos de FCS-P que se suelen mencionar consistentemente en la bibliografía sobre los futuros castellanos (p. ej. Batllori 2016, Bouzouita 2011, 2012, Castillo Lluch 1996, 2002, Eberenz 1991, inter alia) provienen o bien de copias tardías y por tanto filológicamente poco fiables (cf. Fernández-Ordóñez 2006, Rodríguez Molina 2010: 612-28) o bien de fuentes con características dialectales orientales.

2016)<sup>6</sup>. Con base en los datos de E6 se han propuesto una serie de factores que parecen favorecer el uso de los FCS-P frente a los FCA, los cuales forman el eje central de nuestro estudio y se detallan en lo que sigue: la presencia de síncope en la forma verbal (§2.1), el tiempo verbal (§2.2), la presencia de un verbo no finito detrás del FC (§2.3) y el lexema verbal (§2.4). Para el navarroaragonés de los siglos XIII-XIV, el estudio de Primerano/Bouzouita (en prensa) examina los mismos factores mediante una metodología bivariable. A continuación, resumiremos también los resultados de este estudio y precisaremos cómo se ampliará en el presente artículo.

### 2.1 La síncope

El factor de la síncope verbal (p. ej. *prender* > *prendrá*) es sin duda el que más ha sido considerado en la bibliografía sobre la variación que nos ocupa. El primero en resaltar la correlación entre síncope y colocación del pronombre fue Moreno Bernal (2004a), quien, en un estudio panrománico e histórico de la síncope en las formas de FC, propuso que «la pérdida temprana de la vocal final de la raíz del infinitivo está en relación inversa con el uso de las formas analíticas [de futuro y condicional]» (p. 129). Efectivamente, el autor observa que en francés, por ejemplo, la síncope en los FC ha sido un proceso altamente productivo y los FCA no se documentan en la historia de esa lengua, mientras que el portugués, en cambio, siempre ha hecho un uso mucho más moderado de la síncope en los FC y todavía conoce los FCA en la actualidad, en su variedad europea<sup>7</sup>. Por consiguiente, una acusada presencia de la síncope en una variedad determinada parece desfavorecer el uso de las estructuras analíticas, las cuales no admiten formas sincopadas (p. ej. *prender lo á* vs. *\*prendr lo á*). Matute/Pato (2010) siguen en esta línea al discutir la incidencia de este fenómeno en la biblia E6:

Si observamos además las formas que presentan alternancia en la posición del pronombre, vemos que las soluciones con cambio en la base del infinitivo (*pornáslo, serásle, dirásles* o *farélos* frente a *ponerla as, seer l'á, dezir l'as, fazerles edes*) atraen al pronombre en posición de enclisis con más frecuencia que en mesoclis (37 ejemplos de enclisis, el 73%, frente a 14 de mesoclis, el 27%). Al contrario, si el pronombre acompaña a un infinitivo que se ha mantenido invariable, la mesoclis (77 casos) se iguala a la enclisis (76 casos) (Matute/Pato 2010: 62)<sup>8</sup>.

La distinción hecha por los autores – entre verbos «con cambio en la base del infinitivo» y verbos «con infinitivo invariable» –, no viene nutrida por la morfología de la ocurrencia de FC en concreto, sino por la *posibilidad o no* del verbo de aparecer con un radical sincopado, ya que los 14 FCA recogidos con verbos de la primera

6 Esta biblia castellana exhibe, de forma llamativa, numerosos orientalismos lingüísticos (Enrique-Arias 2010).

7 Conviene precisar que los FCA en portugués europeo actual se consideran formas arcaizantes, propias de los registros formal y literario, y tienden a desaparecer en el uso diario de la lengua (Martins 2016).

8 Precisamos que el término *enclisis* se emplea allí en sentido sintáctico y no fonológico, es decir, como sinónimo de posposición verbal.

categoría («con cambios en la base del infinitivo») exhiben un radical completo, como *dezir l'as* o *fazerles edes* (Matute en comunicación personal), y, por tanto, podrían considerarse asincopados. Por consiguiente, si recalculamos estas proporciones en función de la morfología de la forma de FC concreta – es decir, si el token concreto es sincopado o asincopado –, en las ocurrencias sincopadas los FCS-P suben hasta el 100% (37/37) de los casos en el corpus de Matute/Pato (2010), mientras que, en los casos asincopados, estas variantes representan el 45,5% (76/167) de las ocurrencias, frente a un 54,5% (91/167) de FCA (los cálculos son nuestros<sup>9</sup>). Esta clara tendencia de los FCS-P a aparecer con verbos sincopados se confirma en Bouzouita (2016: 284-89), quien documenta para los verbos de la 2ª conjugación<sup>10</sup> un 100% (32/32) de FCS-P con verbos sincopados, como en (7a), y un 100% (31/31) de FCA con verbos asincopados, como en (7b). De forma parecida, para la 3ª conjugación, recoge un 100% (22/22) de FCS-P con formas sincopadas, y con formas no sincopadas un 90,7% (39/43) de FCA, frente a un 9,3% (4/43) de FCS-P, caso ilustrado en (7c). Estas cifras revelan una fuerza de correlación sumamente alta, ya que la *V* de Cramér es 1 para la 2ª conjugación y 0,876 para la 3ª (Bouzouita 2016: 286).

- (7) a. E *sabrà-lo* tod'el pueblo de Efraym (E6, *Isaías*, 9:8, apud Bouzouita 2016: 282)
- b. e acerques e uenga el conseio del sancto Israhel e *saber lo emos* (E6, *Isaías*, 5:19, apud Bouzouita 2016: 282)
- c. e *seruirán le* muchas yentes e grandes reyes (E6, *Jeremías*, 27:6-7, apud Bouzouita 2016: 286)

Primerano/Bouzouita (en prensa) detectan en su análisis bivariable para el navarroaragonés de los siglos XIII-XIV, al igual que Matute/Pato (2010) y Bouzouita (2016), cierta correlación entre la síncope y la colocación pronominal con los FC, dado que, con los verbos de la 2ª y 3ª conjugación, los FCS-P son absolutos en las formas sincopadas (100%, 16/16), mientras que los FCA predominan en las asincopadas con un 97,5% (39/40), frente a un 2,5% (1/40) de FCS-P. El único FCS-P asincopado implica el verbo de la 2ª conjugación *conocer*.

9 En estos nuevos cálculos, los 37 casos sincopados que mencionamos corresponden a las 37 ocurrencias «con cambio en la base del infinitivo» y con pronombre pospuesto identificados por Matute/Pato (2010), mientras que los 167 casos asincopados que obtenemos resultan de la suma de los 76 casos «con radical invariable» y con pronombre pospuesto, las 77 ocurrencias analíticas de este tipo (con pronombre mesoclitico) y las 14 formas analíticas incluidas como casos «con cambio en la base del infinitivo» de estos autores.

10 Bouzouita (2016) separa los resultados por modelo de conjugación, dado que los verbos de la 1ª conjugación no conocen la posibilidad de sincoparse y, por tanto, no pueden exhibir variabilidad al respecto.

A pesar de la correlación observada en estos trabajos<sup>11</sup>, la síncopa en sí no logra elucidar totalmente la razón del uso de los FCS-P, como ya ha notado Bouzouita (2016: 284-89), puesto que se documentan (i) FCS-P con verbos de la 1ª conjugación (8a), la cual no conoce la síncopa, y (ii) casos de FCS-P con formas no sincopadas de la 2ª o 3ª conjugación, como en (8b). Además, queda por explicar por qué en algunos casos se elige una forma sintética sincopada con un pronombre pospuesto, y no una forma analítica con un radical completo y un pronombre mesoclítico, cuando ambas estructuras están disponibles en el sistema lingüístico, como en los ejemplos (7a-b) respectivamente.

- (8) a. Vos sodes carga de dios, *desecharé uos* (E6, *Jeremías*, 23:33, apud Bouzouita 2016: 283)
- b. Trabaié iudgando los e *esparzirá los* con el abentadero en las puertas de la tierra (E6, *Jeremías*, 15:6-7, apud Bouzouita 2016: 286)

## 2.2 El tiempo verbal

El tiempo de la forma verbal, en este caso futuro o condicional, también ha sido señalado como posible factor influyente en el uso de los FCS-P: los pronombres átonos parecen manifestarse en posición pospuesta más frecuentemente con un condicional (9a) que con un futuro (9b) (Bouzouita 2017).

- (9) a. Los mios deuen ser, que son vnos locos y desarman a quantos pasan y *huyriales* alguno (*Celestina*, 215, apud Eberenz 1991: 505)
- b. Pues *dirás les* estas palauras (E6, *Jeremías*, 13:12, apud Bouzouita 2016: 283)

En el estudio de Primerano/Bouzouita (en prensa) sobre el navarroaragonés antiguo, este factor parece ejercer una influencia en la variación entre FCA y FCS-P, si bien solo en el siglo XIV: en ese periodo el condicional se construye con los FCS-P en el 68,4% (13/19) de los casos, al contrario del futuro, que aparece en esta configuración en un menudo 8,2% (4/49) de las ocurrencias, lo cual resulta en una fuerza de correlación moderada ( $V = 0,6244$ )<sup>12</sup>.

11 La correlación entre síncopa y FCS-P ha sido notada también por Graham (2018). No obstante, ese estudio presenta varios problemas metodológicos, por lo que no discutiremos sus resultados en este apartado. Para una crítica, véase Primerano/Bouzouita (en prensa).

12 También Graham (2018: 215-16) concluye que el condicional muestra una fuerte tendencia hacia los FCS-P, mientras que el futuro parece preferir los FCA. No obstante, aparte de los problemas metodológicos de este estudio, los cuales ponen en duda la fiabilidad de los resultados de su análisis (cf. supra), es importante subrayar que para los siglos XIII y XIV el corpus de Graham (2018)

### 2.3 La presencia de una forma verbal no finita pospuesta

Otro factor posiblemente influyente es la presencia de un verbo no finito – un infinitivo, gerundio o participio pasado – detrás de la forma del FC. A partir de datos provenientes de diferentes variedades iberorrománicas y del occitano, Bouzouita (2017) observa, sin entrar en detalles cuantitativos, que a primera vista los FCS-P parecen surgir sobre todo cuando van seguidos por un infinitivo, como en (10a-b).

- (10) a. Mi orden & mis manos *querri<sup>al</sup>as ondrar* (*Cid*, 237)  
 b. Et puesto qu'el león non me pensase mal, usando con él los malos consejeros, *avri<sup>al</sup>os de escuchar* (*Calila e Dimna*<sup>13</sup>, 162)
- (11) puede la retener por .i. dia. & dar la en otro dia a so conpannero que es parçonero en aquella bestia de qui non ha clamos. & *poder lo ha fer* aquesto por siempre (*Fueros de Aragón*, fol. 33r<sup>o</sup>)

Primerano/Bouzouita (en prensa), por su parte, observan que en el navarroaragonés del siglo XIII los pronombres se posponen al FC en el 50% (4/8) de los casos cuando sigue una forma verbal no finita – un infinitivo o un participio pasado – y en el 4,3% (2/47) en la ausencia de tal forma. En el siglo XIV, esa diferencia parece acentuarse, ya que en la presencia de un verbo no finito los FCS-P se usan en el 93,3% (14/15) de las ocurrencias, frente al 5,7% (3/53) cuando no hay tal verbo. Este factor tiene pues una fuerza de correlación moderada en el siglo XIII ( $V = 0,5173$ ) y elevada en el siglo XIV ( $V = 0,8396$ ). Los autores intuyen que el efecto de la forma verbal no finita en la variación podría deberse a razones cognitivas de procesamiento. Efectivamente, una estructura como la de (11), en la que el FCA (*poder lo ha*) constituye el auxiliar de una perífrasis modal (*poder* + infinitivo), podría representar un esfuerzo cognitivo para los y las hablantes, ya que el FCA mismo es una construcción que mantiene rasgos perifrásticos debido a su origen histórico (Company 2006: 378-82, Octavio de Toledo y Huerta 2015, cf. el principio de estratificación de Hopper 1991) y exhibe un orden de constituyentes no canónico, lo cual podría explicar la preferencia por los FCS-P en estas configuraciones (véase Primerano/Bouzouita en prensa para una explicación detallada).

### 2.4 El lexema verbal

El último factor discutido en este trabajo es el lexema verbal. Concretamente, Bouzouita (2017) observa que el uso de FCS-P parece afectar a ciertos verbos más que a

---

contiene solo 5 ejemplos de condicionales, lo cual, en definitiva, pone en duda la contundencia de sus conclusiones.

13 Es de notar que el texto de *Calila e Dimna*, si bien fechado en el siglo XIII, se conserva en dos copias tardías de los siglos XIV-XV.

otros, como *decir*, *haber*, *hacer*, *poder*, *ser*, por ejemplo. Previamente, la misma autora (Bouzouita 2016: 294) había registrado en E6 un nítido predominio de FCS-P con los lexemas *decir* (18/21), *poner* (7/8), *quemar(se)* (8/11) y *tornar* (13/16).

