



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2020

Reflexiones sobre la “lengua vulgar dialectal” y el vulgarismo

De Benito Moreno, Carlota

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-196342>

Book Section

Accepted Version

Originally published at:

De Benito Moreno, Carlota (2020). Reflexiones sobre la “lengua vulgar dialectal” y el vulgarismo. In: Fernández-Ordoñez, Inés. El legado de Ramón Menéndez Pidal (1869-1968) a principios del siglo XXI. Madrid: Libros-CSIC, 19-56.

REFLEXIONES SOBRE LA «LENGUA VULGAR DIALECTAL» Y EL VULGARISMO*

Carlota de BENITO MORENO**

Universidad de Zúrich

RESUMEN: A partir del concepto pidalino de «lengua vulgar dialectal», en este trabajo analizamos cuatro fenómenos del español vulgar (el uso de la preposición *de* en los adjuntos temporales que expresan la edad, la concordancia singular con los numerales complejos acabados en *un(a)*, el orden de los pronombres en las secuencias de clíticos y las formas analógicas del presente de subjuntivo de *haber*) con el objetivo de comprender mejor su evolución histórica y, así, las dinámicas sociolingüísticas que presentan los vulgarismos.

Palabras clave: Menéndez Pidal, vulgarismo, español vulgar, *haiga*, secuencias de clíticos.

ABSTRACT: Taking Menéndez Pidal's concept of «dialectal vulgar speech» as a point of departure, I analyse four features of vulgar Spanish (namely, the use of preposition *de* in temporal adjuncts to express age, singular agreement with complex numerals that end in *un(a)*, the order of pronouns in clitic clusters and the analogical forms of the present subjunctive of *haber*) with the goal of analysing their historical evolution as well as the sociolinguistic dynamics that vulgar features show.

Keywords: Menéndez Pidal, vulgarism, vulgar Spanish, *haiga*, clitic clusters.

1. LA LENGUA VULGAR DIALECTAL

En el cajón número 18 del fondo documental de la *Historia de la lengua española* de Menéndez Pidal encontramos todo un apartado dedicado a la «lengua vulgar dialectal», en el que don Ramón agrupaba sus notas sobre una gran variedad de cuestiones, desde el seseo y el yeísmo («los dos grandes sincretismos fonéticos»¹) hasta notas sobre las variadas lenguas españolas, incluyendo el vasco. En este trabajo me interesa sobre todo la cuestión a la que don Ramón asigna el membrete de «Vulgarismos comunes a España y América». En dicho apartado se

*Debo agradecer a Maxine Hofstetter su ayuda tanto con el cribado de los datos como con las pesquisas bibliográficas: sin dicha ayuda este trabajo nunca hubiera llegado a término a tiempo. También le doy las gracias a Inés Fernández-Ordóñez, por sus consejos, comentarios y por haberme facilitado el acceso al fondo documental pidalino conservado en la Fundación Ramón Menéndez Pidal (FRMP en adelante). Igualmente a Ángel Berenguer Amador, por su ayuda con el judeoespañol, así como a los asistentes de diversos coloquios en los que he presentado versiones previas de este trabajo, cuyos comentarios me han dado pistas de gran interés.

** Correo electrónico: carlota.debenitomoreno@uzh.ch. Código ORCID: <https://orcid.org/0000-0001-9112-5471>.

¹ El uso pidalino del término *sincretismo* no coincide totalmente con el actual, donde se refiere al uso de una única forma o significante para varias funciones o significados, sino que puede tener el significado opuesto: el uso de varias formas o significantes para una única función o significado.

agrupan papeletas bibliográficas sobre distintos fenómenos, mayoritariamente fonéticos, que se encuentran a ambos lados del Atlántico, como la fricativización del grupo /tr/, la simplificación de los grupos cultos (*dotor*, *perfeto*), la pérdida ocasional de /n, d, r/ (*ties* ‘tienes’, *vies* ‘vienes’, *pue* ‘puede’, *quies* ‘quieres’), las confusiones del tipo *concencia-inociencia*, la velarizaciones de labiales (*golver* ‘volver’ o *gomitar* ‘vomitar’),² los cruces del tipo *pelizcar* ‘pellizcar’ (con *piel*) o *témido* ‘tímido’ (con *temer*), los trueques acentuales en hiatos (*máiz*, *bául*), entre varios otros.

Lamentablemente, don Ramón no llegó a desarrollar totalmente esta cuestión en su obra publicada, y apenas podemos vislumbrar algunas de sus ideas en el tratamiento que dio a algunos (pocos) de estos fenómenos en su *Manual de gramática histórica española*.³ Por ello, para entender lo que aúna a esta serie de papeletas del cajón 18 son fundamentales tanto las citas que escoge de las distintas obras que referencia como sus anotaciones sobre ellas. Así, por ejemplo, en la papeleta 1194, sobre *témido-tímido*, cuyas menciones en la bibliografía traza don Ramón desde su «descubrimiento» por Espinosa (1930), encontramos la siguiente nota: «ejemplo de lo mucho dialectal común entre Esp. y América. Una forma que se descubre aislada y luego van apareciendo analogías».⁴ Parece, pues, que lo que le llama la atención es la recurrencia de esta forma en distintas ubicaciones. Todavía más iluminadora es la papeleta 1196, en la que se reseñan las observaciones de Federico de Onís (1961: 566) sobre las diferencias entre los poetas regionales y los universales y cultos. De acuerdo con estas, el lenguaje popular de los primeros es «casi uniforme en sus vulgarismos» —glosa Pidal—,⁵ mientras que en el de los últimos «lo regional es profundo y diferencial»:

Onís observa p. 566 q los poetas *regionales* en dialecto y populares son lo más uniforme y monótono de toda la lit^a moderna. *El lenguaje popular que usan es casi uniforme en sus vulgarismos* [verlos] y el carácter que expresan es el primario humano. En cambio lo regional es profundo y diferencial cuando se trata de los

² Al respecto de este fenómeno de equivalencia o confusión entre la labial y la velar, el *Atlas Lingüístico de la Península Ibérica* —iniciativa, como es bien sabido, de don Ramón— ofrece información interesantísima sobre la forma *abuja* ‘aguja’, estudiada con detenimiento por Navarro Tomás, que demuestra su extensión prácticamente absoluta en el territorio castellanohablante: «Rasgo característico de la historia peninsular de aguja es la sustitución de la g por la b en el habla popular de Castilla y de extensa parte de las provincias castellanizadas. Existe entre la popular *abuja* y la normal *aguja* una total separación social. Ninguna persona medianamente instruida pronuncia tal palabra con b. Se ha visto asimismo hasta qué punto fueron escasos los sujetos que la dijeron con g en el medio social representado por el ALPI» (Navarro Tomás, 1963: 299).

³ Allí, por ejemplo, se trata la equivalencia B=G como un «error lingüístico» debido a una equivalencia acústica (un error de percepción) (Menéndez Pidal, 1962 [1904]: § 72³) y la mayoría de los ejemplos ofrecidos se califican, efectivamente, de vulgares, por oposición de los encontrados en la lengua literaria. La misma oposición entre vulgar y literario se hace respecto del latín: «Pero si el latín vulgar explica la parte más grande y castiza de la lengua española, no puede explicarla toda. Gran porción de nuestro idioma, como de todos los romances, procede del latín literario» (*ibidem*: § 3.1). Este binomio es paralelo al que se establece entre popular y culto: «la voz popular tiene una significación más concreta y material, mientras la culta la tiene más general, elevada o metafórica» (*ibidem*).

⁴ Con «analogías» aquí Pidal parece referirse a la documentación de la misma forma repetidas veces y no a formas relacionadas morfológicamente.

⁵ Nótese que el término «vulgarismo» es de don Ramón, pues Onís (1934) emplea el mucho menos corriente «popularismo».

poetas universales y cultos Unamuno Marquina Machado Ayala [...] (XVIII-1196, Fondo documental de la *Historia de la lengua española*, FRMP).⁶

De acuerdo con esto, se adivina una diferencia entre una lengua vulgar común a todo el dominio hispánico y una serie de variedades regionales diferenciales que se sitúan en un nivel más culto.

Más aún, en la papeleta 1195, que trata la expresión *Una sé de agua*, detectada en la novelita gauchesca *El inglés de los güesos* de Benito Lynch, aparece la nota «la frase satirizada por Quevedo revive en la literatura criollista» y, más abajo, «vive en Español vulgar»: don Ramón vuelve a fijarse en una expresión que aparece a ambos lados del océano, esta vez en tiempos distintos, y da cuenta de esa discontinuidad en la documentación de la expresión por el ámbito en el que se utiliza: el del español vulgar.⁷

Creo que no resulta difícil entender el interés de don Ramón por estas semejanzas transatlánticas, dado que el carácter más o menos general de este español vulgar no encaja bien con otras de las ideas que aparecen en sus obras. De hecho, en su discurso de inauguración ante la Asamblea del Libro Español, «La unidad del idioma», Pidal expresa con claridad la idea contraria a la expresada en la papeleta 1196, indicando que es el habla popular la que difiere regionalmente, frente a la literaria, más uniforme:

Lo mismo debe decirse de las diferencias que separan las varias repúblicas de la América española entre sí, dado que el habla popular en cada una de ellas difiere respecto de las otras y es también divergente del habla literaria que les da cierta aparente unidad (Menéndez Pidal, 1944: s. p.).

La concepción pidalina de la organización de las diversas variedades de la lengua aparece en el capítulo «Varias clases de lengua» de su *Historia de la Lengua Española*:

Según el ámbito geográfico que cada uno de los tipos del idioma alcanza, se distingue una lengua común, la de ámbito más extenso, las más general, a diferencia de las lenguas dialectales, que viven dentro del territorio de la lengua común y emplean formas fonéticas, morfológicas y sintácticas diversas de las comunes, aunque afines a ellas. Hay también lenguas locales o provinciales, que simplemente usan algunos vocablos de uso local y acaso una pronunciación en algo distinta de la común (*idem*, 2007 [1939]: 17).

⁶ Transcribo literalmente las papeletas pidalinas sin intervenciones de puntuación, resolución de abreviaturas, etc. El paréntesis [*verlos*] es también de la mano de don Ramón.

⁷ Tampoco es el uso de «vulgar» totalmente inequívoco en las papeletas. En la papeleta 1206, sobre «diftongos vulgares en vez de hiatos», por ejemplo, los usos que se ilustran son propios de la gente educada en México, por lo que aquí debe entenderse como opuesto a *culto* en el sentido de ‘latinizante’ o ‘etimológico’ y no en el sentido de ‘educado’.

La «lengua común» se refiere, por tanto, a una norma suprarregional, propia de los hablantes educados:

La lengua común, formada mediante la exclusión o la parcial asimilación de los particularismos, se elabora en la conversación, en los negocios, en la literatura, por la selección de hablantes que viven las actividades superiores de la comunidad y participan de la más alta cultura; se fragua en las mayores ciudades, que son las menos apegadas a sus localismos propios, porque a ellas concurren gentes de todas partes y suelen los forasteros ser los dirigentes; se propaga en todas direcciones por los caminos del comercio vital que parten de esos mayores centros urbanos (*ibidem*: 18-19).

La existencia de esos vulgarismos extendidos en el espacio, por lo tanto, debe ser explicada históricamente como un «resto» de un cambio que sí triunfó entre las clases educadas. Es lo que encontrábamos en la papeleta 1195, así como en la 1204, que recoge las observaciones de Amado Alonso referentes a la diptongación de hiatos *aí* > *ái*, *eí* > *éí*.

La pronunciación *cáido*, *páis*, *máistro*, *créia* prosperó durante el siglo XIX por casi toda América y por la mitad norte de España (no en Andalucía); contra ella han reaccionado las clases cultas de todos los países reponiendo *caído* *creía* etc Bs As reacciono también pero con algun retraso. Hoy dicen *páis* las hijas de las madres que dicen *páis* Amado Alonso El Problema d la lengua en América 1935 p. 48 (XVIII-1204, Fondo de la *Historia de la lengua española*, FRMP).

En la *Gramática Histórica* encontramos otro ejemplo: el de la simplificación de los grupos cultos, que se originó en la «lengua popular» como consecuencia de la tendencia a simplificar grupos de consonantes de pronunciación difícil, pero que fue al principio «admitida en el habla literaria». Sin embargo, «en el siglo XVIII reaccionó el cultismo e impuso la pronunciación de todas las letras latinas, salvo en voces muy divulgadas como *delito*, DELICTUM, *luto* frente a *luctuoso*, *fruto* frente a *fructífero*, *respeto* junto a *respecto*, *sino* junto a *signo*» (Menéndez Pidal, 1962 [1904]: § 3.2). De hecho, la capacidad de las clases cultas de imponer su comportamiento lingüístico se explicita en otros momentos de la obra pidalina: «Es forzoso que una reacción correctiva empiece por las clases más educadas, pero de ellas se propaga a las clases que tienen menos tiempo para su educación» (Menéndez Pidal, 1944: s. p.).

La pidalina es la concepción clásica sobre la lengua vulgar, que puede presentar fenómenos comunes a las hablas vulgares de todo el dominio hispánico, cuyo origen parece hallarse en variantes generales que son desterradas del habla culta. Así lo sostiene García de Diego, que le dedica un apartado al español vulgar como «dialecto interno vertical»:

La lengua vulgar tiene gran difusión y una cierta coincidencia en las diversas regiones castellanas, no sólo de la Península, sino de los pueblos americanos. Por ser subestimado frente a la lengua culta, el castellano vulgar ha sido muy imperfectamente estudiado. En los pueblos americanos, tomándolo como hecho diferencial, ha sido recogido con más diligencia, y en estos estudios se aprecia que mucho de lo que se creía característico en su fonética, léxico y gramática, y como innovación nacional, no es sino supervivencia de hechos que viven en los fondos rurales o incultos de España (García de Diego, 1978: 369).

Lapesa también le dedica un apartado al «habla vulgar y rústica» en su *Historia de la lengua*, pero añade otra posible explicación a la existencia de estas formas: que se trate de desarrollos espontáneos:

Aparte de las modalidades más llanas del lenguaje correcto, existen usos cuyo radio de acción está hoy limitado a gentes iletradas de las aldeas y a las capas más populares de las ciudades. Muchos de estos vulgarismos se extienden con intensidad varia por todas o casi todas las regiones de lengua española. Algunos gozaron de mayor aceptación en épocas pasadas, e incluso penetraron en la literatura; *otros, que nunca lo lograron, son desarrollo de tendencias espontáneas del idioma refrenadas por la cultura en el uso normal* (Lapesa, 1981 [1942]: § 116, el resaltado es nuestro).

