

La colocación de los clíticos pronominales en la libre variación <CL + V mod + INF> frente a <V mod + INF_{CL}> en los textos científicos en español y catalán

Ilpo Kempas

Universidad de Helsinki, Universidad de Turku, Universidad de Ciencias Aplicadas de Seinäjoki
ilpo.kempas@seamk.fi, ilpo.kempas@helsinki.fi

Resumen

En el artículo se estudia empíricamente la colocación de los pronombres reflexivos en las perífrasis modales de infinitivo en español y en catalán. El corpus (N=3.588), compuesto por 10 pares de oposiciones sintácticas en ambos idiomas, representa el discurso científico. Se estudió la posición del pronombre reflexivo con respecto al infinitivo (p. ej., *se puede decir* frente a *puede decirse*) y las posibles diferencias entre el español y el catalán. Para el español, los resultados confirman el fuerte arraigo de la anteposición del pronombre clítico también en textos científicos, pese a su asociación con el lenguaje oral. Además, en su mayoría, los resultados demuestran mucha homogeneidad entre ambas variantes sintácticas. Por el contrario, los resultados sobre el catalán presentan diferencias estadísticamente significativas tanto entre sí como respecto al español. El catalán tiende a la anteposición del pronombre, pero cuando el verbo modal aparece con un sujeto personal, la distribución es mucho más pareja.

Palabras claves: español, catalán, sintaxis pronominal, discurso científico, clíticos pronominales, pronombre reflexivo, libre variación

Abstract

The article is an empirical study on the position of reflexive pronouns in modal periphrases with infinitive in Spanish and Catalan. The corpus (N=3 588), composed of 10 pairs of syntactic oppositions in both languages, represents scientific discourse. The research question was the position of the reflexive pronoun in relation to the infinitive (e.g., *se puede decir* vs. *puede decirse*) and possible differences between Spanish and Catalan. Regarding Spanish, the results confirm that the pre-infinitival placement of reflexive pronouns is frequent in scientific texts, despite being associated with oral language. Furthermore, the results show a high degree of homogeneity between both syntactic variants. The results for Catalan, by contrast, show statistically significant differences both between the different cases analysed and with regard to Spanish. Catalan tends to prefer the pre-infinitival placement of reflexive pronouns, but when a modal verb is used with a personal subject, their distribution is much more even.

Keywords: Spanish, Catalan, pronominal syntax, scientific discourse, clitic pronouns, reflexive pronouns, free variation

1. Introducción

En el artículo se estudia la colocación de los clíticos (“CL”) pronominales, concretamente los pronombres reflexivos, en las perífrasis modales (“PMs”) de infinitivo en español y en catalán. Las lenguas se caracterizan por la libre variación sintáctica de los CLs en las perífrasis de infinitivo y gerundio, pero aquí solo se tratan las de infinitivo (“INF”), porque la colocación de los CLs pronominales en las perífrasis de INF es un tema de múltiples aspectos, imposible de abarcar de forma satisfactoria en el marco de un único estudio empírico, y hemos optado por concentrarnos en

determinados casos (Vs modales, un único CL). El tema es interesante, porque la colocación de los CLs en los PMs no se ha estudiado antes desde una perspectiva contrastiva en ESP y en CAT.

Por consiguiente, como variantes iguales aparecen tanto *se puede decir* como *puede decirse* y *es pot dir* y *pot dir-se*, respectivamente (Tabla 1). Cuando coexisten dos variantes sintácticas en libre variación, a menudo una de ellas es más común que la otra. Con este propósito, se recogió mediante Startpage (un motor de búsqueda que permite obtener resultados del buscador de Google, pero que es más apto que Google para la creación de corpus) en unas mismas fechas un corpus compuesto por todas las ocurrencias disponibles de determinados casos en españoles y catalanes (N=3.588), en los que los mismos CLs pronominales aparecen en ambas posiciones (ver apartado 4). Los ejemplos representan un mismo género, el lenguaje científico, que representa el lenguaje formal y desprovisto de fenómenos del ámbito oral. No obstante, también este género se encuentra en relación de interdependencia con la totalidad de las producciones lingüísticas, y los distintos registros no pueden considerarse totalmente por separado.