Primerano/Bouzouita (en prensa) recogen en su corpus navarroaragonés unos 64 verbos diferentes en estructuras de FCA y FCS-P. Entre estos 64 verbos encontramos unos de semántica muy general, tales como *ser*, *hacer*, *decir*, por ejemplo, y otros más específicos, como *morder*, *sahumar*, *ahinojar*, entre otros. La gran mayoría de estos, concretamente 53 de los 64 (p. ej. *abrir*, *comer*, *oír*, *perder*), se documenta en su corpus únicamente en FCA, mientras que unos 4 (*conocer*, *convenir*, *morder*, *querer*) solo ofrecen ejemplos de FCS-P y unos 7 (*decir*, *haber*, *hacer*, *poder*, *prender*, *ser*, *venir*) alternan entre una y otra estructura. Dicho de otro modo, los autores observan los FCA con 60 verbos diferentes y los FCS-P con 11. En vista de ello, parece que los FCS-P, en el navarroaragonés de los siglos XIII-XIV, se dan con un número restringido de verbos, mientras que los FCA aparecen con un abanico más amplio, lo cual sugiere que ciertos verbos lideran el cambio. Aun así, cabe reconocer que el estudio de Primerano/Bouzouita (en prensa) maneja una cantidad baja de datos, ya que se basa en tan solo 100 ocurrencias de FCA y 23 de FCS-P, a pesar del gran tamaño del corpus empleado por los autores (910.231 palabras). Esto se debe, en parte, a que la abrumadora mayoría de los casos de FC incluidos en su corpus lleva un pronombre antepuesto (92,9%, 1.608/1.731), haciendo de los FCA y los FCS-P estructuras de ocurrencia más bien marginal (7,1%, 123/1.731). Desafortunadamente, fue imposible en ese estudio llevar a cabo un análisis abarcador y concluyente de la variable del lexema verbal con tan pocos datos, por lo que el presente estudio pretende ampliar la cantidad de datos analizados para cada uno de estos lexemas verbales a fin de examinar esta variable de forma más rigurosa. Además, se desconoce si la supuesta preferencia de los FCS-P por un grupo reducido de verbos encontrada por Primerano/Bouzouita (en prensa) se debe simplemente al lexema que representan o si se puede explicar acudiendo a los otros tres factores (síncopa, tiempo verbal y verbo no finito). Por ejemplo, la tendencia de un hipotético verbo  $\alpha$  a ocurrir en estructuras de FCS-P en el corpus, podría ser el resultado de la alta frecuencia de  $\alpha$  en forma sincopada en el mismo corpus. Por otra parte, los factores de la síncopa, el tiempo verbal y el verbo no finito se analizaron siguiendo una metodología exclusivamente bivariable, mediante la cual el efecto de cada factor se examinó de manera individual. Por ello, no fue posible medir la interacción entre los diferentes factores. En este estudio nos proponemos remediar este vacío mediante técnicas cuantitativas multivariadas.

En el apartado siguiente, formulamos las preguntas de investigación que se responderán mediante un estudio de corpus y métodos cuantitativos, así como la metodología de recogida y anotación de los datos puesta en práctica en este estudio (§3).

### 3. Objetivos, corpus y metodología

Lo antes referido nos lleva a especificar los objetivos de este trabajo: examinar la posible influencia del lexema verbal en la variación entre FCA y FCS-P en navarroaragonés antiguo, refinando y ampliando el análisis de esta variable ya iniciado en Primerano/Bouzouita (en prensa), además de la interacción entre ella y los demás factores. Para este fin, aplicamos métodos cuantitativos multivariados, los cuales ofrecen medidas más exactas del efecto de cada una de las variables. Las preguntas de investigación que forman el eje central del artículo son las siguientes:

- I. ¿Influye el lexema verbal en la alternancia bajo estudio? De ser el caso, ¿cómo y en qué medida explica la variación?
- II. ¿Cómo interactúa el lexema verbal con las otras variables? ¿Se puede considerar un epifenómeno de ellas?
- III. Respecto a los otros factores identificados en la bibliografía, ¿continúan siendo significativos cuando los integramos en un análisis multivariado más complejo?

Para realizar este estudio, se han utilizado los datos del *Old Spanish Textual Archive* (OSTA, cf. Gago/Pueyo 2018a, 2018b). Hemos tomado como base la lista de 64 lexemas que aparecen en estructuras de FCA y FCS-P identificados en el corpus de Primerano/Bouzouita (en prensa). De cada uno de esos verbos hemos extraído todas las formas de FCA y FCS-P de los textos de los siglos XIII-XIV marcados en el OSTA como aragoneses, navarroaragoneses y navarros (cf. Apéndice 1). La circunscripción de nuestro corpus al ámbito geográfico navarroaragonés<sup>14</sup> se justifica por dos motivos. De una parte, el uso más común de los FCS-P en las variedades peninsulares orientales en aquel periodo frente al castellano (cf. §2) permite obtener suficientes datos para efectuar análisis estadísticos contundentes. Por otra parte, este enfoque geográfico ofrece la posibilidad de comparar nuestros resultados con los de estudios anteriores (p. ej. Primerano/Bouzouita en prensa) y avanzar sobre la base de los trabajos ya realizados. Con todo, deberá entenderse que los resultados obtenidos

14 En la tradición filológica hispánica algunos autores han considerado el navarroaragonés como una entidad dialectal con variantes diatópicas (p. ej. Ynduráin 1945, Alvar 1953, Buesa 1980), mientras que otros han preferido caracterizarlo como dos dialectos distintos, navarro y aragonés, que coinciden en ciertas propiedades (p. ej. González-Ollé 1970, 1983, 1996, Saralegui 1977, 1992, Enguita 2008). En el caso de los FC, Primerano/Bouzouita (en prensa) señalan que los textos provenientes de Navarra podrían exhibir un grado de gramaticalización de estos paradigmas más alto que los de Aragón. Los datos de este estudio, al contrario, parecen indicar lo opuesto: los textos de Aragón contienen un 29,7% (88/296) de FCS-P y un 70,3% (208/296) de FCA, mientras que en los de Navarra se registra un 14,6% (7/48) de FCS-P y un 85,4% (41/48) de FCA ( $X^2 = 4,01$ ;  $p = 0,0452$ ). Sin embargo, téngase en cuenta que las fuentes textuales aragonesas son bastante más numerosas que las navarras en nuestro corpus. Además, el volumen de datos que manejamos no nos permite discriminar sistemáticamente entre navarro y aragonés. Por lo tanto, no entraremos en detalle en esta cuestión aquí, ya que nuestro enfoque principal no es diatópico, y seguiremos empleando el término «navarroaragonés» por razones de brevedad.

sobre las variables que condicionan la variación bajo estudio serán válidos para el navarroaragonés de los siglos XIII-XIV, sin que necesariamente se apliquen del mismo modo a otras variedades iberorrománicas u otros periodos.

Para la extracción de los datos se procedió mediante el etiquetado gramatical y la lematización que ofrece el OSTA, así como las posibilidades de búsqueda por forma paleográfica<sup>15</sup>. Como era de esperar, dado que el OSTA es un corpus compuesto por transcripciones paleográficas, ciertas formas con grafías poco comunes, en ocasiones, no salen bajo el lema que les corresponde, por lo cual ha sido necesario combinar ambos métodos<sup>16</sup>. Nuestro corpus comprende todos los datos recopilados del OSTA, a los cuales hemos agregado los ejemplos encontrados previamente en Primerano/Bouzouita (en prensa) que proceden de textos no incluidos en el OSTA (cf. Apéndice 1) o de textos sí incluidos en el OSTA, pero que no salen en las búsquedas. En total, nuestro corpus contiene 3.090.120 palabras. Dentro de este conjunto de datos, cada ocurrencia fue anotada para las variables siguientes:

- a) Síncopa: asincopado (12a-b) o sincopado (12c);
- b) Tiempo verbal: futuro (13a) o condicional (13b);
- c) Forma no finita: con un verbo no finito pospuesto (14a) o sin verbo no finito pospuesto (14b);
- d) Perífrasis: si el FC funciona como auxiliar de una perífrasis verbal (15a), si es un verbo léxico (15b) o si entra en una construcción causativa (15c);
- e) Lexema verbal: basado en los lexemas de Primerano/Bouzouita (en prensa) y anotado con ortografía normalizada.

- (12)
  - a. mas *fazer lia* medeçina (*Rams de Flors*, fol. 189r°)
  - b. e *faras lo* bien picar (*Libro de Palladio*, fol. 16r°)
  - c. A todo mjo poder *querra la* guardar (*Libro de Apolonio*, fol. 41v°)

15 Los datos incorporados en nuestro corpus de estudio se han sacado usando las siguientes búsquedas: (i) futuro de indicativo de lexema  $\alpha$ , (ii) condicional de lexema  $\alpha$ , (iii) infinitivo de lexema  $\alpha$ . Cuando el número de resultados para esas búsquedas era demasiado elevado, hemos acotado, mediante el etiquetado, buscando solo aquellas formas seguidas por un pronombre clítico. También hemos tomado en cuenta aquellas estructuras en las que el pronombre átono va unido gráficamente con el verbo (p. ej. *ueranlo* [*Crónica de conquiridores II*, fol. 88r°]), aquellas en las que el pronombre va unido con el morfema derivado de *haber* (p. ej. *adormir sea* [*Rams de Flors*, fol. 139r°]), así como formas analíticas en que los tres componentes se escriben unidos (p. ej. *dizirtele* [*Vidal Mayor*, fol. 267r°]).

16 A modo de ejemplo, el FCA *safumarsea* (*Secreto Secretorum*, fol. 292v°) no sale dentro de los resultados de la búsqueda bajo el lema *sahumar*.

- (13) a. e *dezir uos he* de lures fechos (*Libro de Marco Polo*, fol. 67v°)  
 b. e el *dezir le ye* la verdat (*Libro de Apolonio*, fol. 41v°)
- (14) a. e *fer le has beuer* el vjno (*Libro de Palladio*, fol. 121r°)  
 b. E *mudar las has* en otro vaso (*Libro de Palladio*, fol. 208r°)
- (15) a. E *podras lo mezclar* con la simjente (*Libro de Palladio*, fol. 169r°)  
 b. e *mostrar uos he ligado* el uuestro princep (*Crónica del tudense*, fol. 64r°)  
 c. e *fazer la as bien ensangrentar* (*Grant Crónica de Espanya I*, fol. 24v°)
- (16) a. Et *dir uos he* las cosas que son en la part contraria (*Tucidides*, fol. 11r°)  
 b. no guardedes por aquestas cosas et *far loedes* por uuestro prouecho (*Tucidides*, fol. 55v°)

Nótese que para la variable de la síncopa se han codificado como asincopadas las formas de FC «reducidas» de los verbos *hacer*, *decir*, *ver* y *ser*. Estos verbos se documentan en los textos navarroaragoneses medievales bajo una variedad de infinitivos (*fazer*, *far*, *fer*; *dezir*, *dizir*, *dizer*, *dir*; *veer*, *veyr*, *veyer*, *ver*; *seer*, *seyr*, *seyer*, *ser*), unos más largos que otros, lo que dificulta determinar si una forma de FCS de estos verbos es sincopada o asincopada<sup>17</sup>. A modo de ilustración, el FC *faras* en (12b) podría interpretarse como una forma sincopada/reducida si partimos de un infinitivo *fazer* o asincopada si consideramos que el infinitivo es *far*. Hemos decidido resolver esta ambigüedad morfológica anotando estas formas como asincopadas por las siguientes razones: (i) contienen una vocal delante del morfema TAM *-ré*, *-rás*, *-rá*, *-ría*, *-rias*, etc.<sup>18</sup> (p. ej. *farás*), rasgo que comparten con las asincopadas (p. ej. *perderás*), pero no con las sincopadas (p. ej. *perdrás*); (ii) los morfemas *dir-*, *far-*, etc., se dan también en los FCA, como puede observarse en (16a-b), indicando que estos radicales pueden participar en la variación entre FCA y FCS, al contrario de los sincopados que solo aparecen en los FCS (p. ej. *perdrás* vs. *\*perdr lo as*). Además, algunos autores opinan que los FCS de *hacer* y

17 Este problema no se plantea con los FCA, ya que no admiten la síncopa y, por tanto, siempre exhiben un radical completo. Por ejemplo, en la forma *fazer vos hemos* (*Crónica de conquistadores*, fol. 413r°), se reconoce sin lugar a duda el infinitivo *fazer*.

18 Consideramos que la consonante vibrante forma parte del sufijo TAM en los FCS (Hualde et al. 2010: 156). Como lo expone Company (2006: 374), «[I]a frontera gramatical estaba etimológicamente entre el infinitivo y *habere*, pero la gramaticalización provocó que la vibrante, que es segmento integrante y definitorio del infinitivo, y por tanto parte de la base léxica de la construcción originaria, pasara a integrar el morfema flexivo desinencial, de manera que en el español actual las formas de futuro aparecen identificadas en las gramáticas de referencia como formas en *-ré*, *-ría*, y no formas en *-é*, *-ía*, como sería lo esperado desde un punto de vista etimológico».

*decir* «proceden muy probablemente de formas del latín tardío *fare* y *dire*» (Moreno Bernal 2004a: 136, cf. también Saralegui 1983, Dubert-García 2021).

En el caso de la variable de la perífrasis, optamos por una clasificación tripartita, entre estructuras no perífrásticas, perífrásticas y causativas. La razón para asignarles a las estructuras causativas un nivel aparte radica en que, a pesar de que no se suelen considerar como perífrasis verbales, comparten numerosos rasgos comunes con ellas (cf. Gómez Torrego 1999: §51.1.5, Alfonso Vega 2006).