En este sentido, la concepción de la lengua vulgar y de sus hablantes es bastante uniforme en la bibliografía, que suele subrayar su conservadurismo (Fontanella de Weinberg, 1970: 17), fruto de su aislamiento (Olbertz, 2013: 26). Algunos autores, notando la ubicuidad de algunos rasgos vulgares, distinguen entre una variedad urbana y una rural (Donni de Miranda, 1992: 662). Otros notan que el habla rural no es necesariamente equiparable a la regional o local (Martín Zorraquino, 2010: 406-407). Un excelente acercamiento a las ideologías sobre la lengua rural y la urbana se encuentra en Britain (2017).

Nuestro objetivo en este trabajo es estudiar algunos fenómenos morfosintácticos que, por su distribución lingüística, pueden atribuirse —o se han atribuido— al español vulgar, con el objetivo de entender mejor tanto su distribución como su origen, con el objetivo de testar esta concepción sociohistórica. Debido a que solo contamos con corpus de habla rural en España, nuestro estudio no puede aspirar a documentar adecuadamente la vertiente americana de los fenómenos que se establecen a ambos lados del charco, como los que interesaban a don Ramón. Por ello, pretendemos únicamente ofrecer un modesto primer acercamiento a estos fenómenos.⁸

⁸ Si bien dicho acercamiento es cuantitativo, la estadística que se ofrece es meramente descriptiva (es decir, sobre los corpus), pues la escasez de las ocurrencias no permite otra cosa (con menos de 15 ocurrencias resulta aventurado ofrecer estadística inferencial, es decir, sobre la población, especialmente porque las observaciones no son siempre independientes, ya que pueden ser del mismo hablante o texto). Lamentablemente, se cumple plenamente la máxima de que la lingüística histórica es «the art of making the best use of bad data» (Labov, 1994: 11): nuestro objetivo ha sido el de reunir la mayor cantidad disponible de indicios para interpretar lo mejor posible la historia de los fenómenos.

En el siguiente apartado se reseñan algunos fundamentos teóricos de la concepción sociolingüística del español vulgar —algunos de los cuales ya se han avanzado en esta introducción—. Los apartados § 3-6 se dedican al estudio de cuatro fenómenos sintácticos y morfológicos cuya distribución geográfica y sociolingüística los hacen candidatos a ser vulgarismos: los adjuntos temporales etarios con *de* (§ 3), la concordancia singular con numerales complejos acabados en *un(a)* (§ 4), el orden de los pronombres en las combinaciones de clíticos (§ 5) y las formas analógicas de *haber* en *baiga* (§ 6). En § 7 se ofrecen algunas conclusiones.

2. FUNDAMENTOS TEÓRICOS DEL ESPAÑOL VULGAR COMO VARIEDAD

Existen dos conceptos teóricos fundamentales para apuntalar la existencia de cualquier variedad unitaria: el de comunidad de habla y el de difusión lingüística. Parodi y Santa Ana (1997) han aplicado el concepto de comunidad de habla al estudio del español rural, en uno de los poquísimos trabajos que teoriza sobre las dinámicas sociolingüísticas del español rural, combinándolo con el de red social (Milroy y Milroy, 1992) para proponer «un modelo teórico que abarque la interacción entre micro y macro estructuras en un área rural, partiendo de la difusión de la lengua estándar» (Parodi y Santa Ana, 1997: 307). Para dar cuenta de la (co)existencia de distintas comunidades de habla en las áreas rurales, proponen un modelo jerárquico e incluyente (figura 1), en el que, a medida que avanza la numeración, decrece, por un lado, la densidad de las redes sociales y la movilidad de los hablantes y, por otro, el conocimiento lingüístico de cada grupo. Así, la comunidad de habla IV conoce y puede evaluar los usos de su propia comunidad y de las otras tres; la comunidad de habla III conoce y puede evaluar los usos de su propia comunidad y de la II y la I, pero no de la IV, etc.

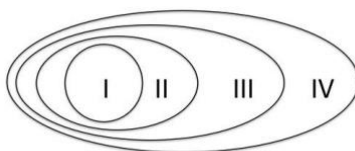


FIGURA 1. Comunidades de habla mutuamente incluyentes (adaptado de Parodi y Santa Ana, 1997)

El carácter jerárquico del modelo no se observa en las concepciones más tradicionales, como la pidalina, en la que la justificación para elegir una u otra modalidad lingüística depende de nociones como la afectividad o la eficiencia:

Cada hablante utiliza, además de la lengua común, otra u otras de esas variedades, y siempre la que tiene ámbito más reducido le da elementos de más personal afectividad; pero como esos elementos carecen de

eficiencia fuera de su reducido ámbito particular, el hablante tiene que renunciar a ellos cuando sale de la familia para tratar con sus convecinos, o cuando sale de la vecindad o del grupo para tratar con el resto de la comunidad: ninguna concepción ambiciosa y noble se resigna al particularismo de una expresión local (Menéndez Pidal, 2007 [1939]: 17-18).

La jerarquía de Parodi y Santa Ana, por el contrario, se basa en el prestigio de los distintos niveles de lengua y predice que las variantes empleadas por una comunidad de habla pueden extenderse a las comunidades incluidas en ella, pero que lo contrario no puede ocurrir. De hecho, los autores consideran que «el español estigmatizado está constituido por remanentes léxicos, fonológicos y morfológicos de la antigua koiné del español americano que se formó en el Nuevo Mundo durante el siglo XVI» (Parodi y Santa Ana, 1997: 314): otra vez los vulgarismos generalizados se conciben como restos o remanentes.

Este modelo, además, da cuenta del aislamiento que se ha atribuido al español vulgar o rural, pues los integrantes de la comunidad de habla I son los que presentan redes sociales más densas: sus relaciones sociales se limitan a «los miembros de su familia nuclear, su familia extendida y [a] un número limitado de personas de la localidad en que viven» (Parodi y Santa Ana, 1997: 317). Esta concepción encaja perfectamente dentro del modelo gravitacional de difusión de las innovaciones lingüísticas (Britain, 2004), de acuerdo con el cual las innovaciones no se difunden solo en función de la distancia, sino también del tamaño de la población: avanzando desde los mayores núcleos urbanos a otros de menor tamaño.

De esta concepción teórica se desprende que los rasgos del español vulgar, rural o estigmatizado son elementos que han desaparecido (o están haciéndolo) de las ciudades. Su distribución geográfica, ubicua y discontinua, es un efecto de esto: el hecho de que fueran variantes generales explica su ubicuidad y el hecho de que exista un proceso de cambio irradiado desde los núcleos urbanos, su discontinuidad.

Los fenómenos que nos interesan en este trabajo son precisamente los que muestran dicho patrón geográfico: una gran extensión en el espacio no restringida a áreas dialectales conocidas, pero ausentes de las ciudades y el habla culta. Nótese que un patrón geográfico semejante puede obtenerse también en los casos que Andersen denomina «de difusión sin adopción», en los que «internally motivated innovations [...] arise independently of any external stimulus» (1988: 154). Es decir, algunas innovaciones surgen en distintos sitios en distintos momentos: son poligenéticas. La sociolingüística anglosajona ofrece algunas contribuciones teóricas a este respecto en la discusión de los llamados *vernacular universals* (Chambers, 2008). Aunque la teoría original ha sido revisada y criticada (Trudgill, 2009), permanece una idea fundamental: algunos

tipos de cambio (como las regularizaciones o las pérdidas de redundancia) son particularmente probables y, por lo tanto, es comprensible que ocurran en varios lugares de forma independiente.

El resto de este trabajo se dedica al estudio sincrónico y diacrónico de cuatro fenómenos cuya distribución geográfica presenta estas características, con el objetivo de entender mejor sus circunstancias sociohistóricas. Puesto que consultamos numerosas fuentes, en el apéndice 1 se ofrece un resumen de las características de los corpus consultados.

3. ADJUNTOS TEMPORALES QUE EXPRESAN LA EDAD CON LA PREPOSICIÓN *DE*

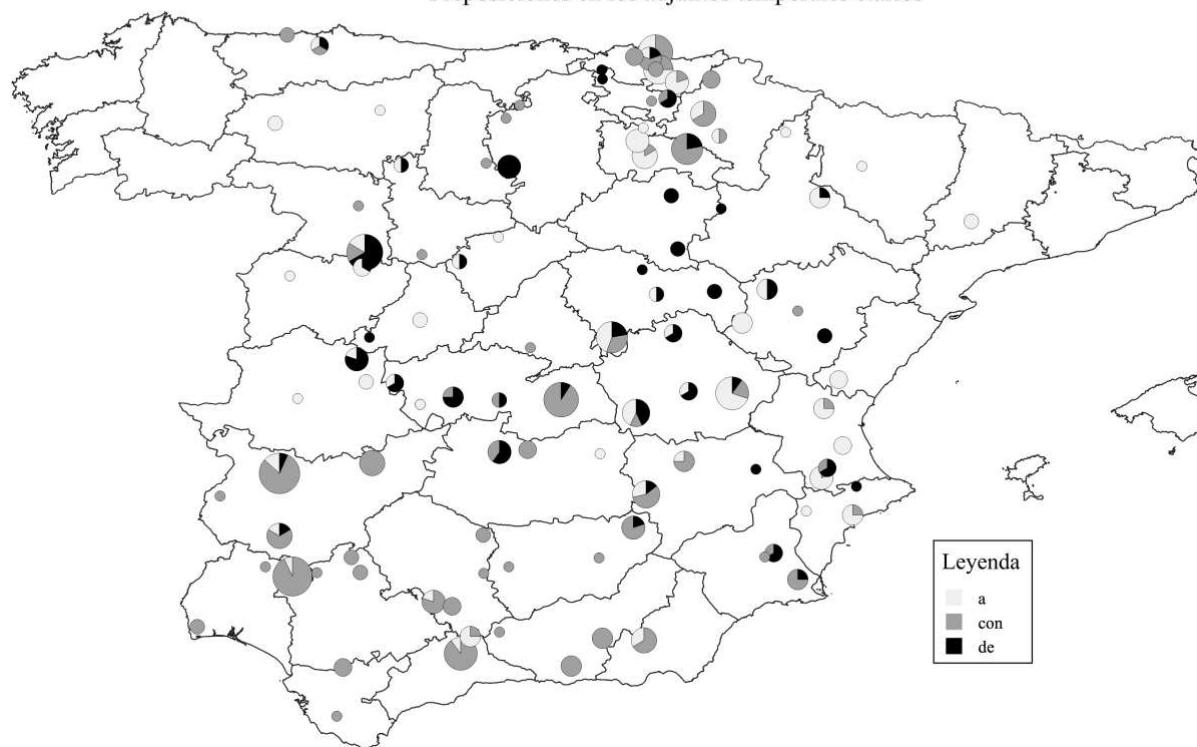
En este apartado abordamos un tipo de adjunto oracional que ha recibido poca atención en la bibliografía: los sintagmas preposicionales que expresan la edad de uno de los participantes o del hablante (adjuntos etarios en lo que sigue). Estos complementos circunstanciales de tiempo pueden estar introducidos por tres preposiciones distintas, a saber, *con*, *a* y *de* (1). Una diferencia sintáctica importante entre estas tres preposiciones es que, cuando la edad está expresada en años, meses o días (contexto al que nos restringimos aquí), solo la preposición *a* exige que estos vayan precedidos por el artículo definido:

- (1) a. [en la escuela] [D]esde que entrábamos *a los seis años* hasta los siete estábamos un par de años con una profesora que nos enseñaba a leer y a escribir.
b. Me casé *con veinte años*. Porque él tenía veintinueve.
c. Yo me casé *de veinte años* porque no tenía madre ni | no tenía madre (Povedilla, Badajoz, *COSE*R 0716).

La distribución sociolingüística y geográfica de estas tres variantes, sin embargo, no es uniforme. En el español peninsular rural (*COSE*R) las tres formas muestran vitalidad, siendo la forma mayoritaria la introducida por *con* (175/359, 48,7 %), seguida de la introducida por *a* (110/359, 30,6 %) y, en último lugar, de la introducida por *de* (74/359, 20,6 %)⁹. Estas variantes tampoco se distribuyen uniformemente en el espacio: como muestra el mapa 1, los adjuntos etarios introducidos por *de* apenas se documentan en las variedades más meridionales (andaluzas), donde es claramente mayoritario el uso de *con*.

⁹ Las búsquedas en el *COSE*R se han hecho empleando expresiones regulares, buscando {*a los/de/con*} seguido de una, dos o tres palabras cualesquiera y estas seguidas de {*año(s)/año(s)*}.

Preposiciones en los adjuntos temporales etarios



MAPA 1. Preposiciones que introducen adjuntos temporales etarios en el *COSEER*¹⁰

La situación cambia radicalmente fuera del habla rural. En el *PRESEEA* apenas encontramos casos de los adjuntos etarios introducidos por *de* en España: un único caso en Santiago de Compostela, que supone un 1 % del total de ejemplos documentados en esta ciudad (tabla 1).¹¹ Es decir, la variación en el habla urbana peninsular está reducida a dos únicas preposiciones (*a* y *con*)¹². En América, sin embargo, la situación es distinta y los adjuntos etarios con *de* se documentan con cierta frecuencia, parece que especialmente en Centroamérica (hay casos en todas las ciudades mexicanas encuestadas y en Ciudad de Guatemala) y en la región más norteña del continente sudamericano (Caracas, Lima y Medellín).

¹⁰ El tamaño de los diagramas de sectores del mapa 1 depende del total de ocurrencias documentadas en cada localidad: el máximo fueron 18.

¹¹ La tabla no muestra las ciudades que no documentaron ejemplos con *de*, a saber, Alcalá de Henares (n=18), Granada (n=21), La Habana (n=11), Madrid (n=16), Málaga (n=28), Montevideo (n=15), Santiago de Chile (n=15) y Valencia (n=37). El número tan bajo de ocurrencias se debe a que el *PRESEEA* solo pone a disposición de la comunidad en abierto una pequeña parte de sus materiales (compárese con Santiago de Compostela, enteramente disponible).