La anteposición (“AP”) del CL se conoce como la subida del CL (ingl. *clitic climbing*). La posición enclítica (“PP”), ligada a un verbo (“V”) no finito, se considera como la “posición de base” del CL, de donde, si presentado en forma de árbol sintáctico, el CL “ha subido” a la posición proclítica, delante de un V finito. En este estudio, nos limitamos a aquellas PMs en las que el clítico incontestablemente puede ocupar ambas posiciones sintácticas.

2. Sobre la colocación de los clíticos pronominales en ESP

Las gramáticas y los manuales de ESP concuerdan en presentar ambas variantes morfosintácticas como iguales en las PMs de INF. Butt et al. (2019, p. 147) recalcan la igual aceptabilidad de ambas, pero señalan que la AP sería *más común en el habla espontánea, la lengua de inmediatez* (Koch y Oesterreicher, 2007), que difiere de la *lengua de distancia*, representada por los registros más formales y caracterizada por el uso frecuente de la tercera persona. El lenguaje científico pertenece a esta última. La preferencia por la AP de los CLs en el lenguaje oral espontáneo encuentra apoyo en los estudios de corpus (p. ej., Davies, 1995; Gabor, 2002; Iglesias, 2012; Gudmestad, 2015). Por otro lado, no se pueden encontrar posturas según las cuales la AP sería de *menor uso* en los registros más formales. Illamola y Vila (2015) sintetizan los resultados de varios estudios sobre el discurso oral espontáneo desde la perspectiva panhispánica. Concluyen (pp. 37-38) que:

- 1) las formas proclíticas son absolutamente más frecuentes que las enclíticas en todas las zonas examinadas.

- 2) entre las diferentes áreas estudiadas hay diferencias significativas: el porcentaje de las formas proclíticas oscila entre el 67% (Lima) y el 81% (SJ de Puerto Rico).
- 3) los resultados sobre España son más iguales entre sí (Las Palmas 78%, Sevilla 76%, Madrid 78%) que aquellos sobre Hispanoamérica.

Nosotros analizamos para el presente estudio esos datos estadísticamente, con la prueba de χ^2 , un procedimiento estadístico utilizado para determinar si existe una diferencia significativa entre los resultados esperados y los observados en una o más categorías, contrastando primero Hispanoamérica y España. Entre ambas resultó no existir ninguna diferencia estadística. En el presente estudio nos ocupa un género distinto de la lengua oral, pero este resultado es importante, ya que significa que no es de esperar que los autores hispanoamericanos y los españoles difieran en su manera de colocar los CLs en las PMs de INF. Tampoco hay diferencia estadística entre Las Palmas, Sevilla y Madrid.

Con ciertas perífrasis de INF, los CLs tienden a aparecer en una determinada posición sintáctica. La perífrasis *hay que* se combina normativamente con un pronombre **enclítico** (NGLE 2010, p. 2148):

- (1) La decisión **hay que** tomarla hoy (en lugar de *...la hay que tomar...)

Según la NGLE (p. 2109), también la perífrasis incoativa <ponerse a + INF> requiere la PP del CL:

- (2) Se puso a leer**lo** (*Se **lo** puso a leer).

En (1) y (2) es la PP que se percibe y se presenta como correcta, pero se puede encontrar un caso contrario. Gómez Torrego (1999, p. 3332) señala que tanto la AP como la PP del CL *se* despronominalizado en *las pasivas reflejas con perífrasis* (3a, 3b) es gramatical, pero que su PP es dudosa o agramatical en *las impersonales sintácticas* (4a, 5a):

- (3a) **Se** {pueden/deben/tienen que/han de/empiezan a...} discutir los problemas.

- (3b) {Pueden/deben/tienen que/han de/empiezan a...} discutir**se** los problemas.

- (4a) Aquí **se** {puede/debe/tiene que/ha de/empieza a...} estar bien.

- (4b) *Aquí {puede/debe/tiene que/ha de/empieza a...} estarse bien.

- (5a) Aquí no **se** {puede/debe/tiene que/ha de/suele...} fumar.