A continuación, exponemos de forma descriptiva los resultados de nuestras búsquedas (§4.1), medimos el efecto de los diferentes factores por medios cuantitativos (§4.2 y §4.3) y evaluamos el posible papel de los efectos de frecuencia en la variación (§4.4).

## 4. Resultados

### 4.1 Descripción de los datos y perfil variacional

En total, extrajimos del OSTA 344 ocurrencias, de las cuales 249 (72,4%) son FCA y 95 (27,6%) son FCS-P. Esta cantidad de datos puede parecer bastante reducida a primera vista, pero recuérdese que el corpus empleado contiene más de 3 millones de palabras, por lo que estamos ante dos estructuras simplemente poco frecuentes en los textos<sup>19</sup>. Además, dado que este estudio se organiza en torno a una lista de lexemas verbales concretos, ha de subrayarse la dificultad de recopilar un número importante de observaciones para cada uno de estos verbos.

En estudios anteriores (Primerano/Bouzouita en prensa) se observó que no todos los verbos parecen alternar entre FCA y FCS-P, sino que algunos lexemas tienden a mostrar una preferencia, a veces categórica, por una u otra construcción (cf. §2.4). Por eso y a fin de comprobar estos rasgos diferenciales entre lexemas, presentamos en el Cuadro 1 la distribución de los FCA y los FCS-P por lexema verbal y por siglo. Hemos clasificado los lexemas en tres categorías según el perfil que muestran en el corpus: (i) los que se construyen únicamente con los FCA, (ii) los que exhiben variación entre FCA y FCS-P, y (iii) los que se construyen exclusivamente con los FCS-P<sup>20</sup>.

19 Los FCS con anteposición parecen constituir la mayoría de las estructuras de FC, por lo menos en los datos navarroaragoneses antiguos: en Primerano/Bouzouita (en prensa) representan un 92,9% (1.608/1.731) de las construcciones de FC.

20 El Cuadro 1 debe leerse de la siguiente manera: para cada lexema en una línea determinada se ha encontrado la distribución especificada en esa línea. Por ejemplo, para los lexemas *acaloñar*, *pechar*, *pendrar* y *pesar* (segunda línea), hemos encontrado dos FCA en el siglo XIII y ninguna ocurrencia en el siglo XIV.

Perfil general	Lexema verbal	Siglo XIII		Siglo XIV		
		FCA	FCS-P	FCA	FCS-P	
FCA absolutos	<i>comprender, enardecer, encalzar, heredar, pagar, quitar</i>	100% (1/1)	0% (0/1)	-	-	
	<i>acaloñar, pechar, pendrar, pesar</i>	100% (2/2)	0% (0/2)	-	-	
	<i>emendar</i>	100% (3/3)	0% (0/3)	-	-	
	<i>aborrecer, abstener, adormir, ahinojar, ajustar, cerrar, clamar, cobijar, entristecer, gozar, mojar, pasar, perfumar, poblar, recordar, rehacer, rentar, sahumar, semejar</i>	-	-	100% (1/1)	0% (0/1)	
	<i>alegrar, comer, enseñorear, menospreciar, perder, posar, trovar, venir</i>	-	-	100% (2/2)	0% (0/2)	
	<i>abrir, mudar, seguir</i>	-	-	100% (3/3)	0% (0/3)	
	<i>echar, sacar</i>	-	-	100% (8/8)	0% (0/8)	
	<i>esforzar, oír</i>	100% (1/1)	0% (0/1)	100% (1/1)	0% (0/1)	
	<i>ayudar</i>	100% (1/1)	0% (0/1)	100% (2/2)	0% (0/2)	
	<i>matar</i>	100% (1/1)	0% (0/1)	100% (3/3)	0% (0/3)	
	<i>llevar</i>	100% (2/2)	0% (0/2)	100% (5/5)	0% (0/5)	
	<i>mostrar</i>	100% (4/4)	0% (0/4)	100% (10/10)	0% (0/10)	
	<b>Total de lexemas</b>	<b>76,6% (49/64)</b>				
	<b>Total de ocurrencias</b>	<b>109 (0,35 por 10.000 palabras)</b>				

Variación entre FCA y FCS-P	<i>conocer</i>	-	-	66,7% (2/3)	33,3% (1/3)
	<i>convenir</i>	-	-	33,3% (1/3)	66,7% (2/3)
	<i>dar</i>	100% (9/9)	0% (0/9)	88,9% (8/9)	11,1% (1/9)
	<i>decir</i>	83,3% (10/12)	16,7% (2/12)	85% (17/20)	15% (3/20)
	<i>demandar</i>	100% (2/2)	0% (0/2)	0% (0/1)	100% (1/1)
	<i>haber</i>	80% (4/5)	20% (1/5)	18,2% (2/11)	81,8% (9/11)
	<i>hacer</i>	100% (3/3)	0% (0/3)	80,4% (45/56)	19,6% (11/56)
	<i>poder</i>	60% (3/5)	40% (2/5)	4,1% (2/49)	95,9% (47/49)
	<i>prender</i>	100% (1/1)	0% (0/1)	50% (1/2)	50% (1/2)
	<i>ser</i>	0% (0/1)	100% (1/1)	25% (1/4)	75% (3/4)
	<i>tener</i>	100% (6/6)	0% (0/6)	75% (6/8)	25% (2/8)
	<i>tornar</i>	100% (2/2)	0% (0/2)	93,3% (14/15)	6,7% (1/15)
	<i>ver</i>	100% (1/1)	0% (0/1)	0% (0/1)	100% (1/1)
	<b>Total de lexemas</b>		<b>20,3% (13/64)</b>		
<b>Total de ocurrencias</b>		<b>228 (0,73 por 10.000 palabras)</b>			
FCS-P absolutos	<i>morder</i>	-	-	0% (0/1)	100% (1/1)
	<i>querer</i>	0% (0/3)	100% (3/3)	0% (0/2)	100% (2/2)
	<b>Total</b>	<b>3,1% (2/64)</b>			
	<b>Total de ocurrencias</b>	<b>6 (0,02 por 10.000 palabras)</b>			

Cuadro 1. El perfil variacional de los lexemas verbales en el corpus

El Cuadro 1 muestra que en la totalidad del corpus los FCA ocurren con un abanico bastante más amplio de verbos que los FCS-P: 62 lexemas verbales (49 con FCA absolutos + 13 con variación) vs. 15 (13 con variación + 2 con FCS-P absolutos), respectivamente. Dicho esto, no debe obviarse que los FCA (72,4%, 249/344) son también más frecuentes que los FCS-P (27,6%, 95/344) en nuestro corpus. Es, por tanto, metodológicamente más preciso calcular la frecuencia relativa de los lexemas por FCA y FCS-P. Los resultados confirman nuestra primera impresión: aparece 0,25 lexema diferente por cada FCA frente a 0,16 por cada FCS-P, por lo que los FCS-P ocurren efectivamente con un número de lexemas diferentes ligeramente inferior al de los FCA<sup>21</sup>.

Del Cuadro 1 se desprende también que el perfil variacional de la mayoría de los lexemas verbales presentes en los datos (76,6%, 49/64) es categórico: se documentan únicamente en formas de FCA. El perfil de FCS-P absolutos, por su parte, se manifiesta con tan solo 2 de los 64 lexemas verbales (el 3,1%): *morder* y *querer*. El verbo *morder* aparece en un solo ejemplo del siglo XIV, mientras que el verbo *querer*, aunque no se documenta en más de cinco ejemplos en total, parece mostrar, aun así, una preferencia por los FCS-P en ambos siglos. Dentro del grupo de variación entre FCA y FCS-P, que contiene 13 lexemas (el 20,3%), se puede constatar una posible evolución diacrónica. En el siglo XIII, la cuasitotalidad de estos verbos tiende a aparecer en formas analíticas. Con siete de ellos, los FCA son hasta absolutos: *dar*, *demandar*, *hacer*, *prender*, *tener*, *tornar* y *ver*. Si examinamos su distribución en el siglo XIV, vemos que la mayoría suele aún preferir los FCA, pero ya no de manera absoluta: el 88,9% (8/9) de los casos para *dar*, el 80,4% (45/56) para *hacer*, el 50% (1/2) para *prender*, el 75% (6/8) para *tener* y el 93,3% (14/15) para *tornar*. Tómese en cuenta, sin embargo, que el número de ejemplos para cada verbo es relativamente bajo, lo que invita a interpretar estas cifras con cautela. Los verbos *demandar* y *ver*, en cambio, solo se recogen con un ejemplo en el siglo XIV, lo cual nos impide sacar conclusiones de esta distribución. Respecto a los verbos caracterizados por un predominio no absoluto de los FCA en el siglo XIII, como *decir* (83,3%, 10/12), *haber* (80%, 4/5) y *poder* (60%, 3/5), solo *decir* mantiene este predominio en la centuria siguiente, con un porcentaje similar (85%, 17/20), al contrario de *haber* y *poder*, que pasan a favorecer los FCS-P, de modo bastante flagrante (el 81,8%, 9/11, para *haber* y el 95,9%, 47/49, para *poder*). Cabe añadir que no encontramos verbos que tiendan a los FCS-P en el siglo XIII y a los FCA en el XIV. Los perfiles observados en el grupo de variación parecen así estar en línea con la tendencia diacrónica general, por la cual los FCS-P van progresivamente superando a los FCA en términos de frecuencia (cf. §2).

En general, los FCS-P parecen estar todavía restringidos en el siglo XIII, ya que aparecen con solo cinco lexemas diferentes: *decir*, *haber*, *poder*, *querer* y *ser*. En el siglo XIV, todos estos verbos se repiten y aparecen otros diez con FCS-P, a saber:

21 Estas cifras se obtuvieron dividiendo el número de lexemas verbales que aparecen con los FCA (62) por el número de ocurrencias de FCA (249), por una parte, y el número de lexemas que aparecen con los FCS-P (15) por el número de ocurrencias de FCS-P (95), por otra.

*conocer, convenir, dar, demandar, hacer, morder, prender, tener, tornar y ver.* Parece producirse así, a primera vista, una ampliación de los lexemas que admiten los FCS-P, por lo cual esta colocación pronominal surgiría con un número cada vez más elevado de verbos con el paso del tiempo. De ser así, podríamos estar ante un fenómeno de *host-class expansion* (Himmelmann 2004, cf. también Lehmann 1985): una construcción en vías de gramaticalización se emplea con un número creciente de palabras y admite cada vez más elementos con que combinarse (*hosts*). Sin embargo, si calculamos la ratio tipo-caso (*type-token ratio*), es decir, el número de lexemas diferentes por cada ocurrencia de FCS-P en cada siglo, obtenemos un resultado diferente: equivale a 0,11 en el siglo XIII y desciende a 0,08 en el XIV. El número de verbos que admiten las construcciones de FCS-P no parece, por ende, aumentar entre el siglo XIII y el XIV, sino más bien disminuir<sup>22</sup>. Pese a ello, cabe recordar que los verbos que aparecen como FCS-P en nuestros datos del siglo XIII parecen fortalecer su uso de esta estructura en la centuria siguiente, lo cual sí podría indicar un incremento.

En resumen, hemos comprobado en estos datos numéricos que los FCS-P parecen efectivamente, en el navarroaragonés de los siglos XIII-XIV, ser característicos de algunos verbos en particular, puesto que los FCA aparecen con un número de lexemas superior al de los FCS-P. Esta observación ha quedado confirmada por la ratio tipo-caso de ambas colocaciones pronominales. También hemos visto que, aunque a primera vista parece haber un aumento diacrónico en el número de verbos que aparecen en estructuras de FCS-P, esta impresión viene desmentida por las ratios tipo-caso de ambos siglos, las cuales muestran incluso un descenso en el número de verbos. Sin embargo, los datos del Cuadro 1 reflejan una tendencia por la cual unos verbos que en el siglo XIII ocurrían mayoritariamente o hasta siempre en FCA llegan a ocurrir en FCS-P en el siglo XIV, a veces de forma predominante. En lo siguiente, analizamos los datos aquí presentados con técnicas estadísticas para medir el efecto de cinco factores en la variación que se estudia: sincopa, tiempo verbal, presencia de alguna forma no finita o perífrasis y lexema (cf. §3). De este modo, pretendemos descubrir si la preferencia de ciertos lexemas verbales por los FCS-P se debe en realidad a la influencia de otros factores.

#### 4.2 Regresión lineal de efectos mixtos

En este apartado, cuantificamos la influencia de los cinco factores examinados (sincopa, tiempo verbal, verbo no finito, perífrasis, lexema) en la variación entre FCA y FCS-P. Como se ha mencionado (cf. §3), para llegar a este objetivo, aplicaremos un modelo estadístico de regresión lineal generalizada de efectos mixtos y un árbol generalizado de efectos mixtos<sup>23</sup>. La regresión lineal generalizada de efectos mixtos

22 Además, el corpus del siglo XIV es de tamaño mayor (2.595.850 palabras) que el del XIII (494.270 palabras).

23 Estos análisis se han efectuado en R (R Core Team 2022).

(Tagliamonte/Baayen 2012) evalúa si, en un conjunto de datos dado, el valor que adopta alguna variable llamada independiente o predictor (p. ej. sincopado vs. asincopado) influye en la probabilidad de que la variable dependiente – es decir, el resultado – adopte un valor determinado (p. ej. FCA vs. FCS-P). En otras palabras, este método permite ver si la variación bajo estudio viene condicionada por una serie de factores preestablecidos. La ventaja de esta técnica es que, al medir el efecto de cada variable independiente, se tiene en cuenta también el efecto de todas las demás variables independientes al mismo tiempo. De este modo, se pueden evitar medidas posiblemente engañosas que, por ejemplo, en un análisis bivariable, predecirían un efecto para cierta variable que en realidad es epifenómeno de la distribución que tiene otra variable en el corpus<sup>24</sup>. Además, estos modelos permiten cuantificar la interacción entre diferentes variables en el conjunto de datos.