¹² Puesto que el *PRESEEA* no está lematizado ni etiquetado, ni permite el uso de expresiones regulares, buscamos la palabra *años* y analizamos los resultados para encontrar los ejemplos que nos interesan. El corpus de Santiago de Compostela está disponible completo en abierto (corpus *ESLORA*) y permite búsquedas más complejas: las realizadas fueron $\{de/con/a\ los\} \{*/**\} años$. Puesto que las búsquedas están limitadas a 5 palabras, para obtener los resultados de $a\ los\ ***\ años$, buscamos $a\ los\ ***$ y filtramos la columna siguiente para quedarnos solamente con las filas que comenzaban por *años*.

Ciudad	A	Con	de	Total
Medellín	2 (29 %)	0 (0 %)	5 (71 %)	7 (100 %)
Lima	6 (67 %)	0 (0 %)	3 (33 %)	9 (100 %)
Mexicali	1 (33 %)	1 (33 %)	1 (33 %)	3 (100 %)
Ciudad de Guatemala	12 (71 %)	0 (0 %)	5 (29 %)	17 (100 %)
México	17 (74 %)	0 (0 %)	6 (26 %)	23 (100 %)
Caracas	7 (88 %)	0 (0 %)	1 (12 %)	8 (100 %)
Guadalajara	6 (75 %)	1 (12 %)	1 (12 %)	8 (100 %)
Monterrey	6 (67 %)	2 (22 %)	1 (11 %)	9 (100 %)
Santiago de Compostela	47 (34 %)	90 (65 %)	1 (1 %)	138 (100 %)

TABLA 1. Frecuencia de las distintas preposiciones en adjuntos temporales etarios en el *PRESEEA*

El *CORPES XXI* da resultados similares para España:¹³ se documentan dos únicos adjuntos preposicionales etarios introducidos por *de*, que resultan no ser casos de usos españoles o contemporáneos: uno es una cita literal del músico cubano Bebo Valdés, mientras que el otro reproduce la inscripción que acompaña a un grabado de Mariano Brandi (1779-1824). Este corpus sí documenta algunos ejemplos de adjuntos etarios con *de* en América (tabla 2)¹⁴, aunque muy escasos y no limitados al área en que se documentan en *PRESEEA* —nótese que Chile presenta seis ejemplos, lo que nos hace pensar que una consulta a todo el corpus *PRESEEA* de Santiago de Chile podría arrojar ejemplos de interés—. ¹⁵

País	a	Con	de	Total
Nicaragua	7 (88 %)	0 (0 %)	1 (12 %)	8 (100 %)
República Dominicana	9 (82 %)	1 (9 %)	1 (9 %)	11 (100 %)
Chile	63 (86 %)	4 (5 %)	6 (8 %)	73 (100 %)
Ecuador	22 (88 %)	1 (4 %)	2 (8 %)	25 (100 %)
Colombia	49 (88 %)	3 (5 %)	4 (7 %)	56 (100 %)
Venezuela	22 (81 %)	4 (15 %)	1 (4 %)	27 (100 %)

¹³ Debido al volumen del corpus, hemos limitado las búsquedas a las secuencias (de lemas) {*de/con/a* + art.} + numeral cardinal + *año* precedidas directamente de un verbo.

¹⁴ En la tabla 3 se ofrecen solo los resultados de países que contienen ejemplos de *de*: aquellos que no mostraron ningún caso son Argentina, Bolivia, Costa Rica, Cuba, El Salvador, Estados Unidos, Guatemala, Guinea Ecuatorial, Honduras, Panamá, Paraguay, Perú, Puerto Rico y Uruguay.

¹⁵ La comparación entre *PRESEEA* y *CORPES XXI* deja asomar otra diferencia interesante, aunque fuera del ámbito que incumbe a este trabajo: mientras que en el *CORPES XXI* la preposición claramente preferida en todos los países (aunque algunos tienen pocos datos) es *a*, en *PRESEEA* *con* es mayoritaria en todas las ciudades españolas (y en La Habana, aunque allí hay pocos ejemplos en total), en consonancia con los resultados hallados en el *COSER*. En los textos del *CORPES XXI*, España es el país con mayor frecuencia de *con* (30 %), pero aún así esta está muy por debajo de *a* (70 %). Esto sugiere la existencia tanto de una diferencia geolectal como de una diferencia entre lengua oral y lengua escrita, merecedora de ser explorada más adelante.

México	102 (96 %)	1 (1 %)	3 (3 %)	106 (100 %)
España	205 (70 %)	87 (30 %)	2 (1 %)	294(100 %)

TABLA 2. Frecuencia de las distintas preposiciones en adjuntos temporales etarios en el
CORPES XXI

Es el análisis histórico de estas construcciones el que arroja luz sobre la distribución actual. Como puede verse en la tabla 3, los adjuntos etarios más antiguos son los introducidos por *de*, que son mayoritarios en los textos del *CDH* hasta el siglo XVII. Prácticamente hasta el siglo XV son la única opción para expresar este contenido, pues los casos con *a* son muy escasos (dos ejemplos en el siglo XIII) y con *con* no se documentan hasta el siglo XVII. La caída progresiva que vienen sufriendo las variantes introducidas por *de* es ya muy evidente en el siglo XVIII; en el siglo XIX el uso de esta preposición resulta anecdótica y ya no se documenta en los textos del XX.¹⁶

	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX	XXI
<i>A los</i>	0 (0 %)	2 (4,2 %)	0 (0 %)	4 (2,6 %)	8 (11,9 %)	9 (15 %)	53 (67,9 %)	116 (92,1 %)	490 (96,1 %)	73 (100 %)
<i>Con</i>	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)	1 (1,7 %)	5 (6,4 %)	5 (4 %)	9 (1,8 %)	0 (0 %)
<i>De</i>	14 (100 %)	46 (95,8 %)	20 (100 %)	151 (97,4 %)	59 (88,1 %)	50 (83,3 %)	20 (25,6 %)	5 (4 %)	11 (2,2 %)	0 (0 %)
Total	14 (100 %)	48 (100 %)	20 (100 %)	155 (100 %)	67 (100 %)	60 (100 %)	78 (100 %)	126 (100 %)	510 (100 %)	73 (100 %)

TABLA 3. Frecuencia de las distintas preposiciones en expresiones temporales etarias a lo largo del tiempo
(*CDH*)

El *CDH* además nos permite entender mejor la procedencia de esa preposición *de*, que es la misma que tenemos ahora en los sintagmas nominales del tipo *un niño de doce años*¹⁷. Como muestra la tabla 4, hasta el siglo XVII, el contexto mayoritario en el que aparecen los sintagmas preposicionales que indican la edad introducidos con *de* es el de atributo del verbo *ser* (es decir, no son adjuntos temporales), como se ve en (2).¹⁸

	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
Otros	0 (0 %)	6 (13 %)	5 (25 %)	20 (13,2 %)	23 (39 %)	20 (40 %)	16 (80 %)	4 (80 %)	8 (72,7 %)

¹⁶ Resulta interesante que la variante que emplea *con*, mayoritaria en el *COSER*, no llega nunca a superar el 10 % en este corpus, ni siquiera en el siglo XX, condictiendo con los resultados del *CORPES XXI* reseñados arriba.

¹⁷ Las búsquedas en el *CDH* se han realizado solo en *CDH* nuclear, para aprovechar su lematización más refinada. Las búsquedas se hicieron buscando numerales que tuvieran a su derecha el lema *año* (distancia 1) y a su izquierda los lemas *de/con* con la categoría «preposición» (a distancia 1) o el lema *a* con la categoría preposición (distancia 2) y la forma *los* (distancia 1).

¹⁸ Esta posibilidad combinatoria prácticamente se ha perdido hoy en día (en el conjunto de corpus sincrónicos examinados encontramos un único ejemplo en el *PRESEEA*, en Mexicali).

<i>Ser</i>	14 (100 %)	40 (87 %)	15 (75 %)	131 (86,8 %)	36 (61 %)	30 (60 %)	4 (20 %)	1 (20 %)	3 (27,3 %)
Total	14 (100 %)	46 (100 %)	20 (100 %)	151 (100 %)	59 (100 %)	50 (100 %)	20 (100 %)	5 (100 %)	11 (100 %)

TABLA 4. Contextos verbales en los que aparecen las expresiones temporales etarias introducidas por *de* a lo largo del tiempo (*CDH*)

(2) *De xxii annos era* Amon quando compeço a rregnar (*Liber Regnum*, c. 1194-1211, *CDH*).

El *CODEA+2015* presenta pocos ejemplos, pero interesantes, pues parecen mostrar un uso bastante conservador de la lengua en comparación con los datos del *CDH*.¹⁹ De los 58 ejemplos que nos ofrece este corpus, la mayoría son tardíos: tres de ellos (5 %) pertenecen al siglo XIII, pero todos los demás se reparten entre los siglos XVI (12 ejemplos, 21 %), XVII (22 ejemplos, 38 %) y XVIII (21 ejemplos, 36 %). Todos ellos muestran la preposición *de* —en combinación con *ser*—, a excepción de un único caso de finales del siglo XVIII, también en un contexto atributivo (con *ballarse*), que muestra la preposición *con*.

El corpus *Post Scriptum*, a pesar de su muy diferente composición textual, presenta una situación muy similar, aunque las ocurrencias en este caso son muy escasas: solo encontramos trece ejemplos de sintagmas preposicionales que introducen la edad, todos con la preposición *de* (dos en el siglo XVI, uno en el XVII, seis en el XVIII y cuatro en el XIX)²⁰. La mayoría de estos son atributos del verbo *ser* (once, es decir, el 85 %). Los dos ejemplos que no lo son, sin embargo, son claros casos de adjuntos temporales etarios, en los que el verbo *ser* no puede entenderse como omitido.

Los datos aquí reseñados sugieren que el uso de los sintagmas preposicionales etarios con *de* en el habla rural de la península se debe a la conservación de un uso antiguo que ha desaparecido del habla de las ciudades y, prácticamente, de la lengua escrita, es decir, del habla culta. Este caso encaja perfectamente en la concepción pidalina y tradicional de vulgarismo bosquejada arriba, pero nótese que el ritmo de desaparición no es el mismo en todos sitios, pues en parte de América estos sintagmas preposicionales con *de* no son vulgarismos, sino que siguen perteneciendo al habla urbana y la lengua escrita, como muestran los datos del *PRESEEA* y el *CORPES XXI*.

¹⁹ Las búsquedas en *CODEA+2015* consistieron en la búsqueda de las secuencias *{de/con/a los} {*/**/**}* *{años/annos}*.

²⁰ Los datos del *Post Scriptum* se han extraído por medio de expresiones regulares aplicadas a una versión tokenizada de los archivos TXT disponibles en la web del corpus, buscando las combinaciones de *{de/con/a los} + años* (con sus distintas versiones gráficas, de acuerdo con las formas que ofrecen los resultados de búsqueda en la versión estandarizada) con una, dos y tres tokens en medio (para admitir así numerales de hasta tres palabras).

4. CONCORDANCIA DE NÚMERO CON LOS NUMERALES ACABADOS EN *UN(A)*

Un fenómeno que sobresale en las entrevistas del *COSEER* es la singular concordancia (nunca mejor dicho) que desencadenan los numerales acabados en *un(o)/una* en el sustantivo y sus modificadores. Como puede verse en los ejemplos de (3), dichos numerales pueden desencadenar la concordancia singular, no solamente en el sustantivo adyacente (3a), sino también en los adjetivos que modifican a este (3b), incluso en contextos predicativos (3c).

- (3) a. Simplemente sal, y luego ya *a los veintiún día* le cuelgas en el desván y hasta que quieras (Arrabal de Portillo, Valladolid, *COSEER*-4401).
b. Y tengo *ochenta y un año cumplió* (Burgohondo, Ávila, *COSEER*-0604).
c. *Noventa y uno, cumplido* el 1 de julio (Sant Joanet, Valencia, *COSEER*, notas de encuesta).

Este fenómeno no parece haber sido descrito con anterioridad; de hecho, es tan desconocido que ni siquiera ha sido digno de sanción académica.²¹ Nos hallamos ante un caso de concordancia con el elemento más cercano de conjunto coordinado (Corbett, 2000: 196),²² como demuestra el ejemplo (3a), donde el artículo precedente al numeral está en plural (*los*):

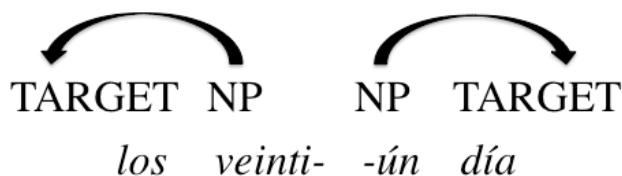


FIGURA 2. Concordancia con el elemento más cercano de un conjunto coordinado (adaptado de Corbett, 2000)

Las encuestas del *COSEER* arrojan frecuencias mayoritarias de la concordancia singular: de 86 ejemplos hallados en las entrevistas analizadas,²³ 57 (66,3 %) mostraban concordancia en singular y solo 29 (33,7 %) concordaban con el sustantivo en plural. Pero lo más interesante es que esto es así en otras lenguas iberorrománicas. En portugués, el *CORDIAL-SIN*, aunque contiene

²¹ La *NGLE* (§ 21.2u) sí sanciona la discordancia de género del tipo *veintiún páginas*.

²² Si bien esta concordancia no está descrita para el español, Corbett indica que «[i]n compound numerals the last element determines the form of the nouns» en ruso y en manés (Corbett, 2000: 212).

²³ Debido a que la aspiración y caída de *-s* final propia de las variedades meridionales hace difícil discernir si el sustantivo está en singular o plural, hemos descartado todas las entrevistas que mostraban dicha aspiración o caída en los contextos de los ejemplos encontrados. Es decir, no consideramos las entrevistas de Andalucía ni Badajoz, así como algunas de Cáceres (Talaván y Naval Moral de la Mata), Toledo (Tembleque) y Ciudad Real (Argamasilla de Alba y Malagón). Se buscaron los distintos numerales acabados en *un* y *una*, teniendo en cuenta también posibles variaciones fonéticas: *veintiún, ventiún, vintiún, {treinta / cuarenta / cincuenta / sesenta / setenta / ochenta / noventa / mil}* y *un, ciento un*.

menos ejemplos, ofrece resultados igualmente sorprendentes: de 14 ejemplos de numerales complejos acabados en *um*, 13 (93 %) muestran concordancia en singular con el sustantivo (4).²⁴

(4) Eu conheço uma senhora que esteve *vinte e um ano* a trabalhar por criada numa casa (Cedros, isla de Faial, Azores, *CORDIAL-SIN*).