- (5b) *Aquí no {puede/debe/tiene que/ha de/suele...} fumarse.

Davies (1995, p. 374) ofrece una prueba contundente de la influencia de la semántica del V auxiliar sobre la colocación de los CLs en las perífrasis de INF: *ir a* favorece más la AP, mientras que *haber*

que no aparece en los resultados (ej. 1). Los Vs modales aquí examinados (*poder, deber/deure, querer/voler, tener que, haber de/haver de*) también pertenecen a aquellos que favorecen la AP.

Gudmestad (2015) estudió la colocación de los CLs con *ir* y *querer* en el lenguaje oral de Caracas. La AP del CL resultó ser significativamente más común con *ir*. La PP del CL era más común con CLs de referencia inanimada, así como con un CL doble (p. ej., *me lo*).

Iglesias (2012) también recalca el papel de los factores léxico-semánticos. Su resultado de interés (p. 188) es que la colocación de los CLs ha experimentado cambios a lo largo de los siglos. En el siglo XIII, la AP casi llegaba al 100%. A partir del s. XV empezó a disminuir, hasta alcanzar un mínimo en el s. XIX, durante el cual los CLs usados con el V modal *deber* aparecen antepuestos tan solo en el 10,8% de los casos (*poder* 19,7% y *querer* 21,8%). Después, la AP ha estado en incremento, y esta tendencia continuará hasta nuestros días.

Davies (1995, p. 376) trata la colocación de los CLs reflexivos. Concluye que los Vs reflexivos asumen la AP del clítico en menor medida que los demás Vs y que entre ambos hay una diferencia significativa ($p=,001$). Los valores obtenidos para los Vs reflexivos por Davies para la AP son el 56,3% y la PP el 43,7% (N=1.683). Los valores para los Vs no reflexivos son para la AP el 63,8% y la PP el 36,2% (N=4.865). Se ve que también en este caso la AP es más común en términos absolutos, si bien la tendencia es más marcada entre los Vs no reflexivos.

El carácter \pm animado del referente del pronombre se ha señalado como un factor que influye en la colocación del CL. Según Myhill (1988a, pp. 241-245; 1988b, pp. 357-361), los CLs animados se anteponen más fácilmente que los inanimados. El autor propone una jerarquía de animacidad (1988a, 242), que fomenta la probabilidad de la subida del CL. Además de Gudmestad (2015) y Myhill (1988a, 1988b), la animacidad es mencionada por Davies (2015). Para la lengua escrita, Davies arrojó 466 (38,6%) ejemplos de la AP y 740 (61,4%) de la PP con referentes animados, mientras que con referentes inanimados, los resultados tienden aún más a la PP (27,2% frente al 72,8%).

3. Sobre la colocación de los clíticos pronominales en catalán

El CAT comparte con el ESP la variación libre en la colocación de los CLs en las PMs de INF. Esta postura es compartida por las presentaciones de la gramática normativa, como *Llibre d'estil de la Universitat Pompeu Fabra*, Gavarró y Laca (2008, p. 2677) y *Gramàtica del català contemporani* (2008) y *Gramàtica de la llengua catalana* (2023).

Si bien en líneas generales la colocación de los CLs en CAT corresponde al ESP, ciertas diferencias morfosintácticas/gramaticales lo distinguen del ESP y, por lo menos en teoría, pueden influir en los resultados cuando se contrastan ambas lenguas. En primer lugar, el CAT posee los adverbios

pronominales *en* e *hi*, ausentes en ESP. En segundo lugar, en CAT la mayoría de los CLs antepuestos y pospuestos no tienen la misma forma (p. ej., *em - me/et - te/el - lo/ens - nos/us - vos/ els - los/ es - se/en - ne*). En tercer lugar, de acuerdo con la gramática normativa, la PP de los CLs conlleva cambios fonéticos que para muchos catalanoparlantes resultan artificiales. En la mayoría de las variedades de CAT, la *-r* final de los INFs es muda, pero reaparece en el caso de un CL pospuesto. Este fenómeno pertenece a la lengua normativizada y es frecuentemente rechazado por los hablantes en situaciones de comunicación oral. De ahí que en la lengua hablada surjan variantes consideradas incorrectas (pero muy frecuentes en el habla espontánea) como *hem d'aixeca'ns* en lugar de *hem d'aixecar-nos*.¹ En definitiva, es necesario tomar en cuenta las diferencias anteriores a la hora de evaluar los resultados.