El modelo es denominado *de efectos mixtos* porque distingue entre dos tipos de efectos o variables. De una parte, tenemos efectos *fijos*, que incluyen las variables con un número preestablecido y cerrado de valores. Un caso prototípico de estos son las variables gramaticales que codifican el uso de dos estructuras alternantes: por ejemplo, una variable  $\alpha$  puede adoptar el valor  $x$  o  $y$ , pero ningún otro. Por otra parte, se pueden medir efectos *aleatorios*, los cuales incluyen variables cuyos valores no están representados de forma exhaustiva en la muestra. Típicamente, son variables como el/la participante en un experimento lingüístico, el texto de donde proviene el ejemplo en estudios de corpus, entre otros. Se tratan generalmente como efectos aleatorios aquellas variables que no permiten entrever una explicación obvia para su incidencia en la variación. En nuestro estudio, hemos incluido cuatro efectos fijos – (i) síncopa, (ii) tiempo verbal, (iii) verbo no finito, (iv) perífrasis – y un efecto aleatorio, lexema.

Antes de pasar a la aplicación del modelo estadístico propiamente dicho, hemos descartado los casos sincopados del conjunto de datos, ya que no existen ocurrencias de FCA sincopadas (cf. §2.1). Este factor presenta, por tanto, un caso de *cuasiseparación* en el que hay incompatibilidad entre un valor de la variable dependiente (en este caso, FCA) y un valor de la variable independiente (en este caso, sincopado). Este desequilibrio nos señala que, con los FC sincopados, no puede haber variación entre FCA y FCS-P. En un estudio variacionista como el presente estos ejemplos sincopados, por consiguiente, deben descartarse de cara a un análisis estadístico preciso, ya que solo se pueden considerar contextos en los que la variación puede manifestarse. Aun así, la cuasiseparación observada en nuestros datos sugiere también que la síncopa es un factor muy potente para explicar la alternancia bajo estudio, dado que en uno de sus valores (FC sincopado) elimina por completo la variación. Con el otro valor (FC asincopado), sí pueden surgir ambas estructuras.

24 Efectivamente, los análisis bivariados basados en la prueba del chi cuadrado ( $\chi^2$ ) presuponen que todas las observaciones del corpus son independientes, es decir, que no tienen nada en común aparte de la variable que se está analizando. En un estudio que considera múltiples factores, este escenario es imposible, ya que dos observaciones diferenciadas por el valor distinto que adoptan para una variable  $\alpha$  podrían coincidir en el valor que adoptan para otra variable  $\beta$ .

Una vez descartados los casos sincopados del corpus, nos quedan 274 observaciones asincopadas de 62 lexemas verbales diferentes, de las cuales 249 (90,9%) son FCA y 25 son FCS-P (9,1%). Es cierto que este recuento final nos deja con una cantidad de datos bastante reducida, por lo cual se podría cuestionar la validez y el sentido de aplicar un modelo estadístico complejo a un corpus tan humilde. A pesar de estas limitaciones, hemos decidido llevar a cabo los análisis propuestos por ser estos, en definitiva, los más refinados. Además, se ha demostrado en otros estudios (p. ej. Marttinen Larsson 2023) que los análisis multivariados son adecuados para estudiar los fenómenos de baja frecuencia e identificar sus condicionamientos. Tras haber eliminado la síncopa, nos quedaban tres efectos fijos – (i) el tiempo verbal, (ii) el verbo no finito y (iii) la perífrasis – y un efecto aleatorio, el lexema. Luego, siguiendo la metodología recomendada por Zuur et al. (2009) y Gries (2015), pasamos por un procedimiento de selección del modelo, quitando uno por uno los efectos fijos a partir de una prueba de razón de verosimilitud para modelos jerárquicos<sup>25</sup>. Este procedimiento resultó en un modelo más sencillo con un solo efecto fijo, la perífrasis, y un efecto aleatorio, el lexema. A continuación, presentamos los resultados para el efecto fijo (§4.2.1) y el efecto aleatorio (§4.2.2).

#### 4.2.1 Efecto fijo: la perífrasis

Los resultados del análisis de regresión están resumidos en el Cuadro 2<sup>26</sup>. En un análisis de regresión, los efectos de las variables independientes se deben interpretar de manera esencialmente relativa o contrastiva: el resultado que el modelo nos devuelve es de cuánto aumenta o disminuye la probabilidad de obtener un resultado determinado si una variable independiente  $\alpha$  adopta un valor  $x$  en comparación con otro valor  $y$ , tomado como el nivel de referencia. Para reformular con la variable concreta incluida en nuestro análisis, el modelo calcula si la probabilidad de obtener un FCS-P cuando la variable de la perífrasis adopta el valor «perifrástico» o «causativo» es inferior o superior a esta misma probabilidad cuando la variable tiene el valor «no perifrástico», el cual funciona como nivel de referencia, y si esta diferencia es significativa o no. Si la diferencia sale como significativa, este resultado indica que la variable analizada influye en la variación bajo estudio. Esto se puede observar en la última columna del Cuadro 2, donde se dan los valores  $p$ : si este valor queda por debajo del umbral de 0,05, el efecto es significativo. En el caso contrario, la variable no influye. La dirección del efecto, es decir, cuál de los valores de la variable dependiente viene favorecido, se deduce a través del coeficiente: si es negativo, el nivel de referencia es favorecido; si es positivo, el otro valor es favorecido.

25 Este procedimiento se ha efectuado manualmente por medio de las funciones *drop1* y *anova* en R.

26 Para el análisis de regresión, hemos utilizado la función *glmer* del paquete *lme4* (Bates et al. 2015).

Variable	Valor	N	Coeficiente	E.E.	Z	p	
Variable dependiente: FCA (nivel de referencia, N = 249) vs. FCS-P (N = 25)							
(Intercepto)			-4,0458	0,9475	-4,270	<0,001	***
Perifrasis	No perifrástico	240	Nivel de referencia				
	Perifrástico	17	2,5381	1,0162	2,498	0,0125	*
	Causativo	17	1,7073	0,7256	2,353	0,0186	*
<b>Evaluación del modelo:</b>							
Índice C de concordancia: 0,87							
Dxy de Somer: 0,75							
R <sup>2</sup> marginal: 9,7%							
R <sup>2</sup> condicional: 52,8%							
Precisión de clasificación: 92%							

Cuadro 2. Resultados del análisis de regresión lineal generalizada de efectos mixtos (efecto fijo)

Con base en los datos presentados en el Cuadro 2, podemos afirmar que la variable de la perifrasis tiene un efecto significativo en la variación entre FCA y FCS-P en nuestro corpus. Efectivamente, los FCS-P están favorecidos cuando el FC funciona como el auxiliar de una perifrasis verbal (cf. 15a) o como verbo causativo (cf. 15c), puesto que en ambos casos el coeficiente es positivo y el valor de *p* queda por debajo de 0,05. El efecto de esta variable se visualiza en la Figura 1. Este resultado contrasta con las conclusiones del estudio de Primerano/Bouzouita (en prensa), donde se registra un efecto significativo también para el tiempo verbal y el verbo no finito (si bien únicamente en el siglo XIV, y no en el XIII). Ahora bien, el análisis que hemos expuesto presenta algunas diferencias importantes con respecto al de Primerano/Bouzouita (en prensa). En primer lugar, frente a la metodología esencialmente bivariable de aquel trabajo previo, hemos empleado aquí un método multivariado, que, como se ha referido, es metodológicamente más preciso. Recuérdese, además, que el conjunto de datos en que se basa el presente análisis solo contiene formas asincopadas, ya que las sincopadas se han descartado por motivos metodológicos y conceptuales (cf. §4.2). Al contrario, el examen tanto del tiempo verbal como del verbo no finito en Primerano/Bouzouita (en prensa) junta FC sincopados y asincopados. Por lo tanto, resulta imposible distinguir, en los análisis de aquel trabajo, entre el efecto de la síncopa y el de otras variables, como el tiempo verbal y el verbo no finito. Por consiguiente, creemos que los resultados expuestos en el presente estudio gozan de más fuerza explicativa. No obstante, nuestro resultado confirma la intuición de Primerano/Bouzouita (en prensa) de que el supuesto efecto de la forma verbal no finita pospuesta se

debe en realidad a la construcción verbal en la que participa el FC, a saber: perifrástica, no perifrástica o causativa<sup>27</sup>. Pasamos ahora al examen del efecto aleatorio.

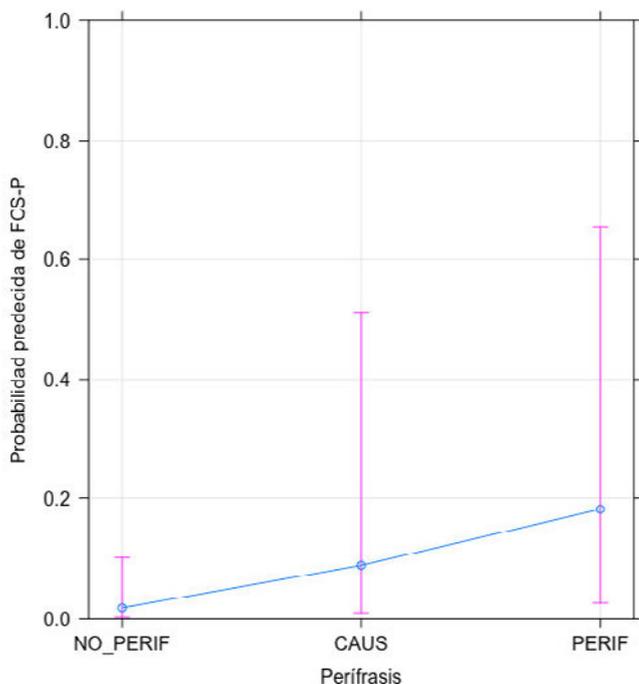


Figura 1. El efecto de la perifrasis en el modelo de regresión

#### 4.2.2 Efecto aleatorio: el lexema verbal

Para responder a una de las preguntas de investigación formuladas en §3 y saber si el lexema verbal influye en la variación entre FCA y FCS-P, tenemos que volver la mirada hacia el efecto aleatorio de nuestro modelo: el lexema. En nuestro análisis de regresión, los verbos que exhiben una preferencia significativa por los FCS-P son *ser*, *decir* y *hacer* (como ya observó Bouzouita 2017). En cambio, no hay verbos que muestren una preferencia significativa por los FCA. El intervalo de confianza para

<sup>27</sup> El procedimiento de selección del modelo que hemos aplicado, si bien es moneda común en los estudios lingüísticos con un enfoque cuantitativo y viene recomendado por ciertos autores (p. ej. Zuur et al. 2009, Gries 2015, 2021), sigue siendo el objeto de bastante controversia en la estadística. Para este estudio, conviene precisar que el modelo inicial que incluía tres efectos fijos y un efecto aleatorio, es decir, antes del procedimiento de selección, no devolvió ningún efecto significativo, ni siquiera de la variable de la perifrasis.

el coeficiente de todos los demás lexemas presentes en nuestros datos se sitúa alrededor del 0, lo cual impide determinar su efecto sobre la variación. De nuevo queremos recordar que, debido al propio objetivo y a la metodología de este trabajo, no es tarea fácil recopilar suficientes datos para todos los lexemas verbales como para permitir que el modelo saque conclusiones fiables respecto a cada uno de ellos.

Este resultado confirma que los FCS-P tienden a ser significativamente más frecuentes con algunos verbos que con otros. Además, el  $R^2$  condicional (el porcentaje de varianza explicado por el modelo entero) mucho más alto que el  $R^2$  marginal (el porcentaje de varianza explicado solo por los efectos fijos) (cf. Cuadro 2) indica que el efecto del lexema no puede simplemente reducirse a las otras variables. El hecho de que el  $R^2$  condicional sea muy superior al  $R^2$  marginal indica que la inclusión de un efecto aleatorio mejora considerablemente la calidad predictiva del modelo<sup>28</sup>.

### 4.3 Árbol generalizado de efectos mixtos

Otra técnica de análisis estadístico multivariado que puede servir para nuestros propósitos son los llamados *árboles generalizados de efectos mixtos* (Fokkema et al. 2018, 2021, Bernaisch 2022, en adelante AGEM). Al contrario de los modelos de regresión, los AGEM – al igual que otros árboles de decisión – no miden el efecto de cada uno de los predictores por medio de un coeficiente, sino que identifican aquellos cuyo efecto es significativo ( $p < 0,05$ ) y los clasifican de más a menos influyente en forma de un árbol binario. Concretamente, esta técnica efectúa una división o un «corte» en los datos, que genera dos subconjuntos con máxima homogeneidad interna y máxima heterogeneidad entre sí, con base en el más significativo de los predictores. Dicho de otro modo, el modelo selecciona el predictor que mejor explica la variación y separa los datos en dos grupos según los valores que puede adoptar este predictor. Luego, se repite este procedimiento en cada subconjunto con el segundo predictor más significativo hasta que no queden predictores significativos. En comparación con los más tradicionales *árboles de inferencia condicional*, los AGEM tienen la ventaja de incluir efectos aleatorios.