Para el catalán, el *COD* solo ofrece un ejemplo de numeral complejo acabado en *un* (las entrevistas de este corpus también son muy breves), que también muestra concordancia en singular (5)²⁵. Nótese que, aunque este corpus no es exclusivamente rural, el ejemplo documentado procede de la pequeña localidad ilerdense de Sort. Lamentablemente, para el gallego, el *AGO* no recoge ningún ejemplo de numerales complejos acabados en *un/una*.²⁶

(5) Vai començar *as vint-i-un any* (Sort, Pallars, Sobirà, Lérida, *COD*).

Al cartografiar estos resultados (mapa 2) vemos que la concordancia «por contigüidad» no solo es mayoritaria en todo el territorio del que disponemos datos, sino que la concordancia estándar está recluida geográficamente a una franja norteña en el caso del castellano.²⁷ Asimismo, llama la atención que, al menos en los puntos del *COSER*, todos los hablantes mostraron sistemáticamente el mismo tipo de concordancia (aunque el número de casos total no sea abundante), sin variación.

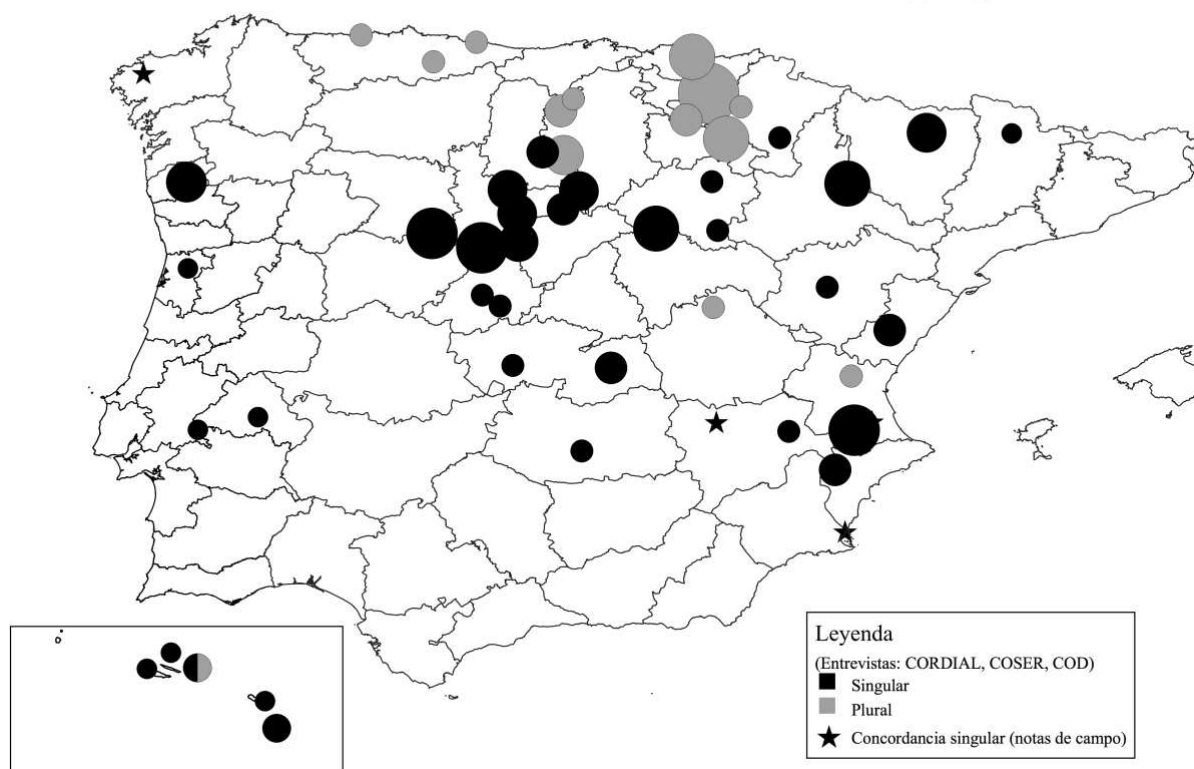
²⁴ Las búsquedas se hicieron con las grafías de todos los numerales acabados en *um(a)*.

²⁵ Las búsquedas se han hecho buscando las decenas en los archivos pdf del *COD*.

²⁶ En mis notas de la entrevista *COSER* en Zas (La Coruña), anoté un ejemplo de la concordancia en singular: la entrevista se realizó en español a un hablante cuya lengua materna era el gallego.

²⁷ El tamaño de los diagramas de sectores del mapa 2 depende del total de ocurrencias documentadas en cada localidad: el máximo fueron 7. Las estrellas indican documentaciones sueltas tomadas de mis notas de campo.

Concordancia con los numerales acabados en "un(a)/um(a)"



MAPA 2. Concordancia con los numerales complejos acabados en *un(a)/um(a)* en España y Portugal (COSER, COD, CORDIAL-SIN)

La comparación con otros corpus, históricos o sincrónicos, es muy interesante. En el *CORPES XXI* encontramos 581 casos de numerales complejos acabados en *un(a)* seguidos de un sustantivo.²⁸ 569 de ellos están en plural y 12 en singular, pero de esos 12 solo uno es un caso válido de sustantivo concordado en singular con el numeral, que es además una referencia al título de una obra sobre Cuba del siglo XIX. Es decir, el *CORPES XXI* no documenta ningún ejemplo de la concordancia por contigüidad que nos interesa en el siglo XXI. Aunque no podemos detenernos en la situación de otras lenguas vecinas, conviene mencionar que las frecuencias de la concordancia en singular también son muy bajas en portugués (4/1040 en el *CRPC*: un 0,4 %) y algo menos en gallego (5/85 en el *TILG*: 6 %) y catalán (23/281 en el *CTILC*: 2,1 %).

En las ciudades peninsulares del *PRESEEA*,²⁹ tanto Madrid como Alcalá de Henares muestran un 25 % de casos en singular, que en números absolutos corresponden a uno de cuatro

²⁸ Estas búsquedas incluyen todos los numerales cardinales acabados en *un(a)*, incluidas todas las decenas, centenas y millares. Las búsquedas se restringieron señalando que los numerales debían estar seguidos por un sustantivo.

²⁹ Las búsquedas en *PRESEEA* se han hecho de la siguiente manera: en la web del corpus hemos consultado los materiales de todas las ciudades americanas, así como de Madrid y Valencia. Los de esta última ciudad se han completado con los materiales disponibles (en pdf) en la web de *PRESEVAL*. El corpus de Alcalá de Henares se ha consultado a través de su versión publicada del corpus (también en pdf). Los datos de Santiago de Compostela se

y dos de ocho, respectivamente. Valencia documenta un ejemplo en singular de los seis que presenta (17 %), mientras que en Santiago de Compostela se atestiguan dos casos en singular de 27 (7 %). Las pocas ocurrencias nos impiden evaluar estas proporciones adecuadamente. La distribución de los ejemplos en singular por edad y grupo educativo indica que ninguno de los ejemplos en singular se dan ni en los hablantes más jóvenes ni en los de nivel educativo más elevado (universitario), aunque, otra vez, los pocos ejemplos de la mayoría de los subcorpus impiden llegar a conclusiones definitivas. Solo en uno de los hablantes encontramos variación entre ambas concordancias (un caso de singular por dos de plural), pero es justo decir que solo otros dos informantes de los que presentaron la concordancia por contigüidad mostraron más de un ejemplo —dos cada uno—.

En las entrevistas *PRESEEA* realizadas en ciudades americanas,³⁰ localizamos 22 ejemplos de estos numerales, solo uno de los cuales (de un hablante de los grupos etario y educativo intermedios de Montevideo) concuerda con un sustantivo en singular. Debemos notar, en cualquier caso, que en varias de las ciudades americanas entrevistadas (Montevideo incluida) encontramos fenómenos de aspiración y caída de la *-s* final que pueden dificultar la correcta transcripción de los datos.³¹

Por lo que respecta a la situación histórica, de 1047 resultados (válidos) seguidos de un sustantivo en el *CDH*, 1032 muestran el sustantivo en plural y 15 de ellos (1,4 %) lo presentan en singular.³² Seis de ellos proceden del scriptorium alfonsí —es decir, de la segunda mitad del siglo XIII— y son los únicos que se documentan en dicho siglo: por lo tanto, la concordancia por contigüidad es la única documentada en este periodo en el *CDH*. Los demás se reparten cronológicamente en un ejemplo del siglo XVI, otro del XVII, dos del XVIII, dos del XIX y tres del XX (uno de ellos de Guatemala). En la tabla 5 puede observarse que, salvo en el primer periodo

han consultado en la web del *ESLORA*, que contiene los materiales completos. Hemos excluido Granada y Málaga, por el mismo motivo por el que excluimos Andalucía de los datos del *COSER*. Las búsquedas se realizaron por las siguientes formas: *veintiún*, *ventiún*, *veintiuna*, *treinta y un*, *cuarenta y un*, *cincuenta y un*, *cincuenta y un*, *sesenta y un*, *setenta y un*, *ochenta y un*, *noventa y un*, *ciento un*, *mil un* (estas búsquedas incluyen los casos femeninos, puesto que el *PRESEEA* no hace las búsquedas sobre una versión tokenizada del corpus).

³⁰ Las ciudades americanas consultadas son las disponibles en el corpus online en junio de 2019: Caracas, Ciudad de Guatemala, Ciudad de México, Guadalajara, La Habana, Lima, Medellín, Mexicali, Monterrey, Montevideo y Santiago de Chile, pero no se han documentado casos de numerales complejos acabados en *un(a)* en Ciudad de Guatemala, Lima ni Medellín.

³¹ Más aún, este fenómeno de concordancia parece resultar poco saliente para muchos hablantes, incluso si tienen formación lingüística: los datos del *COSER* los revisamos uno a uno con el audio para descubrir que muchos de ellos estaban mal transcritos: el transcriptor escribía el plural a pesar de que el informante pronunciaba claramente el singular. Esta baja saliencia puede explicar que el fenómeno no se haya documentado hasta ahora.

³² Debido al volumen del corpus, hicimos búsquedas únicamente con el término *veintiún*, por ser el más frecuente en otros corpus. Las búsquedas se realizaron aprovechando la lematización del corpus: buscamos el lema *veintiuno* seguido inmediatamente de un sustantivo en todo el corpus.

—donde los ejemplos son muy exiguos—, la concordancia en singular es muy minoritaria en estos textos.³³

	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX	XXI
plural	0 (0 %)	4 (100 %)	14 (100 %)	36 (97 %)	52 (98 %)	20 (91 %)	144 (87 %)	742 (93 %)	20 (100 %)
singular	6 (100 %)	0 (0 %)	0 (0 %)	1 (3 %)	1 (2 %)	2 (9 %)	2 (13 %)	3 (7 %)	0 (0 %)
Total	6 (100 %)	4 (100 %)	14 (100 %)	37 (100 %)	53 (100 %)	22 (100 %)	146 (100 %)	745 (100 %)	20 (100 %)

TABLA 5. Concordancia del sustantivo con el numeral *veintiún* en el *CDH*

En el *CODEA+2015*, encontramos 148 ejemplos de numerales complejos acabados en *un(a)*. De estos, diecisiete (11,5 %) concuerdan en singular. Aunque el número de casos no es elevado, en la tabla 6 se observa que estos usos son mayoritarios en el corpus durante el siglo XIII, pero ya no en el XIV.

	XIII	XIV	XV	XVI	XVII	XVIII
plural	2 (15 %)	15 (75 %)	33 (100 %)	38 (100 %)	34 (100 %)	9 (90 %)
singular	11 (85 %)	5 (25 %)	0 (0 %)	0 (0 %)	0 (0 %)	1 (10 %)
Total	13 (100 %)	20 (100 %)	33 (100 %)	38 (100 %)	34 (100 %)	10 (100 %)

TABLA 6. Concordancia con los numerales complejos acabados en *un(a)* en el *CODEA+2015*

El corpus *Post Scriptum* ofrece muchos menos resultados, pero lo hace tanto para el español como para el portugués. En el subcorpus del español se documenta solo un ejemplo con la concordancia en singular, de un total de diecisiete ejemplos de numerales complejos acabados en *un(a)* (es decir, el 6 %). Dicho ejemplo se documenta en el siglo XV y supone el 20 % de los cinco casos encontrados. En el subcorpus del portugués solo se documentan seis ejemplos en total, dos de ellos en singular (33 %): uno en el siglo XVI (de dos ejemplos) y otro en el XVII (el único documentado). Las ocurrencias son tan escasas que no pueden sacarse conclusiones de su distribución por siglos.

Por último, resulta de gran interés investigar la situación del judeoespañol. De los nueve resultados hallados en el *CORHIJE*, repartidos equitativamente entre los siglos XVIII, XIX y XX, solo uno de ellos está en plural, en el periodo más moderno (tabla 7).

³³ No debe descartarse que existan algunos ejemplos de concordancia en singular que se han transcrito en las ediciones incluidas en el *CORDE* en plural por haberse considerado errores por parte de los editores.

	XVIII	XIX	XX
plural	0 (0 %)	0 (0 %)	1 (33 %)
singular	3 (100 %)	3 (100 %)	2 (67 %)
Total	3 (100 %)	3 (100 %)	3 (100 %)

TABLA 7. Concordancia con los numerales complejos acabados en *un(a)* en el *CORHIJE*

Globalmente, nuestros resultados indican que la concordancia por contigüidad es un fenómeno muy antiguo, como sugiere su documentación en todas las lenguas iberorromances (judeoespañol incluido). Este fenómeno, sin embargo, solo se documenta con frecuencia elevada en la lengua rural (tanto en español como en portugués) y es en general evitado en la lengua escrita.³⁴ El español peninsular urbano parece mostrar una situación intermedia, que quizá muestre una división entre las clases más populares y las de nivel educativo más alto, aunque los ejemplos con los que contamos no permiten asegurarlo. Llama la atención que la concordancia por contigüidad es la forma mayoritaria en los (escasos) ejemplos del siglo XIII que tenemos: un corpus más amplio podría permitirnos saber si en dicho siglo existía variación en la escritura, ya que todos nuestros ejemplos proceden de dos únicas obras alfonsíes. En cualquier caso, parece que la concordancia por contigüidad fue erradicada muy temprano y bastante tajantemente de la escritura, creando una división casi perfecta entre variedades orales «vulgares» y variedades cultas y escritas que se ha mantenido plurisecularmente.