El AGEM generado a partir de los datos de nuestro corpus se presenta en la Figura 2. Hemos incluido en la fórmula los tres efectos fijos (tiempo verbal, forma no finita y perífrasis) y un efecto aleatorio (lexema verbal)<sup>29</sup>. Como se puede ver, el AGEM identifica la perífrasis como único predictor significativo ( $p < 0,001$ ): las estructuras peri-

28 Ambas medidas se han obtenido por medio de la función *r.squaredGLMM* del paquete MuMIn (Bartoń 2022) en R.

29 En teoría, se hubiera podido incluir la síncope como efecto fijo, puesto que los árboles de decisión no son sensibles a los problemas de cuasiseparación, al contrario de los modelos de regresión (Tagliamonte/Baayen 2012). Sin embargo, esto habría resultado simplemente en un primer corte en el árbol que habría separado las ocurrencias sincopadas por un lado y las asincopadas por otro, puesto que los casos sincopados son invariablemente sintéticos. Por ello, hemos decidido descartar esta variable y simplificar la fórmula.

frásticas (PERIF) y causativas (CAUS) prefieren los FCS-P significativamente más que las no perífrásticas (NO\_PERIF). Este resultado coincide plenamente con el del análisis de regresión lineal. Respecto al lexema, incluido como efecto aleatorio en el AGEM, los tres verbos que salen como significativamente proclives a aparecer como FCS-P son los mismos que en el análisis de regresión, a saber: *ser*, *decir* y *hacer*<sup>30</sup>.

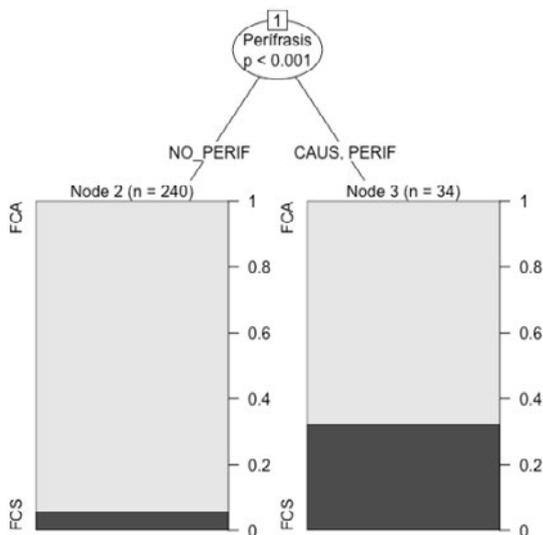


Figura 2. Árbol generalizado de efectos mixtos para la variación entre FCA y FCS-P

A modo de resumen, hemos visto que las dos técnicas estadísticas empleadas – la regresión lineal de efectos mixtos y el árbol generalizado de efectos mixtos – ofrecen resultados altamente similares, si no idénticos: la variable de la perífrasis es la única que influye en la variación en cuestión, ya que las construcciones en las que el FC funciona como auxiliar de una perífrasis o verbo causativo tienden a preferir los FCS-P. También resulta de estos análisis que ciertos verbos son más propensos al uso de los FCS-P que otros, en especial, *ser*, *decir* y *hacer*, lo cual corrobora intuiciones previas (p. ej. Bouzouita 2017). La síncope, por su parte, a pesar de no haberse incluido en los modelos, se presenta como el factor más decisivo en la variación objeto de estudio, puesto que una de las variantes examinadas, el FCA, es simplemente incompatible con los FC sincopados.

Ahora bien, si volvemos a examinar el perfil variacional de los 64 lexemas que componen nuestro corpus (cf. Cuadro 1), constatamos lo siguiente. En primer lugar, no todos los verbos que aparecen en estructuras de FCS-P pueden participar en construcciones perífrásticas o causativas. Solo es el caso para *haber* (*haber* + parti-

<sup>30</sup> En cuanto a la evaluación del modelo, la precisión de clasificación, calculada con la función *confusionMatrix* del paquete *caret* (Kuhn 2022) es de 92% y el índice *C* de concordancia, obtenido con la función *somers2* del paquete *Hmisc* (Harrell 2023) es 0,88.

cipio pasado, *haber* (*a/de* + infinitivo), *hacer* (*hacer* + infinitivo), *poder* (*poder* + infinitivo), *ser* (*ser* + participio pasado), *tener* (*tener* + participio pasado, *tener* (*a/de*) + infinitivo), *tornar* (*tornar a* + infinitivo) y *querer* (*querer* + infinitivo), pero no de *conocer*, *convenir*, *dar*, *decir*, *demandar*, *prender*, *ver* y *morder*. En segundo lugar, tampoco todos los verbos documentados en FCS-P pueden sincoparse (p. ej. *dar*, *demandar*, *tornar*) y bastantes otros lexemas que pueden tener FCS sincopados no se documentan en estructuras de FCS-P y, por tanto, solo aparecen con FCA, es decir, siempre con radical completo (*comprender*, *enardecer*, *aborreecer*, *abstener*, *entristecer*, *rehacer*, *comer*, *perder*, *venir*, *seguir*, *oír*). Todo ello sugiere que ni la perífrasis ni la síncopa logran explicar completamente la variación bajo estudio, sino que ofrecen meramente unas tendencias<sup>31</sup>. Además, llama la atención el hecho de que los tres verbos que han salido como particularmente proclives a los FCS-P en el análisis de regresión y el AGEM no sean ni siquiera los que exhiben los porcentajes más altos de FCS-P: mientras que *ser* sí aparece con un 80% (4/5) de FCS-P, *decir* y *hacer*, en cambio, lo hacen en un 15,6% (5/32) y un 18,6% (11/59) de los casos, respectivamente. Verbos como *querer* (100%, 5/5), *poder* (90,7%, 49/54) y *haber* (62,5%, 10/16) no salen en los modelos como significativamente propensos a los FCS-P. Esto se debe indudablemente a la decisión metodológica de haber descartado los casos sincopados tanto para la regresión como para el AGEM, debido a que, en nuestro corpus, todos los FCS-P de *querer*, *poder* y *haber* son sincopados. En vista de esto, podría darse el caso que ciertos verbos tiendan más que otros a emplearse en forma sincopada y que la frecuencia de este fenómeno se correlacione con su disposición a aparecer como FCS-P. Por eso, en el apartado que sigue (§4.4), indagaremos en nuestra última pista: los efectos de frecuencia, primero de la síncopa en los diferentes verbos (§4.4.1), y luego la frecuencia global de uso de los lexemas (§4.4.2).

#### 4.4 Efectos de frecuencia

##### 4.4.1 La frecuencia de la síncopa

Puesto que la síncopa parece presentarse como uno de los factores más decisivos en la variación entre FCA y FCS-P, quisimos comprobar si existe una correlación entre el uso de los FCS-P y la frecuencia de la síncopa en determinados verbos. Para esto, recopilamos en el OSTA las frecuencias relativas de formas sincopadas y asincopadas de FCS de todos los verbos de nuestros datos que permiten la síncopa, o sea, casi todos los lexemas con infinitivo de la 2ª o 3ª conjugación<sup>32</sup>. Estas frecuencias se presentan en

31 Recuérdense, en este sentido, las observaciones críticas acerca del papel de la síncopa formuladas en §2.1.

32 No hemos incluido en esta lista los verbos *abrir* y *adormir*, cuya configuración fonotáctica los hace candidatos poco idóneos para la síncopa en formas de FCS: sus radicales presentan, como secuencia final, o bien consonante + líquida (*abrir* > *abrirá* > \**abbrá*) o bien líquida + consonante (*adormir* > *adormirá* > \**adormrá*) (cf. Moreno Bernal 2004a: 132-33).

la segunda y tercera columna del Cuadro 3<sup>33</sup>. En gris claro se han marcado los verbos cuyos FCS sincopados predominan frente a los asincopados. Para fines de comparación, en la cuarta y quinta columna se repiten las frecuencias relativas de FCA y FCS-P para cada verbo en nuestro corpus, juntando los datos de los dos siglos tratados.

Lexema	FCS sincopados	FCS asincopados	FCA	FCS-P
<i>convenir</i>	100% (34/34)	0% (0/34)	33,3% (1/3)	66,7% (2/3)
<i>querer</i>	100% (497/497)	0% (0/497)	0% (0/5)	100% (5/5)
<i>tener</i>	100% (102/102)	0% (0/102)	85,7% (12/14)	14,3% (12/14)
<i>poder</i>	99,9% (1.914/1.916)	0,1% (2/1.916)	9,3% (5/54)	90,7% (49/54)
<i>haber</i>	99,4% (1.787/1.798)	0,6% (11/1.798)	37,5% (6/16)	62,5% (10/16)
<i>venir</i>	98,5% (193/196)	1,5% (3/196)	100% (2/2)	0% (2/2)
<i>prender</i>	76% (19/25)	24% (6/25)	66,7% (2/3)	33,3% (1/3)
<i>morder</i>	50% (1/2)	50% (1/2)	0% (0/1)	100% (1/1)
<i>comer</i>	40% (8/20)	60% (12/20)	100% (2/2)	0% (0/2)
<i>perder</i>	25% (12/48)	75% (36/48)	100% (2/2)	0% (0/2)
<i>decir</i>	1,6% (4/256)	98,4% (252/256)	84,4% (27/32)	15,6% (5/32)
<i>conocer</i>	0% (0/21)	100% (21/21)	66,7% (2/3)	33,3% (1/3)
<i>oír</i>	0% (0/35)	100% (35/35)	100% (2/2)	0% (2/2)
<i>seguir</i>	0% (0/39)	100% (39/39)	100% (3/3)	0% (3/3)
<i>ver</i>	0% (0/159)	100% (159/159)	50% (1/2)	50% (1/2)
<i>hacer</i>	0% (0/1.289)	100% (1.289/1.289)	81,4% (48/59)	18,6% (11/59)
<i>ser</i>	0% (0/3.318)	100% (3.318/3.318)	20% (2/5)	80% (4/5)
<b>Total</b>	<b>46,9%</b> <b>(4.571/9.755)</b>	<b>53,1%</b> <b>(5.184/9.755)</b>	<b>51,8%</b> <b>(117/226)</b>	<b>48,2%</b> <b>(109/226)</b>

Cuadro 3. La frecuencia de la síncopa con los diferentes verbos, comparada con la variación entre FCA y FCS-P

El Cuadro 3 destaca que los cinco verbos cuyo porcentaje de FCS sincopados supera el 99% son todos verbos documentados en nuestro corpus en estructuras de FCS-P

33 No incluimos en esta tabla los verbos cuya búsqueda en el OSTA no ha devuelto ningún caso de FCS: *aborrecer, abstener, comprender, enardecer, entristecer y rehacer*.

(*convenir, querer, tener, poder, haber*). El verbo *venir*, que también muestra porcentajes muy elevados de síncope (98,5%, 193/196), en cambio, no aparece como FCS-P en el corpus. Asimismo, el verbo *prender* también suele emplearse, por lo general, en forma sincopada, pero con un porcentaje relativamente más bajo que los otros verbos con mayoría de síncope (76%, 19/25). Surge una vez (33,3%, 1/3) como FCS-P en nuestros datos. También se puede observar que, en general, los lexemas caracterizados por un porcentaje alto de FCS-P suelen situarse en la parte superior del Cuadro 3 (*convenir, querer, poder, haber*), es decir, tender a la síncope. Una excepción a este patrón es *tener*, que aparece sobre todo en FCA, pese a la regularidad absoluta de la síncope con los FCS de este verbo. Lo anterior deja vislumbrar una correlación entre la frecuencia de la síncope y el uso de los FCS-P. Ahora bien, hay otros verbos documentados en nuestro corpus en construcciones de FCS-P que, no obstante, se emplean en los textos del OSTA en forma predominantemente asincopada (*decir, conocer, ver, hacer, ser*) o se caracterizan por un equilibrio entre variantes sincopadas y asincopadas (*morder*). Si miramos más de cerca estos verbos, destaca que algunos de ellos comparten un rasgo común: el radical de estos FCS asincopados es monosilábico (*dezir* > *dirá*<sup>34</sup>; *ver* > *verá*; *fazer* > *fará*; *ser* > *será*)<sup>35</sup>. Esto explicaría también por qué el verbo *dar* aparece con FCS-P en nuestro corpus (cf. Cuadro 1) a pesar de rechazar la síncope, como los demás verbos de la 1ª conjugación (*dar* > *dará*). Los verbos que no se conforman a este patrón son *conocer* (> *conocerá*) y *morder* (> *morderá*), que no tienen un radical monosilábico en forma asincopada y, a pesar de ello, se usan en el corpus en FCS-P.

En vista de lo anterior, parece que la síncope desempeña un papel sumamente relevante en la variación entre FCA y FCS-P. Como vimos anteriormente (§2.1), si bien es verdad que los FCS-P forman la única opción viable para las formas sincopadas, ya que estas son incompatibles con los FCA<sup>36</sup>, no se explica todavía por qué

34 Estas distribuciones parecen estar sujetas a variación diatópica, puesto que, en la *General Estoria IV*, un texto castellano occidental, la mayoría de FCS del verbo *decir* son claramente sincopados, como en *dizremos* por ejemplo (Moreno Bernal 2004b: 109-10).

35 Los verbos *ser* y *ver* pueden aparecer en los textos también bajo los infinitivos *seer, seyr, seyer* y *veer, veyr, veyer*, respectivamente, los cuales podrían, en teoría, dar lugar a FCS con radical bisilábico (*seerá, seyrá, seyerá; veerá, veyrá, veyerá*). Pese a esto, no hemos encontrado ningún caso de estas formas en los textos del OSTA y todos los FCS que hemos recogido exhiben un radical monosilábico *se-* o *ve-* (cf. Moreno Bernal 2004b sobre la *General Estoria IV* castellana). Cabe especificar que, en las búsquedas, el OSTA devuelve un caso de *seerán* en la *Crónica de conquistadores I*: «et que aquellos deuián seyer perseguidos/Los quales *seeran* desexidos del reyno delos persanos» (fol. 87v<sup>o</sup>). No obstante, hemos comprobado consultando el manuscrito (BNM, MSS/2211) que esta palabra aparece separada por un final de línea, con *se* al final de la primera línea y *eran* al inicio de la siguiente, por lo que esta secuencia podría interpretarse como la combinación de un pronombre reflexivo y el verbo *ser* en imperfecto de indicativo, es decir: *se eran*. Opinamos que esta lectura es más probable que la de futuro, ya que, dentro del contexto narrativo, este predicado hace claramente referencia a un evento pasado y, además, puede emplearse como reflexivo (Mackenzie 1984, s. v. *desexir*).

36 Esto es, fuera de los entornos de anteposición del pronombre (cf. §1).

el o la hablante decide emplear un FCS sincopado con un pronombre pospuesto en lugar de un FCA asincopado. Los datos que acabamos de exponer podrían ofrecer una pista para una respuesta relacionada con la diversa frecuencia de la síncopa en cuanto proceso fonológico con los diferentes verbos. El uso frecuente de ciertos verbos (p. ej. *poder*, *haber*) en FCS sincopados podría haber consolidado el arraigo del que gozan estas formas en la mente de los y las hablantes y relegado las variantes asincopadas como formas más bien marginales. Además, dado que, en general, los FCS suelen ser más frecuentes que los FCA (cf. Company 2006)<sup>37</sup>, a largo plazo, los y las hablantes habrían adoptado, en los entornos sintáctico-pragmáticos relevantes, la forma que más prominencia tiene en su sistema cognitivo, es decir, el FCS sincopado, necesariamente con pronombres pospuestos. En cuanto a los verbos con mayoría de FCS asincopados que, sin embargo, ocurren como FCS-P, hemos notado que se caracterizan por un radical monosilábico. Justamente, también encontramos un radical monosilábico en la mayoría de los verbos del Cuadro 3 en forma sincopada: *querer* > *querrá*; *tener* > *tenrá/terná/terrá/tendrá*; *poder* > *podrá*; *aver* > *avrá*; *venir* > *venrá/verná/verrá/vendrá*; *prender* > *prendrá*; *morder* > *mordrá*; *comer* > *combrá*; *perder* > *perdrá*; *seguir* > *sigrá*. Así, los FCS-P de verbos monosilábicos (p. ej. *dar*) se parecen a muchos de los FCS-P sincopados hallados en el corpus por el patrón acentual que exhiben, a saber: [σ'σσ], como en *dará lo* o *prendrá lo*, distinto a la acentuación de los FCA del tipo *prender lo á*, [σ'σσ'σ]. Esta similitud podría explicar que ambas estructuras evolucionen en la misma dirección, es decir, a favor de los FCS-P. El verbo *conocer*, por su parte, sí se documenta con los FCS-P a pesar de rechazar la síncopa en forma de FCS. Esto podría explicarse por la gradualidad con la que suelen proceder los cambios lingüísticos. Ya que el periodo abarcado por este estudio (siglos XIII-XIV) representa todavía la fase inicial del cambio, parece que, si los FCS-P se relacionan efectivamente con la síncopa y los radicales monosilábicos, se habría expandido primero con estos verbos cuya variante sincopada predomina más claramente frente a la asincopada. El FCS-P de *conocer* podría pues representar la fase incipiente de la etapa siguiente del cambio, en la que la forma innovadora se va extendiendo a más verbos y contextos.

#### 4.4.2 La frecuencia global de uso de los lexemas

Hemos visto que el uso de los FCS-P parece correlacionarse con el fuerte arraigo de la síncopa. En este último apartado, queremos examinar otro parámetro relacionado con los efectos de frecuencia, que concierne a la regularidad con la que se emplean ciertos lexemas globalmente en los textos medievales. En efecto, es una opinión relativamente bien asentada que los procesos de gramaticalización se correlacionan

37 Efectivamente, los FCS no solo pueden aparecer con pronombres antepuestos y pospuestos, lo que les da acceso a todos los posibles entornos sintáctico-pragmáticos (cf. §1), sino que también son las únicas formas posibles en la ausencia de un pronombre personal de objeto.

con una alta frecuencia de uso de los ítems implicados en ellos, aunque no existe en la actualidad un consenso sobre cómo nace exactamente esta correlación (cf. Hopper/Traugott 2003, Mair 2004, Bybee 2007, 2010, Enrique-Arias/Bouzouita 2013, Narrrog/Heine 2021: 177-78, inter alia). Algunos autores (p. ej. Hoffmann 2004) también han propuesto que una alta frecuencia de uso no es un requisito imprescindible de la gramaticalización y que el cambio puede producirse sin esta. El objetivo de los análisis que siguen es averiguar si los verbos que se documentan con los FCS-P en nuestro corpus se caracterizan por una frecuencia de uso especialmente alta, ya que la posposición pronominal se relaciona con un alto grado de gramaticalización. Para este fin, hemos extraído todas las ocurrencias de esos lexemas, es decir, en todos los modos y tiempos verbales en los textos navarroaragoneses de los siglos XIII-XIV en el OSTA que componen nuestro corpus y los hemos clasificado en orden de frecuencia decreciente. Los resultados de este procedimiento se presentan en el Cuadro 4<sup>38</sup>. En color gris claro se han marcado los lexemas documentados en el corpus de este trabajo como FCS-P.

Lexema	Frec. /1.000.000 pal.	Rango	Lexema	Frec. /1.000.000 pal.	Rango
<i>ser</i>	30563,86	#1	<i>poblar</i>	168,36	#33
<i>haber</i>	13913,01	#2	<i>mudar</i>	161,51	#34
<i>hacer</i>	9611,37	#3	<i>ajustar</i>	138,24	#35
<i>decir</i>	6864,63	#4	<i>menospre- ciar</i>	125,24	#36
<i>poder</i>	3832,50	#5	<i>cerrar</i>	123,19	#37
<i>querer</i>	3771,59	#6	<i>encalzar</i>	118,05	#38
<i>venir</i>	3015,01	#7	<i>heredar</i>	105,39	#39
<i>dar</i>	3012,28	#8	<i>alegrar</i>	101,29	#40
<i>ver</i>	2477,93	#9	<i>emendar</i>	96,50	#41
<i>tener</i>	1914,54	#10	<i>pechar</i>	89,65	#42
<i>pasar</i>	1504,94	#11	<i>abrir</i>	85,89	#43
<i>clamar</i>	1492,96	#12	<i>pesar</i>	75,63	#44
<i>tornar</i>	1390,99	#13	<i>recordar</i>	51,67	#45
<i>matar</i>	1207,92	#14	<i>semejar</i>	46,88	#46

38 Nótese que las frecuencias proporcionadas en el Cuadro 4 resultan únicamente de los números de formas de FC presentes en los textos del OSTA, divididos por las cantidades de palabras de esos mismos textos. Es decir, hemos descartado las formas incluidas en nuestro corpus base provenientes de textos no representados en el OSTA, así como las cantidades de palabras de estos textos.

<i>llevar</i>	1060,78	#15	<i>quitar</i>	40,38	#47
<i>demandar</i>	983,45	#16	<i>rehacer</i>	25,32	#48
<i>prender</i>	982,42	#17	<i>comprender</i>	21,90	#49
<i>trovar</i>	859,23	#18	<i>morder</i>	16,08	#50
<i>echar</i>	658,37	#19	<i>abstener</i>	13	#51
<i>mostrar</i>	597,12	#20	<i>aborrecer</i>	11,63	#53
<i>perder</i>	567	#21	<i>adormir</i>	11,63	#53
<i>seguir</i>	517,05	#22	<i>mojar</i>	8,21	#54
<i>comer</i>	506,78	#23	<i>gozar</i>	4,79	#55
<i>oír</i>	463,66	#24	<i>ahinojar</i>	3,42	#57
<i>ayudar</i>	377,43	#25	<i>rentar</i>	3,42	#57
<i>pagar</i>	333,97	#26	<i>perfumar</i>	2,39	#58
<i>conocer</i>	320,29	#27	<i>acaloñar</i>	1,37	#59
<i>sacar</i>	319,94	#28	<i>enardecer</i>	1,03	#62
<i>convenir</i>	294,28	#29	<i>enseñorear</i>	1,03	#62
<i>posar</i>	201,21	#30	<i>entristecer</i>	1,03	#62
<i>pendrar</i>	186,15	#31	<i>sahumar</i>	0,68	#63
<i>esforzar</i>	175,20	#32	<i>cobijar</i>	0,34	#64

Cuadro 4. La frecuencia de uso de los lexemas en los textos navarroaragoneses del OSTA (s. XIII-XIV)

Como se puede ver en el Cuadro 4, los lexemas más frecuentes en los textos del OSTA que constituyen nuestro corpus coinciden, a grandes rasgos, con algunos de los verbos que se emplean en estructuras de FCS-P. Concretamente, de los diez verbos con el mayor número de ocurrencias en total (*ser*, *haber*, *hacer*, *decir*, *poder*, *querer*, *venir*, *dar*, *ver* y *tener*), nueve aparecen en nuestros datos de FCS-P; solo *venir* no entra en este patrón. De los demás verbos documentados como FCS-P, también *tornar*, *demandar* y *prender* pueden contarse dentro de los verbos con una frecuencia de uso no desdeñable<sup>39</sup>. Los lexemas *conocer*, *convenir* y sobre todo *morder*, en cambio, ocurren menos frecuentemente. No obstante, recordemos que el 100% de FCS-P registrado con el verbo *morder* resulta de un solo ejemplo, un dato que invita a interpretar estas cifras con cautela. Obsérvese, además, que los verbos que tienen un radical monosilábico en sus FCS asincopados (*dar*, *decir*, *hacer*, *ser*, *ver*) también se cuentan dentro de los verbos globalmente más frecuentes. De estos datos podemos

39 También Bouzouita (2016) identifica *tornar* como uno de los verbos más propicios a aparecer como FCS-P en su corpus basado en E6.

tentativamente concluir lo siguiente: el uso de los FCS-P parece correlacionarse con una alta frecuencia de uso de los lexemas específicos. Esta conclusión corrobora así la idea común de que los procesos de gramaticalización vienen impulsados por una frecuencia elevada de los elementos lingüísticos que participan en ellos.

En resumen, parece que los verbos caracterizados por una alta frecuencia de FCS-P frente a los FCA en nuestro corpus son sobre todo verbos que se usan habitualmente en formas de FCS sincopadas o cuyos FCS asincopados exhiben un radical monosilábico. Así pues, este hallazgo sugiere que el efecto del lexema que parece surgir en nuestros datos podría, en realidad, deberse a otro factor, a saber, la síncopa verbal. Además, se confirman las tesis de Bybee (2010), ya que, en opinión de esta autora, una alta frecuencia de los entornos favorecedores del cambio puede llegar a acelerar el cambio mismo. Esta propuesta encaja bien con las observaciones que hemos presentado: una alta frecuencia de uso de la síncopa en los FCS de algunos verbos, con o sin pronombre átono de objeto, lleva a una mayor proclividad al uso de los FCS-P (sincopados). Aun así, queda inexplicado el uso de los FCS-P con *conocer*, *demandar*, *morder* y *tornar*, pero no podemos perder de vista que solo hemos registrado un FCS-P de cada uno de estos verbos (cf. Cuadro 1). Aparte de esto, hemos visto también que en nuestro corpus los verbos que aparecen en FCS-P suelen ser verbos muy frecuentes globalmente, lo cual parece confirmar que son los ítems de alta frecuencia los que impulsan este proceso de gramaticalización, como se ha notado en otros desarrollos diacrónicos.

## 5. Conclusiones y perspectivas futuras

En este apartado final, recopilamos las principales conclusiones de los análisis efectuados y ofrecemos respuestas (parciales) a las preguntas de investigación formuladas en §3:

1. La síncopa verbal se perfila como el factor que más influye en la variación entre FCA y FCS-P, puesto que, cuando ocurre, se elimina cualquier posibilidad de variación, quedando solo formas de FC sintéticas (cf. Moreno Bernal 2004a, Matute/Pato 2010, Bouzouita 2016, Primerano/Bouzouita en prensa).
2. Dentro del conjunto de formas asincopadas de nuestro corpus, hemos encontrado un efecto significativo para la variable de la perífrasis: los FCS-P son más probables cuando el FC funciona como auxiliar de una perífrasis o como verbo causativo (cf. Primerano/Bouzouita en prensa). El factor del lexema parece influir significativamente en la variación, ya que los verbos *decir*, *hacer* y *ser* salen como particularmente propicios a aparecer como FCS-P. Sin embargo, la ausencia de formas sincopadas en el análisis de regresión y el AGEM llevó a la exclusión de ciertos verbos que tienden claramente a los FCS-P.