5. EL ORDEN DE LOS PRONOMBRES EN LAS COMBINACIONES DE CLÍTICOS

Los verbos y construcciones reflexivas del español pueden tomar, además del pronombre reflexivo, otro pronombre dativo o acusativo, formando una secuencia de clíticos. En la lengua estándar, el primer pronombre que aparece en la secuencia es *se*, permitiéndose únicamente las combinaciones *se me*, *se te*, *se le(s)*, *se la(s)*, *se lo(s)*, *se nos* y *se os*. El orden inverso, sin embargo, se documenta ampliamente, especialmente cuando el pronombre no reflexivo es de primera o segunda persona singular (*me se*, *se te*):³⁵ «[l]a construcción [...] es común y general en el habla vulgar castellana», indica Martín Zorraquino (1979: 350). El orden de las combinaciones de clíticos ha recibido mucha atención desde el punto de vista teórico-formal (Heap, 1998; Heggie y

³⁴ Esto es así en todas las lenguas examinadas, pero las frecuencias ligeramente más elevadas en gallego y catalán son sugerentes: no perdamos de vista que la estandarización de estas lenguas es mucho más tardía.

³⁵ Martín Zorraquino (1979) observa que, en algunas zonas, la inversión del orden se extiende a las combinaciones con los pronombres plurales *os* y *nos*, dando ejemplos de la Ribera salmantina. En las entrevistas del *COJER* hemos documentado este orden tanto con los pronombres de primera y segunda persona plural (*nos se* y *os se*) como con el dativo singular (*le se*), pero con mucha menor frecuencia que los casos que afectan a *me* y *te*, a los que nos limitamos aquí.

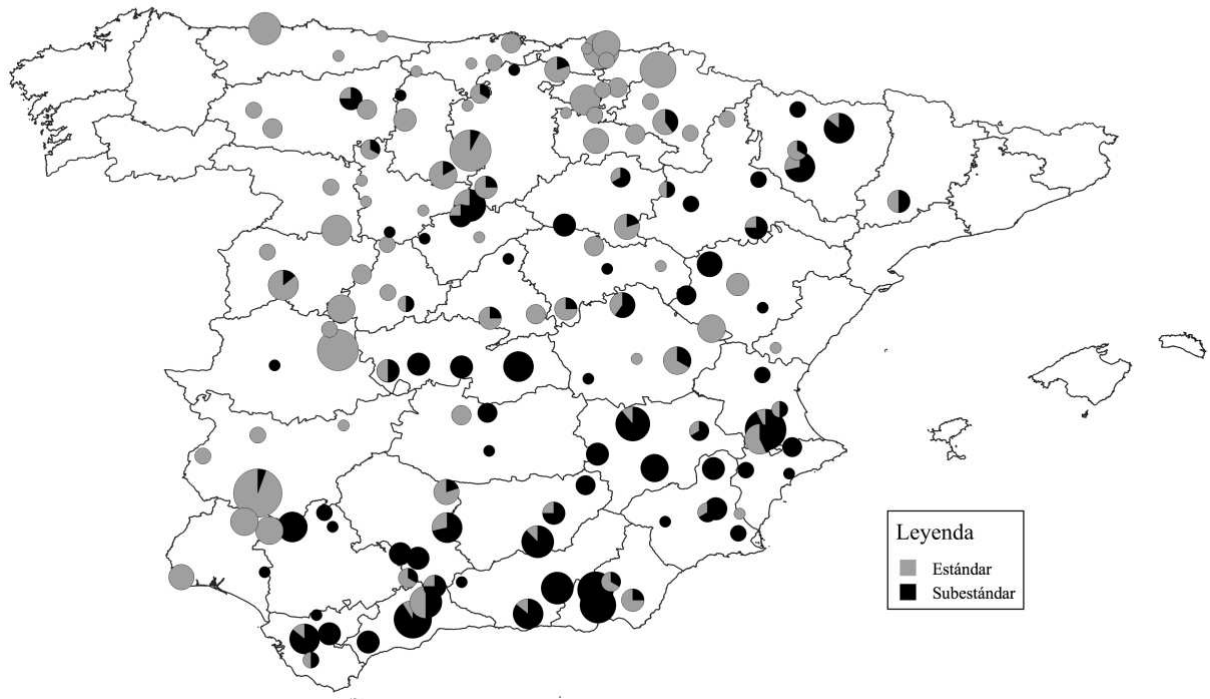
Ordóñez, 2005; Cuervo, 2013, entre otros), al contrario que la distribución geográfica de las secuencias «vulgares», solo estudiada por Heap (2003).

En dicho trabajo, Heap compara la distribución de las «secuencias invertidas» *me se* y *te se* en el *ALPI*, los atlas regionales y el *COSER*, por lo que conviene reseñar muy brevemente sus hallazgos. En primer lugar, Heap observa que el orden subestándar es mucho más frecuente que el estándar en los datos del *ALPI*: de 476 respuestas con combinaciones de clíticos (el *ALPI* contiene dos preguntas para documentar este fenómeno), 412 (86,6 %) muestran el orden subestándar. En segundo lugar, Heap identifica una jerarquía implicativa, de acuerdo con la cual el orden *me se* en la primera persona solo es posible si el orden *te se* se documenta en la segunda. En tercer lugar, de la distribución geográfica hallada en los atlas regionales Heap observa un patrón espacial organizado en torno al eje norte-sur (con mayor frecuencia de las formas subestándar al sur), que confirma con los datos (parciales y no exhaustivos) del *COSER*. De la comparación de estas tres fuentes dialectales, que pueden ordenarse cronológicamente de más antigua (el *ALPI*) a más reciente (el *COSER*), Heap llega a la conclusión de que las secuencias subestándar están retrocediendo. Puesto que encuentra ejemplos del orden subestándar ya en el siglo XII (en el *Corpus del Español*), concluye que «esta variación entre los órdenes que ahora son estándares y vernáculos no es un fenómeno reciente, sino más bien el reflejo de una dinámica que existe desde hace muchos siglos» (Heap, 2003: 795). Las secuencias subestándar, sin embargo, son siempre muy minoritarias, claro indicio de su alto nivel de estigmatización (Heap, 2003). Asimismo, Heap nota que la asimetría entre *te se* y *me se* se documenta en todas las épocas: a pesar del indudable interés de dicha jerarquía, por cuestiones de espacio no podemos detenernos en ella: los datos que se ofrecen a continuación se refieren al conjunto de ambas personas.

En su estado actual (julio de 2019), los datos del *COSER* indican que tanto el orden estándar como el subestándar compiten en el habla rural: de 561 ejemplos, 296 (52,8 %) ofrecen el orden estándar, por 265 (47,4 %) del subestándar³⁶. Dado que el *COSER* abarca ahora un área bastante mayor que aquella a la que pudo acceder Heap, creo conveniente mostrar los resultados en un mapa con gráficos de sectores. El mapa 3 sugiere que el eje norte-sur no es el único relevante, sino que también lo es el eje este-oeste, siendo las secuencias subestándar más comunes en la mitad oriental y en la mitad meridional del territorio.

³⁶ Para encontrar las secuencias proclíticas buscamos (siempre en la consulta básica) las secuencias *se me*, *se te*, *me se*, *te se* en la búsqueda exacta, mientras que para encontrar las enclíticas recurrimos a las búsquedas por expresiones regulares, usando las siguientes: «\b[A-Za-záéíóú]+seme\b», «\b[A-Za-záéíóú]+mese\b», «\b[A-Za-záéíóú]+tese\b», «\b[A-Za-záéíóú]+sete\b».

Orden de los pronombres en las secuencias de clíticos



MAPA 3. Orden de las combinaciones pronominales en el *COSEER* (refl + 1sg/2sg)³⁷

La comparación con el habla urbana resulta muy interesante. En el *PRESEEA* solo se documentan casos de las secuencias invertidas *me se* o *te se* en España (ninguna en América),³⁸ pero con una frecuencia extremadamente baja: solo se documentan catorce ejemplos de 524 secuencias de clíticos en primera o segunda persona del singular (2,7 %). En todas las ciudades muestra proporciones muy semejantes (tabla 8), aunque las ocurrencias son demasiado escasas para evaluar este dato: su mayor frecuencia en Valencia, la ciudad más oriental de las consideradas, solo puede considerarse sugerente.

	Santiago de Compostela	Alcalá de Henares	Madrid	Valencia	Granada	Málaga
Estándar	172 (97,2 %)	124 (98,4 %)	58 (98,3 %)	59 (92,2 %)	50 (98 %)	47 (100 %)

³⁷ El tamaño de los diagramas de sectores del mapa 3 depende del total de ocurrencias documentadas en cada localidad: el máximo fueron 18.

³⁸ Las ciudades americanas consultadas son las disponibles en el corpus online en junio de 2019: Caracas, Ciudad de Guatemala, Ciudad de México, Guadalajara, La Habana, Lima, Medellín, Mexicali, Monterrey, Montevideo y Santiago de Chile. Las ciudades españolas consultadas en el corpus fueron Madrid, Valencia, Málaga y Granada. Los datos de Valencia se han completado con los materiales disponibles (en pdf) en la web de *PRESEVAL*. El corpus de Alcalá de Henares se ha consultado a través de su versión publicada del corpus (también en pdf). Los datos de Santiago de Compostela se han consultado en la web del *ESLORA*, que contiene los materiales completos. Las búsquedas realizadas consisten en las cuatro secuencias (*se me*, *me se*, *se te*, *te se*).

Subestándar	5 (2,8 %)	2 (1,6 %)	1 (1,7 %)	5 (7,8 %)	1 (2 %)	0 (0 %)
Total	177 (100 %)	126 (100 %)	59 (100 %)	64 (100 %)	51 (100 %)	47 (100 %)

TABLA 8. El orden de las secuencias de clíticos en el *PRESEEA* (España) por ciudad

Esa baja frecuencia se encuentra también en los textos escritos, tanto modernos como históricos. El *CORPES XXI* solo ofrece 44 resultados de las secuencias subestándar de un total de 35 194 combinaciones de clíticos (reflexivo + 1sg/2sg), es decir, un 0,13 %.³⁹ Las secuencias se localizan en un número bastante reducido de países, como puede verse en la tabla 9.⁴⁰ La incidencia de las formas subestándar es prácticamente desdeñable.⁴¹

	Colombia	México	Perú	Venezuela	Honduras	España
Subestándar	1 (0,04 %)	1 (0,06 %)	1 (0,1 %)	1 (0,08 %)	5 (1,29 %)	32 (0,25 %)
Estándar	2783 (99,96 %)	4668 (99,94 %)	1023 (99,9 %)	1231 (99,92 %)	382 (98,71 %)	12 649 (99,75 %)
Total	2784 (100 %)	4671 (100 %)	1024 (100 %)	1232 (100 %)	387 (100 %)	12 681 (100 %)

TABLA 9. El orden de las secuencias de clíticos en el *CORPES XXI*

Los resultados del *CDH* son muy similares, con frecuencias extremadamente bajas durante todos los periodos:⁴²

	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX	XXI
Subestándar	0 (0 %)	2 (1,9 %)	0 (0 %)	2 (0,5 %)	1 (0,1 %)	2 (0,3 %)	19 (2,8 %)	21 (1,4 %)	26 (0,6 %)	0 (0 %)
Estándar	4 (100 %)	105 (98,1 %)	45 (100 %)	362 (99,5 %)	1026 (99,9 %)	797 (99,7 %)	660 (97,2 %)	1431 (98,6 %)	5194 (99,3 %)	271 (100 %)
Total	4 (100 %)	107 (100 %)	45 (100 %)	364 (100 %)	1027 (100 %)	799 (100 %)	679 (100 %)	1452 (100 %)	5220 (100 %)	271 (100 %)

TABLA 10. El orden de las secuencias de clíticos en el *CDH*

El judeoespañol vuelve a constituir una pista interesante: en los textos del *CORHIJE*, aunque hay pocas ocurrencias, tan solo se documentan las secuencias invertidas: nueve veces *me*

³⁹ Puesto que las ocurrencias de las secuencias estándar son muy abundantes, para su localización nos fiamos de la lematización del corpus, buscando *se* como pronombre personal de 3.ª persona seguido inmediatamente de *me* o *te* como pronombres personales de 1.ª o 2.ª persona, respectivamente. Para las ocurrencias subestándar, sin embargo, hicimos búsquedas por forma (*me* y *te* seguidos inmediatamente de *se*) y analizamos todos los resultados para comprobar su validez.

⁴⁰ Uno de los ejemplos subestándar aparece en un texto cuya procedencia geográfica no se identifica en el *CORPES XXI*.

⁴¹ Merece la pena notar que cinco de los 32 ejemplos españoles son de textos metalingüísticos, que *hablan* de estas secuencias, no las *usan* y que los cinco ejemplos hondureños de *me se* proceden todos del mismo texto.

⁴² Las búsquedas se realizaron (el 8 de septiembre de 2017) por lema (únicamente en la sección del *CDH* nuclear), buscando como primer lema uno de los pronombres (*me/te/se*) y como segundo lema, a distancia 1 a la derecha, el otro correspondiente. Lamentablemente esta búsqueda solo encuentra casos de pronombres separados por espacios, es decir, mayoritariamente proclíticos.

se y una vez *te se*, en consonancia con lo hallado para otros textos (Schmid y Bürki, 2000, Berenguer Amador, 2017). Por último, el corpus *Post Scriptum* nos da también información valiosísima (tabla 11). En las cartas privadas, aunque las ocurrencias totales son muchas menos que en los grandes corpus de referencia, las proporciones son mucho más elevadas. No cabe duda de que la incidencia de las secuencias subestándar es mucho mayor en este tipo de documentos. Puesto que el corpus *Post Scriptum* también tiene información geográfica sobre el origen de las misivas, podemos analizar la distribución geográfica de los ejemplos. La gran mayoría de los ejemplos de secuencias subestándar se encuentran en la Península Ibérica y no en América —pero téngase en cuenta que el continente americano presenta muchas menos ocurrencias en total—: allí solo uno de 35 (2,9 %) presenta el orden subestándar, frente a los setenta casos de los 462 (15,2 %) ubicados en Europa.