3. Hemos comprobado que la variación bajo estudio parece responder también a efectos de frecuencia: los FCS-P parecen ser más comunes con estos verbos que exhiben los porcentajes más altos de síncopa y con los lexemas caracterizados por una frecuencia de uso global elevada.
4. Los FCS-P parecen preferir, además de los verbos sincopados, aquellos verbos cuyo radical en los FCS es monosilábico, lo cual podría reflejar una preferencia por cierto patrón acentual, a saber, [σ'σσ].
5. Por lo tanto, el efecto del lexema podría ser un epifenómeno de la configuración morfológica y/o acentual de los verbos que ocurren en estructuras de FCS-P y/o de la elevada frecuencia de uso de estos verbos en general (cf. Bybee 2010).

Antes de finalizar, quisiéramos evaluar de forma crítica las ventajas e inconvenientes de la metodología puesta en práctica en este trabajo. Como hemos visto a lo largo del artículo, utilizar un conjunto de datos de tamaño reducido, bastante local, plantea problemas a la hora de aplicar modelos estadísticos complejos, pues no siempre se dispondrá de suficientes muestras como para hacer predicciones fiables. Sin embargo, queremos subrayar la dificultad de componer un corpus amplio y representativo para el estudio de fenómenos morfosintácticos en la diacronía, en especial cuando estos se caracterizan por una baja frecuencia de uso (cf. §4.1). Recuérdese, efectivamente, que las 344 ocurrencias aquí manejadas se han recogido recorriendo un corpus de más de 3 millones de palabras. Además, al trabajar con una lista de lexemas todavía manejable (64 verbos), se hacen posibles análisis profundos y detallados en los que se pueden vincular datos de naturaleza variada, como, por ejemplo, la frecuencia relativa de los FCS-P frente a los FCA para un verbo  $\alpha$  en relación con la preferencia o no del mismo por la síncopa verbal en todas sus formas de FCS. En conclusión, con tiempo y espacio, puede resultar muy provechoso, en los estudios de corpus, la unión de metodologías a mayor y menor escala.

Puesto que por el momento manejamos datos exclusivamente navarroaragoneses, en un futuro sería interesante replicar los análisis presentados aquí con datos provenientes de otras variedades iberorrománicas antiguas (catalán, castellano, asturiano, leonés, gallego y portugués), con el fin de averiguar si los mismos factores influyen en la variación bajo estudio.

## Fuentes bibliográficas

- BAUTISTA, F. 2017: «Anales navarroaragoneses hasta 1239», *e-Spania*. URL: <https://journals.openedition.org/e-spania/26509> [28.09.2023]
- CODEA+ 2015 = GITHE (GRUPO DE INVESTIGACIÓN TEXTOS PARA LA HISTORIA DEL ESPAÑOL): *CODEA+ 2015 (Corpus de documentos españoles anteriores a 1800)*. URL: <https://www.corpuscodea.es/> [28.09.2023]
- COOPER, L. 1960: *El Liber Regum. Estudio lingüístico*, Zaragoza, Institución Fernando el Católico.
- GOROSCH, M. 1950: *El Fuero de Teruel*, Estocolmo, Almqvist och Wiksell.
- NAVARRO TOMÁS, T. 1957: *Documentos lingüísticos del Alto Aragón*, Syracuse University Press.
- ORCÁSTEGUI GROS, C. 1986: «Crónica de San Juan de la Peña (versión aragonesa): edición crítica», *Cuadernos de historia Jerónimo Zurita* 51-52: 419-569.
- OSTA = GAGO JOVER, F./PUEYO MENA F. J. 2020: *Old Spanish Textual Archive*, Hispanic Seminary of Medieval Studies. URL: <http://osta.oldspanishtextualarchive.org/> [28.09.2023]

## Referencias bibliográficas

- ALFONSO VEGA, M. 2006: «Verbos causativos», in: C. COMPANY COMPANY (ed.), *Sintaxis histórica de la lengua española*, México, Fondo de Cultura Económica/UNAM: 969-1052.
- ALVAR, M. 1953: *El dialecto aragonés*, Madrid, Gredos.
- BARTON, K. 2022: MuMIn: Multi-Model Inferencing. R package version 1.47.1. URL: <https://CRAN.R-project.org/package=MumIn> [28.09.2023]
- BATES, D./MAECHLER, M./BOLKER, B./WALKER, S. 2015: «Fitting linear mixed-effects models using lme4», *Journal of Statistical Software* 67: 1-48.
- BATLLORI, M. 2016: «El valor modal de *haber* en los futuros y condicionales analíticos», in: C. DE BENITO MORENO/Á. OCTAVIO DE TOLEDO Y HUERTA (ed.), *En torno a 'haber': construcciones, usos y variación desde el latín hasta la actualidad*, Fráncfort, Peter Lang: 33-78.
- BATLLORI, M. 2021: «La estructura informativa del español antiguo», in: M. LEONETTI/M. V. ESCANDELL-VIDAL (ed.), *La estructura informativa*, Madrid, Visor Libros: 217-49.
- BERNAISCH, T. 2022: «Comparing generalised linear mixed-effects models, generalised linear mixed-effects model trees and random forests. Filled and unfilled pauses in varieties of English», in: O. SCHÜTZLER/J. SCHLÜTER (ed.), *Data and methods in corpus linguistics. Comparative approaches*, Cambridge, Cambridge University Press: 163-93.
- BOUZOUITA, M. 2008a: *The diachronic development of Spanish clitic placement*, Tesis doctoral, King's College London.
- BOUZOUITA, M. 2008b: «At the syntax-pragmatics interface: clitics in the history of Spanish», in: R. COOPER/R. KEMPSON (ed.), *Language in flux: dialogue coordination, language variation, change and evolution*, Londres, College Publications: 221-63.
- BOUZOUITA, M. 2011: «Future constructions in medieval Spanish: mesoclisism uncovered», in: R. KEMPSON/E. GREGOROMICHELAKI/C. HOWES (ed.), *The dynamics of lexical interfaces*, Stanford, CSLI: 89-130.
- BOUZOUITA, M. 2012: «Los futuros analíticos y sintéticos en la 'Fazienda de Ultra Mar'», in: E. MONTERO CARTELE/C. MANZANO ROVIRA (ed.), *Actas del VIII Congreso Internacional de Historia de la Lengua Española*, Santiago de Compostela, Meubook: 695-706.
- BOUZOUITA, M. 2013: «La influencia latinizante en el uso del futuro en la traducción bíblica del código Escorial I.i.6», in: E. CASANOVA HERRERO/C. CALVO RIGUAL (ed.), *Actes del 26è Congrés Internacional de Lingüística i Filologia Romàniques*, Berlín, De Gruyter: 353-64.
- BOUZOUITA, M. 2016: «La posposición preposicional con futuros y condicionales en el código escorialense I.i.6: un examen de varias hipótesis morfosintácticas», in: J. KABATEK (ed.), *Lingüística de corpus y lingüística histórica iberorrománica*. Berlín, De Gruyter: 271-301.

- BOUZOUTA, M. 2017: «La influencia innovadora del occitano en la gramaticalización del futuro y del condicional en las lenguas iberorrománicas», presentado en el *Congrès de l'association internationale d'études occitanes*, Albi, 10-15 de julio.
- BOUZOUTA, M./SENTÍ, A. 2022: «La gramaticalización del futuro y el condicional en el iberorromance del siglo XIV a partir de traducciones bíblicas paralelas: el caso del castellano y el catalán antiguos», in: A. ENRIQUE-ARIAS (ed.), *Traducción bíblica e historia de las lenguas iberorrománicas*, Berlín/Boston, De Gruyter: 243-73.
- BOUZOUTA, M./SITARIDOU, I. 2023: «Cambios en el orden de palabras y la estructura informativa», in: S. N. DWORKIN/G. CLAVERÍA NADAL/Á. S. OCTAVIO DE TOLEDO Y HUERTA (ed.), *Lingüística histórica del español/The Routledge Handbook of Spanish Historical Linguistics*, Londres, Routledge: 221-32.
- BUESA, T. 1980: «Estado actual de los estudios sobre el dialecto aragonés», in: A. UBIETO ARTETA (ed.), *Estado actual de los estudios sobre Aragón*, vol. 1, Zaragoza, Cometa: 355-400.
- BYBEE, J. 2007: *Frequency of use and the organization of language*, Oxford, Oxford University Press.
- BYBEE, J. 2010: *Language, usage and cognition*, Cambridge, Cambridge University Press.
- CASTILLO LLUCH, M. 1996: *La posición del pronombre átono en la prosa hispánica medieval*, Tesis doctoral, Universidad Autónoma de Madrid/Université Paris XIII.
- CASTILLO LLUCH, M. 2002: «Distribución de las formas sintéticas y analíticas de futuro y condicional en español medieval», in: M. T. ECHENIQUE ELIZONDO/J. P. SÁNCHEZ MÉNDEZ (ed.), *Actas del V Congreso Internacional de Historia de la Lengua Española*, Madrid, Gredos: 541-50.
- CASTILLO LLUCH, M. 2004: «L'emphase et la syntaxe du pronom personnel atone en espagnol médiéval dans une perspective romane», in: M. H. ARAÚJO CARREIRA (ed.), *Plus ou moins?! L'atténuation et l'intensification dans les langues romanes*, Saint-Denis, Université Paris 8 Vincennes Saint-Denis: 343-56.
- COMPANY COMPANY, C. 2006: «Tiempos de formación romance II: los futuros y condicionales», in: C. COMPANY COMPANY (ed.), *Sintaxis histórica de la lengua española. Primera parte: la frase verbal*, vol. 1, México, Fondo de Cultura Económica/UNAM: 349-422.
- COMPANY COMPANY, C./MEDINA URREA, A. 1999: «Sintaxis motivada pragmáticamente. Futuros analíticos y futuros sintéticos en el español medieval», *Revista de Filología Española* 79: 65-100.
- DUBERT-GARCÍA, F. 2021: «Las formas reducidas de *facere* en el noroeste de la península ibérica. Convergencia y divergencia geolingüísticas», *Studia Lingüística Romanica* 6: 163-91.
- EBERENZ, R. 1991: «Futuro analítico y futuro sintético en tres obras con rasgos coloquiales: El 'Corbacho', 'La Celestina' y 'La Lozana Andaluza'», in: R. LAPESA (ed.), *Homenaje a Hans Flasche: Festschrift zum 80. Geburtstag 25. November 1991*, Stuttgart, Franz Steiner Verlag: 499-508.
- ENGUITA UTRILLA, J. M. 2008: «Evolución lingüística en la Baja Edad Media: aragonés, navarro», in: R. CANO (ed.), *Historia de la lengua española*, 2ª ed., Barcelona, Ariel: 571-92.
- ENRIQUE-ARIAS, A. (ed.) 2010: *La Biblia Escorial Li.6. Transcripción y estudios*, Logroño, Biblias Hispánicas.
- ENRIQUE-ARIAS, A./BOUZOUTA, M. 2013: «La frecuencia textual en la evolución histórica de los clíticos pronominales en español», *Iberoromania* 77: 29-46.
- FERNÁNDEZ-ORDÓÑEZ, I. 2006: «La historiografía medieval como fuente de datos lingüísticos. Tradiciones consolidadas y rupturas necesarias», in: J. J. DE BUSTOS TOVAR/J. L. GIRÓN ALCONCHEL (ed.), *Actas del VI Congreso Internacional de Historia de la Lengua Española*, Madrid, Visor Libros: 1770-807.
- FERNÁNDEZ-ORDÓÑEZ, I. 2011: *La lengua de Castilla y la formación del español*, Madrid, RAE.
- FLEISCHMAN, S. 1982: *The future in thought and language. Diachronic evidence from Romance*, Cambridge, Cambridge University Press.
- FOKKEMA, M./EDBROOKE-CHILDS, J./WOLPERT, M. 2021: «Generalized linear mixed-model (GLMM) trees: a flexible decision-tree method for multilevel and longitudinal data», *Psychotherapy Research* 31: 329-41.

- FOKKEMA, M./SMITS, N./ZEILEIS, A./HOTHORN, T./KELDERMANN, H. 2018: «Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees», *Behavior Research Methods* 50: 2016-34.
- GAGO JOVER, F./PUEYO MENA, F. J. 2018a: «El *Old Spanish Textual Archive*, diseño y desarrollo de un corpus de textos medievales: lematización y etiquetado gramatical», *Scriptum digital* 7: 25-35.
- GAGO JOVER, F./PUEYO MENA, F. J. 2018b: «El *Old Spanish Textual Archive*, diseño y desarrollo de un corpus de textos medievales: el corpus textual», *Cuadernos del Instituto Historia de la Lengua* 11: 165-209.
- GÓMEZ TORREGO, L. 1999: «Los verbos auxiliares. Las perífrasis verbales de infinitivo», in: I. BOSQUE/V. DEMONTE (ed.), *Gramática descriptiva de la lengua española*, Madrid, Espasa: 3323-89.
- GONZÁLEZ-OLLÉ, F. 1970: «El romance navarro», *Revista de Filología Española* 53: 45-93.
- GONZÁLEZ-OLLÉ, F. 1983: «Evolución y castellanización del romance navarro», *Príncipe de Viana* 44: 173-80.
- GONZÁLEZ-OLLÉ, F. 1996: «Navarro», in: M. ALVAR (ed.), *Manual de dialectología hispánica: el español de España*, Barcelona, Ariel: 305-16.
- GRAHAM, L. 2018: «An analysis of morphosyntactic variation in the Old Spanish future and conditional», *Journal of Historical Linguistics* 8: 192-229.
- GRANBERG, R. 1988: *Object pronoun position in medieval and early modern Spanish*, Tesis doctoral, University of California.
- GRANBERG, R. 1999: «Clitic position in thirteenth-century Spanish: sentences with preverbal subjects», *La corónica* 27: 89-113.
- GRIES, S. Th. 2015: «The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models», *Corpora* 10: 95-125.
- GRIES, S. Th. 2021: *Statistics for linguistics with R. A practical introduction*, Berlín/Boston, Mouton de Gruyter.
- HARRELL, F. Jr. 2023: Hmisc: Harrell Miscellaneous. R package version 4.8-0. URL: <https://CRAN.R-project.org/package=Hmisc> [28.09.2023]
- HASPELMATH, M. 2011: «The gradual coalescence into ‘words’ in grammaticalization», in: H. NARROG/B. HEINE (ed.), *The Oxford Handbook of Grammaticalization*, Oxford, Oxford University Press: 342-55.
- HIMMELMANN, N. P. 2004: «Lexicalization and grammaticalization: opposite or orthogonal?», in: W. BISANG/N. P. HIMMELMANN/ B. WIEMER (ed.), *What makes grammaticalization? A look from its fringes and its components*, Berlín/Nueva York, Mouton de Gruyter: 21-42.
- HOFFMANN, S. 2004: «Are low-frequency complex prepositions grammaticalized? On the limits of corpus data – and the importance of intuition», in: H. LINDQUIST/C. MAIR (ed.), *Corpus approaches to grammaticalization in English*, Ámsterdam/Filadelfia, John Benjamins: 171-210.
- HOPPER, P. 1991: «On some principles of grammaticization», in: E. TRAUOGOTT/B. HEINE (ed.), *Approaches to Grammaticalization*, vol. I. Ámsterdam/Filadelfia, John Benjamins: 17-35.
- HUALDE, J. I./OLARREA, A./ESCOBAR, A. M./TRAVIS, C. E. 2010: *Introducción a la lingüística hispánica*, Cambridge, Cambridge University Press.
- HOPPER, P./TRAUGOTT, E. 2003: *Grammaticalization*, Cambridge, Cambridge University Press.
- KUHN, M. 2022: caret: Classification and Regression Training. R package version 6.0-93. URL: <https://CRAN.R-project.org/package=caret> [28.09.2023]
- LEHMANN, C. 1985: «Grammaticalization. Synchronic variation and diachronic change», *Lingua e Stile* 20: 303-18.
- LEHMANN, C. 2020: «Univerbation», *Folia Linguistica Historica* 41: 205-52.
- LEMA, J./RIVERO, M. L. 1991: «Types of verbal movement in Old Spanish: modals, futures and perfects», *Probus* 3: 237-78.
- MACKENZIE, J. G. 1984: *A lexicon of the 14<sup>th</sup>-century Aragonese manuscripts of Juan Fernández de Heredia*, Madison, Hispanic Seminary of Medieval Studies.

- MAIR, C. 2004: «Corpus linguistics and grammaticalisation theory: statistics, frequency and beyond», in: H. LINDQUIST/C. MAIR (ed.): *Corpus approaches to grammaticalization in English*, Amsterdam/Filadelfia, John Benjamins: 121-50.
- MARTINS, A. M. 2016: «A colocação dos pronomes clíticos em sincronia e diacronia», in: A. M. MARTINS/E. CARRILHO (ed.), *Manual de linguística portuguesa*, Berlín/Boston, De Gruyter: 401-30.
- MARTTINEN LARSSON, M. 2023: «Modelling incipient probabilistic grammar change in real time: the grammaticalisation of possessive pronouns in European Spanish locative adverbial constructions», *Corpus Linguistics and Linguistic Theory* 19: 177-206.
- MATUTE, C./PATO, E. 2010: «Morfología y sintaxis en el códice Escorial Ii.6», en A. ENRIQUE-ARIAS (ed.), *La Biblia Escorial Ii.6. Transcripción y estudios*, Logroño, Biblias Hispánicas: 45-66.
- MORENO BERNAL, J. 2004a: «La morfología de los futuros románicos. Las formas con metátesis», *Revista de Filología Románica* 21: 121-69.
- MORENO BERNAL, J. 2004b: «Estudio lingüístico de la ‘Historia de Nabucodonosor’ (General estoria, cuarta parte) de Alfonso X», *Revista de Filología Románica* 21: 81-119.
- NARROG, H./HEINE B. 2021: *Grammaticalization*, Oxford, Oxford University Press.
- NIJEUWENHUIJSEN, D. 2002: «Variación de colocación de los pronombres átonos en el español antiguo», *Zeitschrift für romanische Philologie* 118: 360-75.
- NIJEUWENHUIJSEN, D. 2006: «La colocación de los pronombres átonos», in: C. COMPANY COMPANY (ed.), *Sintaxis histórica de la lengua española*, México, Fondo de Cultura Económica/UNAM: 1339-404.
- NOCENTINI, A. 2001: «La genesi del futuro e del condizionale sintetico romanzo», *Zeitschrift für romanische Philologie* 117: 376-401.
- OCTAVIO DE TOLEDO y HUERTA, Á. S. 2015: «Futuros que se miran el ombligo: mesoclisys y anteposición de formas no personales en la historia del español», in: M. CASTILLO LLUCH/M. LÓPEZ IZQUIERDO (ed.), *El orden de palabras en la historia del español y otras lenguas iberorromances*, Madrid, Visor Libros: 141-233.
- PATO, E. 2018: «¿Cómo se ha escrito la historia de la lengua española? Sobre los textos medievales y clásicos como fuente de datos lingüísticos», *Hápax: Revista de la Sociedad de Estudios de Lengua y Literatura* 11: 59-116.
- PONS RODRÍGUEZ, L. 2006: «Canon, edición de textos e historia de la lengua cuatrocentista», in: L. PONS RODRÍGUEZ (ed.), *Historia de la lengua y crítica textual*, Madrid/Fráncfort, Iberoamericana/Vervuert: 69-125.
- PRIMERANO, A./BOUZOUITA, M. (en prensa): «La gramaticalización de los futuros y condicionales en el navarroaragonés de los siglos XIII-XIV: una comparación con el castellano medieval», *Revue Romane*.
- PRIMERANO, A./TORRES-LATORRE, A./SENTÍ, A./BOUZOUITA, M. (en prensa): «La gramaticalización del futuro y el condicional en iberorromance antiguo: castellano y catalán en contraste», in: J. R. MORALA/C. EGIDO FERNÁNDEZ/A. JUNQUERA MARTÍNEZ (ed.), *Actas del XII Congreso Internacional de Historia de la Lengua Española*, León, Universidad de León.
- R CORE TEAM 2022: *R: a language and environment for statistical computing*, Viena, R Foundation for Statistical Computing.
- RODRÍGUEZ MOLINA, J. 2010: *La gramaticalización de los tiempos compuestos en español antiguo: cinco cambios diacrónicos*, Tesis doctoral, Universidad Autónoma de Madrid.
- SARALEGUI, C. 1977: *El dialecto navarro en los documentos del monasterio de Irache (958-1397)*, Pamplona, Institución Príncipe de Viana.
- SARALEGUI, C. 1983: «Morfología del futuro y condicional castellanos: polimorfismo antiguo y fijación lingüística», *Medioevo romanzo* 8: 419-59.
- SARALEGUI, C. 1992: «Aragónés/Navarro. Evolución lingüística externa e interna», in: G. HOLTUS (ed.), *Lexikon der romanistischen Linguistik. 6.1: Aragonesisch-Navarresisch, Spanisch, Asturiansch-Leonisch*, Tübinga, Niemeyer Verlag: 37-54.
- SENTÍ, A./BOUZOUITA, M. 2022: «The syntax of Old Catalan clitics: *Llibre dels Fets*», *Catalan Journal of Linguistics* 21: 47-77.

- TAGLIAMONTE, S. A./BAAYEN, R. H. 2012: «Models, forests, and trees of York English: *was/were* variation as a case study for statistical practice», *Language Variation and Change* 24: 135-78.
- TORRES-LATORRE, A. 2023: «La gramaticalització del futur i el condicional al català del segle XIII: formes analítiques i formes sintètiques», *Zeitschrift für Katalanistik* 36: 63-85.
- TORRES-LATORRE, A./SENTÍ, A. 2023: «Clitic placement and the grammaticalization of the future and the conditional in Old Catalan», *Language* 8: 1-20.
- YNDURÁIN, F. 1945: *Contribución al estudio del dialecto navarro-aragonés antiguo*, Zaragoza, Institución Fernando el Católico.
- ZUUR, A. F./LENO, E. N./WALKER, N./SAVELIEV, A. A./SMITH, G. M. 2009: *Mixed effects models and extensions in ecology with R*, Berlín/Nueva York, Springer.

## Apéndice: corpus de estudio

### Textos incluidos en el OSTA

- |  |  |
|--|--|
| 1. <i>Fueros de Aragón</i>                       | 16. <i>Grant Crónica de Espanya</i>                |
| 2. <i>Libro de Apolonio</i>                      | 17. <i>Guerra del Peloponeso</i>                   |
| 3. <i>Libro de los tres reyes de oriente</i>     | 18. <i>Historia contra los paganos</i>             |
| 4. <i>Vida de Santa María Egipcíaca</i>          | 19. <i>Historia de Eutropio</i>                    |
| 5. <i>Fuero General de Navarra</i> <sup>40</sup> | 20. <i>Historia de los godos</i>                   |
| 6. <i>Vidal Mayor</i>                            | 21. <i>Historia Troyana</i>                        |
| 7. <i>Crónica de los conquistadores</i>          | 22. <i>Libro de autoridades/Rams de Flors</i>      |
| 8. <i>Crónica de los conquistadores</i>          | 23. <i>Libro de Marco Polo</i>                     |
| 9. <i>Crónica de los emperadores</i>             | 24. <i>Obra de agricultura</i>                     |
| 10. <i>Crónica de Morea</i>                      | 25. <i>Viajes de Juan de Mandevilla</i>            |
| 11. <i>Crónica de San Isidoro</i>                | 26. <i>Vidas semblantes</i>                        |
| 12. <i>Crónica del tudense</i>                   | 27. <i>Vidas semblantes</i>                        |
| 13. <i>De Secreto Secretorum</i>                 | 28. <i>Vidas semblantes</i>                        |
| 14. <i>Flor de las historias de oriente</i>      | 29. <i>Crónica general de España</i> <sup>41</sup> |
| 15. <i>Grant Crónica de Espanya</i>              |  |

40 No hemos consultado los múltiples testimonios del *Fuero General de Navarra* incluidos en el OSTA por tratarse de copias de los siglos XIII y XIV de un mismo texto. Para este estudio, hemos empleado solo el manuscrito BNM, MSS/17653, representante de la versión A del texto. Ese códice también incluye el texto etiquetado como *Linaje de Rodrigo Díaz el Campeador* en el OSTA.

41 El códice de este texto (El Escorial, X-II-22) también contiene el texto etiquetado como *Genealogía de los reyes de Navarra* en el OSTA.

## Otros textos

1. *Liber Regum*<sup>42</sup>
2. *Anales navarroaragoneses hasta 1239*
3. *Fuero de Teruel*
4. *Crónica de San Juan de la Peña*
5. *CODEA+ 2015*
6. *Documentos lingüísticos del Alto Aragón*

---

42 En el momento en que se efectuó la recogida de datos, el *Liber Regum* no estaba incluido en el OSTA.

## The verbal lexeme and other variables in the grammaticalization of the Old Navarro-Aragonese future and conditional

**Abstract:** This paper deals with the Old Ibero-Romance morphosyntactic variation between analytic future and conditional forms with mesoclitic pronouns (e.g., *cantar lo é*, *cantar lo ía* ‘I will/would sing it’), on the one hand, and synthetic forms with postverbal pronouns (e.g., *cantaré lo*, *cantaría lo* ‘I will/would sing it’), on the other. Because of insufficient systematic research on this variation, little is known about the linguistic factors that condition it and influence the choice between one form and the other. Although a preliminary study (Primerano/Bouzouita in press) examines, using a bivariate methodology, the possible effect of three morphosyntactic variables, to wit, syncope, verb tense, and non-finite verb/periphrastic construction, the influence of the verb lexeme on the variation has not been analysed yet due to the small size of the corpus. In view of this, the present research aims to expand Primerano/Bouzouita’s study (in press) in order to provide quantitative analyses which measure the effect of the lexeme variable, as well as its interaction with others that have been deemed of importance in the variation under study. Further, we will reassess the role of these variables by means of more refined multivariate quantitative methods and investigate possible frequency effects. This paper makes use of data drawn from 13<sup>th</sup> and 14<sup>th</sup> century Navarro-Aragonese texts available in a relatively recent online database, the *Old Spanish Textual Archive* (OSTA). The results of these analyses suggest that syncope and the presence of a periphrastic construction are the most relevant variables for the variation at hand. The lexeme variable equally seems to play a role, which could be epiphenomenal to the regularity of syncope with certain verbs and the frequency of use of the different lexemes.

**Keywords:** future, conditional, mesoclisism, postposition, lexeme, syncope, frequency, Navarro-Aragonese