	XVI	XVII	XVIII	XIX
Subestándar	0 (0 %)	11 (8,5 %)	41 (14,3 %)	32 (27,1 %)
Estándar	85 (100 %)	119 (91,5 %)	246 (85,7 %)	86 (94 %)
Total	85 (100 %)	130 (100 %)	287 (100 %)	118 (100 %)

TABLA 11. El orden de las secuencias de clíticos en *Post Scriptum*

Martín Zorraquino da noticia de algunas menciones de las secuencias subestándar en América, pero no parece que estén tan generalizadas como en España: los autores consultados frecuentemente mencionan lo raras que son dichas secuencias o incluso notan su inexistencia. Más aún, nótese que el *ALEICan* es el único atlas peninsular que no ofrece ningún ejemplo del orden subestándar. Resulta, por lo tanto, muy conveniente consultar el *Corpus Diacrónico y Diatópico del Español de América (CORDIAM)*. Del total de 1980 ocurrencias de las secuencias pronominales, solo 5 (un 0,3 %) siguen el orden subestándar, con una distribución por siglo algo más baja que la mostrada por los datos del *CDH* (pero, otra vez, las bajas ocurrencias nos aconsejan cautela) (tabla 12):

	XVI	XVII	XVIII	XIX
Estándar	729 (99,5 %)	418 (100 %)	632 (99,8 %)	196 (100 %)
Subestándar	4 (0,5 %)	0 (0 %)	1 (0,2 %)	0 (0 %)
Total	733 (100 %)	418 (100 %)	633 (100 %)	196 (100 %)

TABLA 12. El orden de las secuencias de clíticos en el *CORDIAM* por siglo

El estudio pormenorizado de las secuencias de clíticos nos muestra una situación similar a la de la concordancia con los numerales complejos acabados en *un(a)*, en la que parece que tenemos

una variante puramente «vulgar» (restringida a la oralidad del ámbito rural y la inmediatez de los documentos privados).⁴³ A favor de la antigüedad del fenómeno nos hablan las documentaciones del siglo XIII y el hecho de que se halla en judeoespañol. Sin embargo, su naturaleza necesariamente tardía, la distribución espacial —oriental y meridional— que muestra en el español rural peninsular, así como el hecho de que no se encuentre en Canarias y sea tan infrecuente en América, sugiere que el fenómeno no estaba generalizado en el español medieval —¿quizá restringido a un área concreta?—, como debía estarlo la concordancia singular con los numerales complejos acabados en *un(a)*. Es decir, nos hallamos ante un fenómeno que se ha difundido en el espacio sin ayuda del prestigio de los centros urbanos (más aún, en contra de este) y cuya extensión ha quedado restringida a las capas sociales que se suelen considerar conservadoras y aisladas: los hablantes rurales.⁴⁴

6. EL SUBJUNTIVO ANALÓGICO DEL VERBO *HABER*

El último fenómeno al que queremos referirnos es el de las formas analógicas del verbo *haber* en el presente del subjuntivo: *baiga, baigas, baigamos, baigáis, baigan*. Las formas verbales que muestran el llamado incremento velar⁴⁵ han recibido mucho interés en la bibliografía (Menéndez Pidal, 1962 [1904]; Malkiel, 1974; Kania, 2011; O'Neill, 2015), debido a su oscura relación con las desinencias con yod derivativa (Menéndez Pidal, 1962 [1904]: § 113). De acuerdo con Pidal, el incremento surgió en las formas verbales cuya raíz acaba en *-n/-l* seguida de yod (*tengo, salgo* < TENEŌ, SALIŌ). Luego atrajo a los verbos en los que la yod derivativa había influido en la consonante de la raíz: CADŌ > *cayo* > *caigo*, AUDIŌ > *oyo* > *oigo*, que se generalizan en los Siglos de Oro. Otras formas analógicas parecen más tardías y nunca llegaron a penetrar en el habla culta: es el caso de *baiga* por *haya* y *vaiga* por *vaya*. Menéndez Pidal los considera vulgarismos: «Aunque no son hoy usados, deben recordarse también otros presentes en *-ay* y los en *-ny*, que nuestros clásicos hacían alguna vez †*haiga*, †*vaiga*, †*buiga*, †*destruigo*, †*restituigo*; el vulgo sigue aún usando los

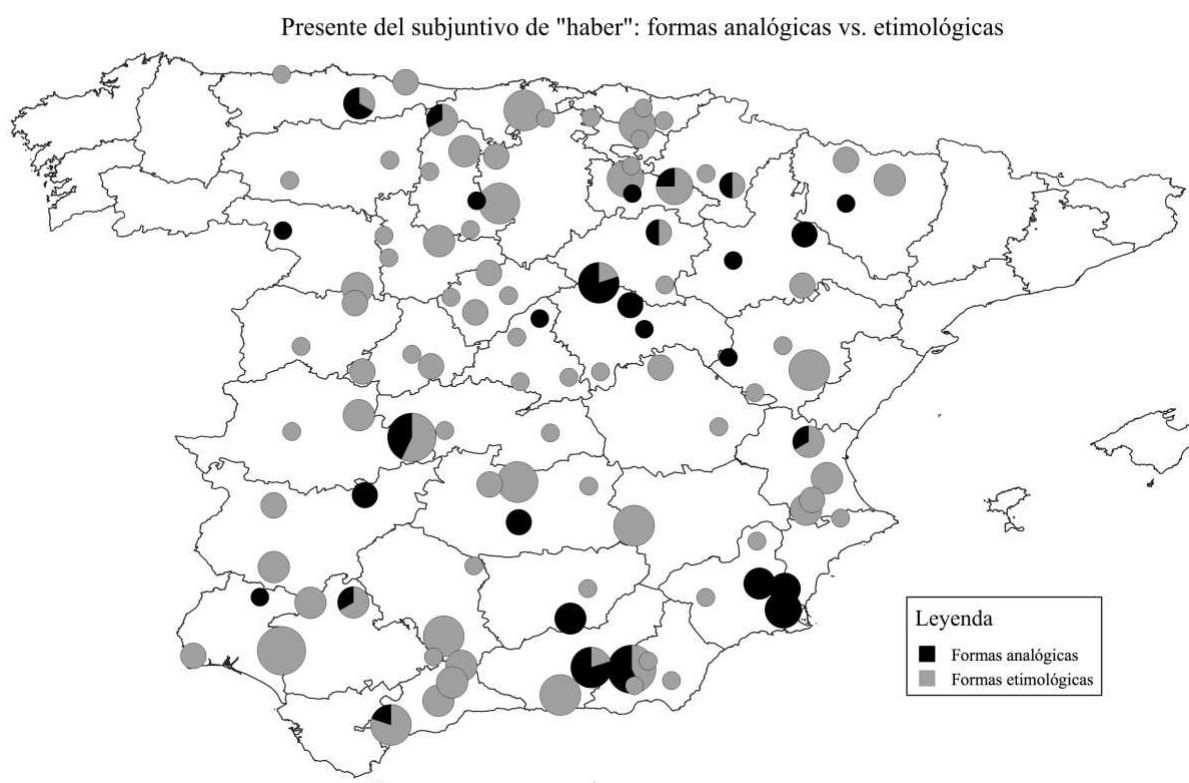
⁴³ No podemos realizar ahora un estudio paniberorrománico del fenómeno. Sin embargo, debemos notar que las secuencias de clíticos en sí mismas son un fenómeno necesariamente románico (y no latino), pues es en las lenguas romances en las que proliferan los usos de *se* y en las que los pronombres se convierten en clíticos. Es decir, no contamos con un modelo latino que sirva de base para las lenguas románicas: de hecho, las distintas lenguas han generalizado soluciones variables: las combinaciones *mi si* y *ti si* son estándar en italiano. Más aún, debemos notar que también en lenguas vecinas se encuentra la variación, por lo menos en gallego (Louredo, 2019).

⁴⁴ En mi opinión, no es evidente que el fenómeno esté en recesión, como cree Heap (2003). Las diferencias entre los datos del *ALPI* y los atlas regionales pueden deberse a los equipos encuestadores y la comparación con el *COSEER* no es adecuada, pues los datos son de naturaleza muy distinta. La desaparición de un vulgarismo generalizado por presión del estándar daría lugar, previsiblemente, a un patrón geográfico caótico y no a uno areal.

⁴⁵ En la primera persona singular del presente de indicativo —no en los verbos *haber* o *ir*, seguramente debido a la extrema irregularidad de su presente de indicativo, que impide que se establezca la analogía— y todo el presente de subjuntivo.

tres primeros y añade otros, por ejemplo, en Bogotá: †*creiga*, †*leiga*, †*reiga* = *ría*» (*ibidem*: § 113, 292-293)⁴⁶.

En el *COSER* se documentan 54 ejemplos de formas de *haiga* (en todas las personas), frente a 232 formas de *baya*, lo que supone un 23,3 % de todas las formas del subjuntivo de *haber*, una cifra relativamente elevada. La forma se documenta desperdigada por todo el territorio encuestado, generalmente en convivencia con el paradigma estándar (mapa 4):⁴⁷



En el habla urbana (*PRESEEA*), la forma analógica se documenta con una frecuencia llamativa, especialmente porque aparece en un elevado número de ciudades, españolas y americanas: en once de las diecisiete disponibles.⁴⁹ Como muestra la tabla 13, los porcentajes más

⁴⁶ Malkiel (1974) está a grandes rasgos de acuerdo con este desarrollo; Kania (2011) ha actualizado la cronología en un estudio cuantitativo reciente.

⁴⁷ Las búsquedas incluyen todas las formas del paradigma, al haberse realizado con las secuencias *haiga* y *baya* en búsqueda aproximada.

⁴⁸ El tamaño de los diagramas de sectores del mapa 4 depende del total de ocurrencias documentadas en cada localidad: el máximo fueron 7.

⁴⁹ Las búsquedas se han realizado en el corpus web (tanto en *PRESEEA* como *ESLORA*) e incluyen todas las formas del paradigma. No se documenta en Alcalá de Henares (n=44), Caracas (n=7), Ciudad de Guatemala (n=24), Madrid (n=42), Málaga (n=33) ni Valencia (n=28).

elevados se encuentran en Montevideo (¡con un 50 % de casos!), seguido de tres de las ciudades mexicanas.

Ciudad	<i>Haiga</i>	<i>Haya</i>	Total
Montevideo	15 (50 %)	15 (50 %)	30 (100 %)
Ciudad de México	11 (20 %)	44 (80 %)	55 (100 %)
Guadalajara	11 (18 %)	49 (82 %)	60 (100 %)
Mexicali	5 (15 %)	29 (85 %)	34 (100 %)
Granada	5 (12 %)	37 (88 %)	42 (100 %)
Medellín	4 (9 %)	42 (91 %)	46 (100 %)
Monterrey	5 (7 %)	69 (93 %)	74 (100 %)
Santiago de Chile	5 (4 %)	107 (96 %)	112 (100 %)
Lima	1 (2 %)	34 (97 %)	35 (100 %)
Santiago de Compostela	3 (3 %)	104 (97 %)	107 (100 %)
La Habana	1 (2 %)	65 (98 %)	66 (100 %)

TABLA 13. *haiga* vs. *haya* en el PRESEEA

La elevada frecuencia de *haiga* en español mexicano ha sido estudiada desde la perspectiva variacionista por Johnson y Barnes (2013), que demuestran que la forma es tanto más frecuente cuanto más bajo el nivel educativo del hablante. Los datos del PRESEEA —aunque escasos— sugieren que esto puede generalizarse para el mundo hispánico (véase la figura 3), pues los hablantes universitarios apenas usan las formas en *haiga* en la mayoría de las ciudades que documentan las formas. Sin embargo, hay que notar que la forma se emplea también por hablantes de nivel educativo elevado en Ciudad de México, en Monterrey y en Montevideo (con una frecuencia bastante sorprendente en esta última).

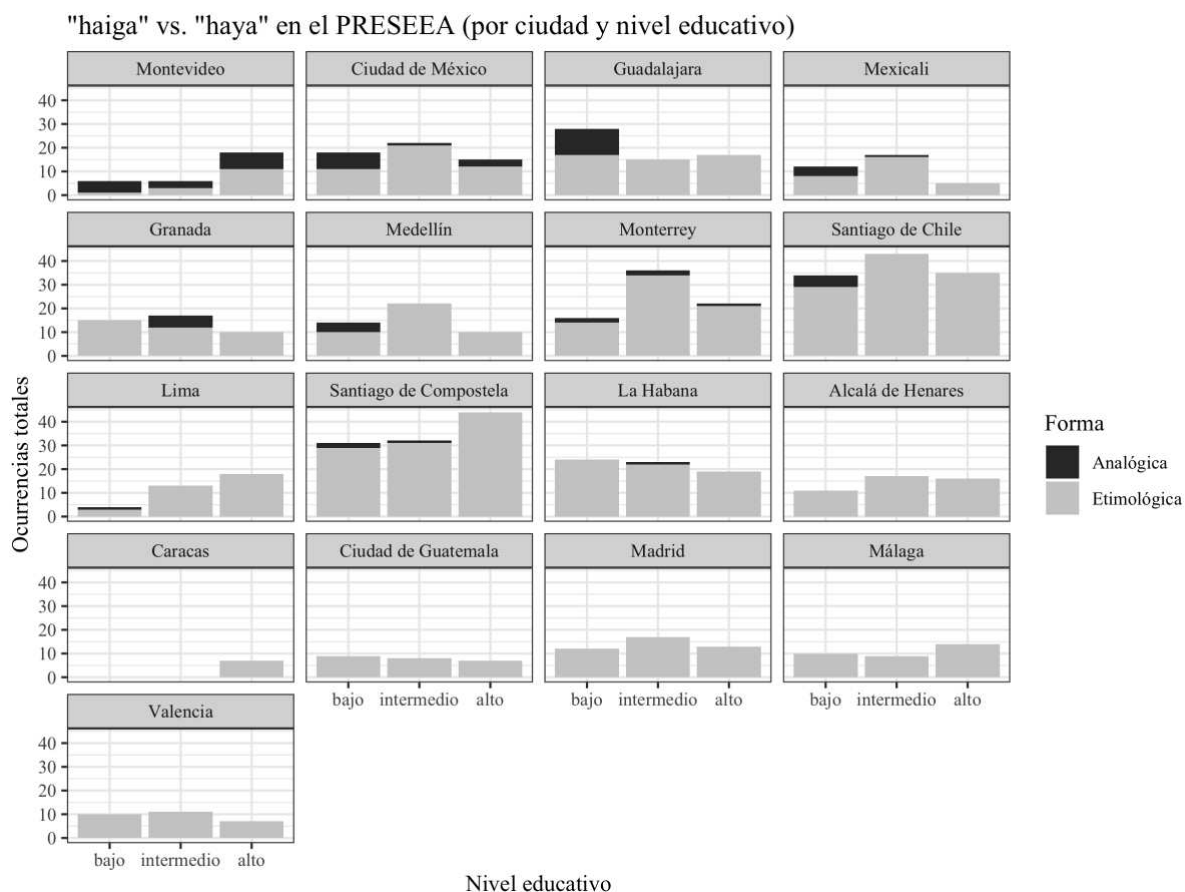


FIGURA 3. *haiga* vs. *haya* en el PRESEEA (por ciudad y nivel educativo)

El CORPES XXI da unas proporciones de la forma analógica mucho más bajas, que apenas superan el 1 % en algún caso (véase la tabla 14). Es más, tres de los ejemplos de *haiga* documentados (uno en España, uno en Colombia y el único de Puerto Rico) son, en realidad, menciones metalingüísticas de la forma.

País	<i>Haiga</i>	<i>Haya</i>	Total
Argentina	1 (0,06 %)	1549 (99,94 %)	1550 (100 %)
Bolivia	3 (1,02 %)	290 (98,98 %)	293 (100 %)
Chile	3 (0,33 %)	894 (99,67 %)	897 (100 %)
Colombia	7 (0,58 %)	1207 (99,42 %)	1214 (100 %)
Costa Rica	1 (0,51 %)	197 (99,49 %)	198 (100 %)
Cuba	1 (0,24 %)	417 (99,76 %)	418 (100 %)
Ecuador	1 (0,29 %)	339 (99,71 %)	340 (100 %)
España	2 (0,03 %)	7354 (99,97 %)	7356 (100 %)
Honduras	3 (1,21 %)	245 (98,79 %)	248 (100 %)
México	15 (0,79 %)	1876 (99,21 %)	1891 (100 %)
Perú	1 (0,24 %)	420 (99,76 %)	421 (100 %)
El Salvador	2 (0,91 %)	218 (99,09 %)	220 (100 %)

Uruguay	1 (0,21 %)	472 (99,79 %)	473 (100 %)
Venezuela	3 (0,46 %)	643 (99,54 %)	646 (100 %)

TABLA 14. *haiga* vs. *haya* en el CORPES XXF⁰

La información de los corpus históricos es crucial. En el CDH (tabla 15) las frecuencias se mantienen muy bajas (aunque algo menos que en el CORPES XXI) y las primeras documentaciones son del siglo XVIII, confirmando que *haiga* es una forma tardía (Malkiel, 1974; Kania, 2011). Pero tanto el CODEA+2015 (tabla 16) como *Post Scriptum* (tabla 17) adelantan esta datación. El CODEA+2015 documenta un caso en el siglo XVII, pero en 1700: solo diez años de los ejemplos más tempranos del CDH. *Post Scriptum* aporta ejemplos del siglo XVII más abundantes y tempranos, pues los primeros datan de 1658. Esto sugiere que el tipo de textos recogidos en este corpus (cartas privadas) favorece la aparición de las formas analógicas.

	XVIII	XIX	XX	XXI
<i>Haya</i>	1151 (96,3 %)	1250 (98,1 %)	2573 (98,2 %)	145 (100 %)
<i>Haiga</i>	44 (3,7 %)	24 (1,9 %)	47 (1,8 %)	0 (0 %)
Total	1195 (100 %)	1274 (100 %)	2620 (100 %)	145 (100 %)

TABLA 15. *haiga* vs. *haya* en el CDH⁵¹

	XII	XIII	XIV	XV	XVI	XVII	XVIII
<i>Haya</i>	2 (100 %)	332 (100 %)	280 (100 %)	294 (100 %)	267 (100 %)	49 (98 %)	31 (72 %)
<i>Haiga</i>	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)	1 (2 %)	12 (28 %)
Total	2 (100 %)	332 (100 %)	280 (100 %)	294 (100 %)	267 (100 %)	50 (100 %)	43 (100 %)

TABLA 16. *haiga* vs. *haya* en el CODEA+2015⁵²

	XVI	XVII	XVIII	XIX
<i>Haya</i>	100 (100 %)	180 (91,4 %)	186 (83 %)	88 (88,1 %)
<i>Haiga</i>	0 (0 %)	17 (8,6 %)	38 (17 %)	11 (11,1 %)
Total	100 (100 %)	197 (100 %)	224 (100 %)	99 (100 %)

⁵⁰ Para encontrar la forma estándar, buscamos el presente de subjuntivo de *haber*. Para la analógica, buscamos las formas que comienza por *(h)aiiga* o *(h)ayga* y analizamos los resultados manualmente.

⁵¹ Las búsquedas incluyen solo la forma sincrética entre la 1.^a y la 3.^a persona singular en el subcorpus CDH nuclear, que se analizaron manualmente. Las grafías buscadas fueron: *(h)aiiga*, *(h)ayga*, *(h)aya*, *(h)aya*. Por ser muy abundantes, no hemos examinado manualmente todos los ejemplos anteriores al siglo XVIII (es decir, antes de que se documente la forma analógica) y por ello no aparecen incluidos en la tabla. No obstante, indico aquí el número de resultados por siglo, para que esa ausencia de la forma analógica se ponga en contexto: siglo XII, 13 resultados (válidos); siglo XIII, 2029; siglo XIV, 448; siglo XV, 1623; siglo XVI, 1382; siglo XVII, 871. La gran mayoría de estos resultados son casos del subjuntivo de *haber*: el porcentaje de ocurrencias en los que *(h)aya* y *(h)aya* corresponden a otros lemas en los siglos XVIII-XXI va del 1,2 % al 4,5 %.

⁵² Las búsquedas incluyen todas las personas del paradigma. Las grafías buscadas fueron: *(h)aiiga*, *(h)ayga*, *(h)aya*, *(h)aiia*.

TABLA 17. *haiga* vs. *baya* en *Post Scriptum*⁵³

El paradigma analógico del subjuntivo de *haber* ilustra, en mi opinión, el caso de la poligénesis. Su distribución geográfica es, por un lado, similar a la de los otros fenómenos tratados anteriormente, pues aparece en todo el territorio peninsular en el habla rural. Sin embargo, su presencia en el *COSER* es bastante esporádica (geográficamente). Por otro lado, frente a la concordancia singular de los numerales complejos acabados en *un(a)* y el orden de los clíticos, *haiga* se documenta con más frecuencia en las entrevistas del *PRESEEA*, aunque es una forma claramente estigmatizada en la lengua escrita, como indica su baja documentación en el *CORPES XXI*. Más aun, no la localizamos hasta la segunda mitad del siglo XVII: una distribución geográfica tan amplia no casa bien con una difusión tan tardía de una forma estigmatizada. Creemos que también apoya la hipótesis de la poligénesis el hecho de que el subjuntivo de *haber* en judeoespañol se forma consistentemente con las formas analógicas (en *haiga*), según los datos del *CORHIJE*: puesto que no se documenta en español medieval, ha debido surgir allí de forma independiente.⁵⁴ El surgimiento múltiple de *haiga* no es sorprendente: se trata de un caso de regularización de la conjugación (hacia una forma también irregular). Este tipo de cambio se considera típico de los «universales vernáculos» (Trudgill, 2009). A lo largo de la historia del español, además, han surgido nuevas formas con el incremento velar —aunque no todas triunfaron en la lengua estándar—, pero no se ha observado que este surgimiento tenga un foco geográfico (Kania, 2011), como esperaríamos de un fenómeno que se hubiera difundido en el espacio. Por último, también habla a favor de la posibilidad de la poligénesis el hecho de que esta analogía se documente en el habla infantil, como atestigua el corpus *CHILDES*:

(6) que raro que haiga para todos (CHILDES, Pamplona, 3 años).

7. CONCLUSIONES

Si bien todos los fenómenos ilustrados en este trabajo pertenecen a la «lengua vulgar dialectal», en tanto en cuanto están ampliamente extendidos en el habla rural y se evitan en la lengua culta o literaria, la historia de algunos no encaja con la visión tradicional de los vulgarismos como producto del conservadurismo de las clases populares —especialmente las

⁵³ Las búsquedas incluyen todas las personas del paradigma y se realizaron aprovechando la búsqueda por forma estandarizada del corpus.

⁵⁴ Esta es la única forma documentada en otras obras en judeoespañol, véanse por ejemplo Schmid y Bürki (2000: 143) y Berenguer Amador (2017: 227-228).

rurales— (discutida en los apartados § 1 y § 2). Solo la historia de los adjuntos temporales etarios introducidos por *de* muestra un caso de preservación en el habla «vulgar» de una forma que gozó de prestigio en la lengua culta, pero fue desplazada por otra(s). Así lo demuestra su presencia mayoritaria en los corpus históricos hasta el siglo XVII (en el *CDH*, más literario) e incluso hasta el XIX (en *Post Scriptum* y el *CODEA+2015*). Quizá por lo tardío del desplazamiento, estos adjuntos no merecen el calificativo de vulgarismo en todo el territorio hispanohablante, pues, aunque en proporciones bajas, se localizan en textos escritos de diversos países americanos (*CORPES XXI*) y se conservan con vitalidad en el habla de varias ciudades americanas (*PRESEEA*).

El caso del subjuntivo analógico de *haber* (*baiga*) resulta representativo de otro de los posibles orígenes del vulgarismo según la concepción tradicional: la poligénesis. Así lo sugiere el hecho de que es una forma tardía e innovadora que nunca alcanzó prestigio en la lengua escrita: no se trata de un arcaísmo. Su amplia extensión geográfica, tanto en América como en España, se justifica por ser un caso de regularización paradigmática y, por tanto, un cambio «probable». Su distribución en el *COSER*, en mi opinión, es muy indicativa de su carácter poligenético, puesto que no muestra ni un uso sistemático ni un patrón geográfico que señale a un foco de difusión concreto. Al ser una forma fuertemente estigmatizada, está prácticamente ausente del habla de las clases educadas urbanas. Sin embargo, en algunas ciudades (como Montevideo) empieza a introducirse también en este sociolecto, mostrando que las dinámicas sociolingüísticas no son necesariamente idénticas en todo el espacio hispanohablante: es el famoso pluricentrismo.

Los dos casos más interesantes, en mi opinión, son los de la concordancia singular con los numerales complejos acabados en *un(a)* y el orden subestándar de las combinaciones de clíticos, puesto que no encajan en la concepción tradicional del «vulgo» como un grupo sociolingüísticamente pasivo: es decir, receptor de innovaciones pero no innovador.⁵⁵ La concordancia por contigüidad con los numerales nos muestra una situación de convivencia de variantes tan antigua que los datos históricos de que disponemos apenas nos permiten reconstruirla: es su sistematicidad en un amplio territorio contiguo iberorrománico (y no meramente castellano), así como las pistas que dan los escasos datos del siglo XIII lo que nos indica dicha antigüedad, pero para poder comprender el origen de su dinámica de variación — que separa con claridad el habla vulgar del habla culta y la lengua escrita, al menos en portugués y

⁵⁵ Creo que el papel que se da al latín vulgar como origen de gran parte de los cambios del latín al romance — observación que he de agradecer a un evaluador anónimo— no es contraejemplo, por dos motivos: el primero, por la gran multiplicidad de sentidos que se han atribuido a este concepto (a veces equivale meramente a latín hablado; Lloyd, 1989), que lo hace difícilmente comparable al uso que *vulgar* tiene cuando acompaña a otras lenguas y, en segundo lugar, porque el papel de este se destaca sobre todo en un escenario de conquista, que no corresponde con el escenario predominantemente peninsular tratado aquí.

castellano— habrá que retrotraerse al latín hispánico (lo dejamos para un trabajo futuro). Por lo tanto, no podemos estar seguros de si se trata de un arcaísmo o de una innovación. El orden no estándar de las secuencias de clíticos, sin embargo, es un fenómeno —necesariamente— más reciente, pero que muestra una distribución similar: una elevada frecuencia en el habla rural frente a una bajísima frecuencia en el habla urbana y la lengua escrita a lo largo de los siglos. En este caso, su frecuencia en el habla rural deja asomar un patrón geográfico continuo que no parece indicativo de la desaparición del orden subestándar como resultado de la extensión del orden estándar a partir de las ciudades. Al contrario, el orden subestándar parece haberse difundido a partir de un foco ¿oriental?, a pesar de ser una forma fuertemente estigmatizada y no contar con el respaldo de las clases educadas urbanas. Tanto la supervivencia casi general de la concordancia por contigüidad en el habla rural por tantos siglos como la difusión del orden subestándar de los clíticos chocan con el modelo de Parodi y Santa Ana (1997), que considera que los hablantes típicamente rurales conforman unas redes sociales muy densas y limitadas que los convierten en una suerte de guardianes del arcaísmo y de receptores pasivos de las innovaciones urbanas.

Lo cierto es, sin embargo, que en el mundo rural también se establecen relaciones sociales laxas y de larga distancia (con otras localidades rurales y con las clases populares urbanas) que permiten tanto la difusión de innovaciones lingüísticas como la conservación plurisecular de formas desprestigiadas: nótese que el territorio que conserva la concordancia por contigüidad es continuo y no parcheado, indicio de que la resistencia de la forma obedece a dinámicas sociolingüísticas activas de la comunidad de habla que emplea esta variante y de que no es un mero remanente condenado a desaparecer por la presión del estándar. Urge el estudio en profundidad tanto de estas dinámicas sociolingüísticas en el mundo rural como de los llamados vulgarismos (fonéticos, morfológicos y sintácticos), aspectos todos que se han visto descuidados hasta ahora por la concepción dominante del mundo rural como un lugar conservador e inmóvil (Britain, 2017).

BIBLIOGRAFÍA

- AGO = Fernández Rei, Francisco (dir.) (2010-): *Arquivo do Galego Oral*, Santiago de Compostela, Instituto da Lingua Galega, disponible en: <http://ilg.usc.es/ago/> [Consulta: 10/06/2019].
- ALEICan = Alvar, Manuel (1975): *Atlas Lingüístico y Etnográfico de las Islas Canarias*, Las Palmas de Gran Canaria, Cabildo Insular.
- ALPI = *Atlas Lingüístico de la Península Ibérica* (1962): Madrid, Consejo Superior de Investigaciones Científicas.

- ANDERSEN, Henning (1988): «Centre and periphery: adoption, diffusion and spread», en Jacek Fisiak (ed.), *Historical dialectology. Regional and Social*, Berlin/New York/Amsterdam, Mouton de Gruyter, pp. 39-83.
- BERENQUER AMADOR, Ángel (2017): *El libro sefardí La Güerta de oro de David M. Atías (Liorna, 1778). Edición y estudio lingüístico del verbo*, Lausana, Sociedad Suiza de Estudios Hispánicos.
- BRITAIN, David (2004): «Geolinguistics and linguistic diffusion», en Ulrich Ammon, Norbert Dittmar, Klaus Mattheier y Peter Trudgill (eds.), *Sociolinguistics: International Handbook of the Science of Language and Society*, Berlin, Mouton de Gruyter, pp. 34-48.
- (2017): «Which way to look?: Perspectives on “Urban” and “Rural” in dialectology», en Emma Moore y Chris Montgomery (eds.), *A Sense of Place: Studies in Language and Region*, Cambridge, Cambridge University Press, pp. 171-188.
- CDH = Instituto de Investigación Rafael Lapesa de la Real Academia Española (2013): *Corpus del Nuevo diccionario histórico (CDH)* [en línea], disponible en: <http://web.frl.es/CNDHE> [Consulta: Edad: 08/09/2017; Numerales: 29/08/2016; Pronombres: 06/09/2017; Subjuntivo: 07/09/2017].
- CHAMBERS, J. K. (2008): «Dynamic typology and vernacular universals», en Bernd Kortmann (ed.), *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*, Berlin, Mouton de Gruyter, pp. 127-146.
- CHILDES = MacWhinney, Brian y Catherine Snow (2003-): *Child Language Data Exchange System*, disponible en: <https://childes.talkbank.org/> [Descargado: 07/09/2016].
- COD = Viaplana, Joaquim, Maria-Rosa Lloret, Maria-Pilar Perea y Esteve Clua (2007): *Corpus Oral Dialectal*, disponible en: <http://www.ub.edu/ccub/corpusoraldialectal-cod.html> [Descargado: 01/09/2013].
- CODEA+2015 = Grupo de Investigación Textos para la Historia del Español (GITHE) (2015-): *Corpus de documentos españoles anteriores a 1800 (CODEA+2015)*, disponible en: <http://corpuscodea.es/> [Consulta: Edad: 01/08/2017; Numerales: 31/07/2017; Subjuntivo: 01/08/2017].
- CORBETT, Greville G. (2000): *Number*, Cambridge, Cambridge University Press.
- CORDIAL-SIN = Martins, A. M. (coord.) (2000-). *CORDIAL-SIN: Corpus Dialectal para o Estudo da Sintaxe / Syntax-oriented Corpus of Portuguese Dialects*, Lisboa, Centro de Linguística da Universidade de Lisboa, disponible en: <http://www.clul.ulisboa.pt/en/10-research/314-cordial-sin-corpus> [Consulta: 13/07/2019].
- CORDIAM = Academia Mexicana de la Lengua: *Corpus Diacrónico y Diatópico del Español de América*, disponible en: www.cordiam.org [Consulta: 22/06/2019].
- CORHIJE = García Moreno, Aitor y Francisco Javier Pueyo Mena: *Corpus Histórico Judeoespañol*, disponible en: <http://recursos.esefardic.es/corhije/> [Consulta: 31/07/2017].
- CORPES XXI = Real Academia Española: Banco de datos (CORPES XXI) [en línea]. *Corpus del Español del Siglo XXI (CORPES)*, disponible en: <http://www.rae.es> [Consulta: Edad: 16/06/2019; Numerales: 02/08/2016; Pronombres: 22/06/2017; Subjuntivo: 25/06/2019].
- COSEER = Fernández-Ordóñez, Inés (dir.) (2005-): *Corpus Oral y Sonoro del Español Rural*, disponible en: www.corpusrural.es [Consulta: Edad: 13/07/2019; Numerales: 14/07/2019; Pronombres: 13/07/2019; Subjuntivo: 13/07/2019].
- CRPC = Centro de Linguística da Universidade de Lisboa: *Corpus de Referência do Português Contemporâneo (CRPC)*, disponible en: <https://www.clul.ulisboa.pt/en/10-research/713-crpc-reference-corpus-of-contemporary-portuguese> [Consulta: 08/09/2017].
- CTILC = Institut d'Estudis Catalans (2014): *Corpus Textual Informatitzat de la Llengua Catalana (CTILC)*, disponible en: <https://ctlc.iec.cat/> [Consulta: 24/06/2019].

- CUERVO, María Cristina (2013): «Spanish clitic clusters», *Borealis. An International Journal of Hispanic Linguistics*, 2, 2, pp. 191-220, disponible en: <https://doi.org/10.7557/1.2.2.2692>.
- DONNI DE MIRANDE, Nélica E. (1992): «El sistema verbal en el español de la Argentina: rasgos de unidad y de diferenciación dialectal», *Revista de Filología Española*, LXXII, 3/4, pp. 655-670, disponible en: <http://doi.org/doi:10.3989/rfe>.
- ESLORA = ESLORA: *Corpus para el estudio del español oral*, versión 1.2.2 de noviembre de 2018, disponible en: <http://eslora.usc.es> [Consulta: Edad: 14/06/2019; Numerales: 11/06/2019; Pronombres: 14/06/2019; Subjuntivo: 16/06/2019].
- ESPINOSA, Aurelio (1930): *Estudios sobre el español de Nuevo México*, 2 vols., edición de Amado Alonso y Ángel Rosenblat, Buenos Aires, Universidad de Buenos Aires.
- FONTANELLA DE WEINBERG, M.^a Beatriz (1970): «La evolución de los pronombres de tratamiento en el español bonaerense», *Thesaurus*, XXV, 1, pp. 12-22.
- GARCÍA DE DIEGO, Vicente (1978): *Manual de dialectología española*, Madrid, Centro Iberoamericano de Cooperación.
- HEAP, David (1998): «Optimizing Iberian clitic sequences», en José Lema y Esther Treviño (eds.), *Theoretical Analyses on Romance Languages*, Amsterdam/Philadelphia, John Benjamins, pp. 227-248.
- (2003): «Sentencias “invertidas” de clíticos: un cambio (¿?) en tiempo real», en José Luis Girón Alconchel y José Jesús de Bustos Tovar (coords.), *Actas del VI Congreso Internacional de Historia de la Lengua Española*, Madrid, Arco Libros, pp. 785-98.
- HEGGIE, Lorie y Francisco ORDÓÑEZ. (2005): *Clitic and Affix combinations*, Amsterdam/Philadelphia, John Benjamins.
- JOHNSON, Mary y Sonia BARNES (2013): «Haya vs. Haiga: an Analysis of the Variation Observed in Mexican Spanish Using a Mixed Effects Model», en Ana M. Carvalho y Sara Beaudrie, *Selected Proceedings of the 6th Workshop on Spanish Sociolinguistics*, Somerville, Cascadia Proceedings Project, pp. 32-40.
- KANIA, Sonia (2011): «The Spread of the Velar Insert /g/ in Medieval Spanish Verbs», *Bulletin of Hispanic Studies*, 88, 2, pp. 129-160, disponible en: <https://doi.org/10.3828/bhs.2011.1>.
- LABOV, William (1994): *Principles of Linguistic Change*, vol I. *Internal Factors*, Oxford, Basil Blackwell.
- LAPESA, Rafael (1981 [1942]): *Historia de la Lengua Española*, novena edición, Madrid, Gredos.
- LLOYD, Paul M. (1989): «On the definition of Vulgar Latin», *Neuphilologische Mitteilungen*, 80, pp. 110-122.
- LOUREDO, Eduardo (2019): «Variación e cambio lingüístico en tempo aparente: o galego do Ribeiro», tesis doctoral, Universidad de Santiago de Compostela.
- MALKIEL, Yakov (1974): «New Problems in Romance Interfixation (I): The Velar Insert in the Present Tense (with an Excursus on -zer / -zir Verbs)», *Romance Philology*, 27, 3, pp. 304-355.
- MARTÍN ZORRAQUINO, María Antonia (1979): *Las construcciones pronominales en español. Paradigma y Desviaciones*, Madrid, Gredos.
- (2010): «El estudio del español regional de Aragón en el quehacer filológico de José María Enguita Utrilla», en Rosa M.^a Castañer Martín y Vicente Lagüéns García, *De moneda nunca usada. Estudios dedicados a José M.^a Enguita Utrilla*, Zaragoza, Institución Fernando el Católico, pp. 401-420.
- MENÉNDEZ PIDAL, Ramón (1944): *La unidad del idioma*, Madrid, Instituto Nacional del Libro Español.
- (1962 [1904]): *Manual de gramática histórica española*, Madrid, Espasa-Calpe.

- (2007 [1939]): «Del lenguaje en general. (Ensayo de una presentación de la Historia de la Lengua)», en Ramón Menéndez Pidal, *Historia de la lengua española*, vol. II, edición de Diego Catalán, Madrid, Fundación Ramón Menéndez Pidal/Real Academia Española, pp. 7-75.
- MILROY, Leslie y James MILROY (1992): «Social network and social class: Toward an integrated sociolinguistic model», *Language in Society*, 21(1), pp. 1-26, disponible en: <https://doi.org/10.1017/S0047404500015013>.
- NAVARRO TOMÁS, Tomás (1963): «Geografía peninsular de la palabra *agujá*», *Romance Philology*, 17(2), pp. 285-300.
- NGLE = Real Academia Española y Asociación de Academias de la Lengua Española (2009): *Nueva gramática de la lengua española*, Madrid, Espasa.
- O'NEILL, Paul (2015): «The Origin and Spread of Velar Allomorphy in the Spanish Verb: A Morphomic Approach», *Bulletin of Hispanic Studies*, 92, 5, pp. 489-518, disponible en: <https://doi.org/10.3828/bhs.2015.29>.
- OLBERTZ, Hella (2013): «Pues en el español rural de la sierra ecuatoriana: ¿interferencia del quichua?», en Christine Felbeck, André Klump y Johannes Kramer (eds.), *America Romana: Perspektiven transarealer Vernetzungen*, Frankfurt, Peter Lang, pp. 179-214.
- ONÍS, Federico de (1934): *Antología de la Poesía Española e Hispanoamericana (1882-1932)*, Madrid, Publicaciones de la Revista de Filología Española.
- PARODI, Claudia y Otto SANTA ANA (1997): «Tipología de comunidades de habla: del español rural al estándar», *Nueva Revista de Filología Hispánica*, XLV, 2, pp. 305-320.
- Post scriptum = Marquilhas, Rita (coord.) (2014-): *P.S. Post Scriptum. Arquivo Digital de Escrita Quotidiana em Portugal e Espanha na Época Moderna*, disponible en: <http://ps.clul.ul.pt> [Descargado: 04/07/2017].
- PRESEEA = *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América* (2014-): Alcalá de Henares, Universidad de Alcalá, disponible en: <http://preseea.linguas.net> [Consulta: Edad: 16/06/2019; Numerales: 12/06/2019; Pronombres: 12/06/2019; Subjuntivo: 13/07/2019].
- PRESEVAL = *Grupo Val.Es.Co: Proyecto para el Estudio Sociolingüístico del Español de Valencia*, disponible en: <https://www.uv.es/preseval/ppal.htm> [Descargado: 15/04/2014].
- SCHMID, Beatrice e Ivette BÜRKI (2000): «*El bacino imaginado*»: *comedia de Molière en versión judeoespañola. Edición del texto aljamiado, estudio y glosario*, Basilea, Romanischen Seminars der Universität Basel.
- TILG = Santamarina, Antón, Ernesto González Seoane y María Álvarez de la Granja (dirs.): *Tesouro informatizado da lingua galega (Versión 4.1)*, Santiago de Compostela, Instituto da Lingua Galega, disponible en: <http://ilg.usc.es/TILG/> [Consulta: 11/06/2019].
- TRUDGILL, Peter (2009): «Vernacular Universals and the Sociolinguistic Typology of English Dialects», en Markku Filppula, Juhani Klemola y Heli Paulasto (eds.), *Vernacular Universals and Language Contacts*, London, Routledge, pp. 304-322.

APÉNDICE 1: BREVE RESUMEN DE LOS CORPUS USADOS

Corpus	Lengua	Área	Periodo (siglos)	Tipos de textos
<i>CORPES XXI (Corpus del Español del Siglo XXI)</i>	Español	España y América	XXI	Corpus de referencia moderno: textos literarios, ensayísticos, científicos, prensa, blogs, etc. Poco componente oral.
<i>PRESEEA (Proyecto para el Estudio Sociolingüístico del Español de España y América) / ESLORA</i>	Español	España y América	XX-XXI	Entrevistas semidirigidas en centros urbanos.
<i>CDH (Corpus del Diccionario Histórico)</i>	Español	España y América	XII-XX	Corpus de referencia histórico: textos literarios, legales, científicos, prensa, etc. Poco componente oral.
<i>CODEA+2015 (Corpus de Documentos Españoles Anteriores a 1800)</i>	Español	España	XII-XVIII	Documentación legal
<i>CORDLAM (Corpus diacrónico y diatópico del español de América)</i>	España	América	XVI-XIX	Documentos, textos literarios y prensa
<i>COSER (Corpus Oral y Sonoro del Español Rural)</i>	Español	España	XX-XXI	Entrevistas semidirigidas en el mundo rural
<i>Post Scriptum</i>	Español, portugués	España y América, Portugal	XVI-XIX	Cartas privadas de distintas procedencias.
<i>CORHIJE (Corpus Histórico del Judeoespañol)</i>	Judeoespañol	–	XVI-XX	Textos escritos de variada índole (narrativos, religiosos, teatro, etc.)
<i>TILG (Tesouro Informatizado da Língua Galega)</i>	Gallego	Dominio gallegohablante	XVII-XXI	Corpus de referencia histórico: textos escritos (literarios, científicos, actas etc.)
<i>AGO (Arquivo do Galego Oral)</i>	Gallego	Dominio gallegohablante	XX-XXI	Fundamentalmente etnotextos. También grabaciones de habla culta.
<i>CRPC (Corpus de Referência do Português Contemporâneo)</i>	Portugués	Portugal	XIX-XXI	Corpus de referencia moderno: textos literarios, ensayísticos, científicos, prensa, blogs, etc.
<i>CORDIAL-SIN (Corpus Dialectal para o Estudo da Sintaxe)</i>	Portugués	Portugal	XX	Entrevistas semidirigidas en el mundo rural
<i>CTILC (Corpus Textual Informatitzat de la Llengua Catalana)</i>	Catalán	Dominio catalanohablante	XIX-XX	Corpus de referencia moderno: textos literarios, ensayísticos, científicos, prensa, legales, etc.
<i>COD (Corpus Oral Dialectal)</i>	Catalán	Dominio catalanohablante	XX	Entrevistas semidirigidas en enclaves urbanos y semiurbanos