En cuanto a las restricciones o tendencias opuestas a la variación libre, *Gramàtica de la llengua catalana* (2023, §8.3.3) hace observar la obligatoriedad de colocar juntos los CLs múltiples. Cabe notar que lo mismo también ocurre en ESP. En consecuencia, son posibles *La Júlia pot haver de portar-me-les-hi*, *La Júlia me les hi pot haver de portar* o *La Júlia pot haver-me-les-hi de portar*, pero no **La Júlia em pot haver de portar-les-hi*. Lo arriba expuesto confirma asimismo que el CAT no tiene ningún inconveniente para anteponer y posponer secuencias de múltiples CLs.

Según *Llibre d'estil* (s.f.), con el *perfet perifràstic* y en determinadas perífrasis, ambas posiciones son adecuadas en el registro estándar, pero la PP (p. ej., *el president va adreçar-se als assistents*) es preferible en los registros marcadamente formales. En los menos formales, orales o coloquiales, se prefiere la AP del CL (*el president es va adreçar als assistents*) (ver tb. Fabra, 1956, p. 88; Paradís-Pérez, 2019, p. 36). Ya se ha visto que también en ESP la AP resulta ser marcadamente más común en los registros orales. Es de esperar, pues, que en los documentos científicos catalanes, la formalidad del registro fomente la PP del CL pronominal.

Al tratar las perífrasis verbales del CAT, Gavarró y Laca (2008, pp. 2677-2678) consideran como su denominador común la posibilidad de la AP del CL. En este respecto, difieren de las *secuencias de Vs principales*, que normalmente no la admiten (**La decideixo veure* en vez de *Decideixo veure-la*). No obstante, las autoras recalcan la presencia de la variación individual en estos últimos casos y mencionan (p. 2680) que al menos para algunos hablantes una construcción como *La procurava veure cada dia* está bien formada.

4. Método y materiales

¹ Agradecemos al revisor esta última información.

El enfoque del estudio empírico consistió en limitarse a un número reducido de casos y variables, pero al mismo tiempo conseguir muestras de un tamaño representativo. En consecuencia, no incluimos los factores sociolingüísticos; esto tampoco habría sido posible porque la mayoría de las veces nos faltaban los datos sobre las personas que habían producido los ejemplos. Como se ha visto en el apartado 2, entre España e Hispanoamérica no existe diferencia estadística en el discurso oral, por lo cual tampoco tomamos nota del origen geográfico de los ejemplos: sería aún menos probable encontrar diferencias en el género textual que nos ocupa.

Para estudiar la posición del CL en los textos científicos, se recogió mediante Startpage (s.f.) un corpus (N=3.588) compuesto por distintos casos, esto es, construcciones con el CL pronominal antepuesto y pospuesto (ESP 2.663 ejemplos, CAT 925). Startpage es un motor de búsqueda que utiliza la misma base de resultados que Google, pero es más apropiado para búsquedas realizadas para estudios científicos, porque, entre otras cosas, no incluye anuncios comerciales. La diferencia en el número de ejemplos entre las lenguas se explica simplemente por un mayor número de ejemplos disponibles en español. Los casos, a su vez, se componen de 10 pares de oposiciones sintácticas en ESP y 10 en CAT. Las recogidas se realizaron durante distintas etapas de 2022, según se indicará a continuación. Se recogieron todas las ocurrencias disponibles de cada caso en el momento de la observación. El objetivo era lograr un corpus representativo también en términos numéricos y que permitiera constatar posibles diferencias y tendencias. Las frases empleadas se eligieron por ser comunes en el discurso científico. Las frases usadas en ambas lenguas (23 a 29-1-2022) son:

(6a) ESP **se** puede decir que – puede decirse que

(6b) CAT **es** pot dir que – pot dir-**se** que

(7a) ESP **se** puede concluir que – puede concluirse que

(7b) CAT **es** pot concloure que – pot concloure's que

(8a) ESP **se** debe interpretar como – debe interpretarse como

(8b) CAT **es** deu/deuen interpretar com – deu/deuen interpretar-**se** com²

(9a) ESP **se** puede explicar por – puede explicarse por

(9b) CAT **es** pot explicar per – pot explicar-**se** per

² En CAT, para obtener un mayor número de casos, también se incluyó la forma plural *deuen*. El verbo *deure* seguido de INF en CAT no tiene el mismo significado que el verbo castellano *deber*. Según el *Gran Diccionari de la Llengua Catalana*, se usa como una construcción epistémica y su uso obligatorio es obsoleto, mientras que en ESP los dos sentidos coexisten.

Sintácticamente, en (6a) y (7a) la oración subordinada introducida por *que* constituye una subordinada sustantiva de sujeto. En consecuencia, la PP del *se* despronominalizado es gramatical en estos casos, equiparables con (3a y 3b).

Para asegurar la obtención de resultados precisamente sobre el lenguaje científico, en el campo de búsqueda de Startpage se añadieron los términos *estudio/estudi* y *artículo/article*. De esta manera solo se recuperaron documentos que incluían estos términos.

Después (1 a 9-2-2022), se usaron las frases (10a) y (10b) con el objetivo de estudiar si la inclusión del perfecto compuesto influye en la colocación del CL pronominal y se creó un corpus de ESP y CAT de 1.738 ocurrencias. El tiempo del V predicado no suele considerarse como un factor que influya en la colocación de los CLs, con la única excepción de Sinnott y Smith (2007), en cuyo estudio resultó ser el factor más importante en Bogotá. Queríamos saber si la longitud del elemento verbal de la perífrasis influye en la posición del CL pronominal. Los INFs mencionados a continuación se reconocieron como los más comunes a través de una búsqueda previa. El equivalente CAT de *tener que* es distinto, y corresponde a la expresión española *haber de*, pero en este caso en CAT es la opción más natural. Además, en ESP *haber de* es marcadamente formal.

(10a) ESP “**Nos** hemos querido basar” – “Hemos querido basar**nos**” (etc.)

Vs modales: *querer* (n=665), *poder* (n=531), *tener que* (n=454)

INFs: *querer*: basarse, centrarse, referirse

poder: percatarse, darse, ocuparse

tener que: enfrentarse, limitarse, centrarse

(10b) CAT “**Ens** hem volgut basar” – “Hem volgut basar**-nos**” (etc.)

Vs modales: *voler* (n=146), *poder* (n=106), *haver de* (n=36)

INFs: *voler*: apropiarse, centrarse, limitarse, aturarse

poder: apropiarse, centrarse, adonar-se, fer-se

haver de: centrarse, limitarse, aturarse, contentarse

También se realizó otra búsqueda sobre el CAT (N=154) (4-2-22). El objetivo era estudiar la colocación del CL pronominal en torno a <*ha de* + INF>, la forma impersonal de *haver*, con los Vs *interpretar*, *dir* y *concloure*, p. ej., en *s’ha de interpretar com* frente a *ha de interpretar-se com*.

Por último, para estudiar el equivalente directo ESP de la perífrasis catalana <*haver de* + INF> desde la perspectiva de la colocación de los CLs pronominales, se recogieron todas las ocurrencias (N=453) de la construcción *hemos de* empleada con los Vs *limitarse*, *basarse* y *darse cuenta* (5-8-22). Completamos los datos más tarde (12-1-23), aplicando la búsqueda a estos mismos INFs cuando aparecen con *tenemos que* (N=798).

En los análisis estadísticos se utilizaron la prueba de χ^2 y la prueba U de Mann-Whitney con el umbral de significación de $<,05$.

4. Resultados

Los resultados de las primeras búsquedas se desprenden de la Tabla 2.

Caso	Español (N=760)			Catalán (N=483)		
	Procl.	Encl.	Total	Procl.	Encl.	Total
1	104 (57,5)	77 (42,5)	181	132 (67,7)	63 (32,3)	195
2	115 (56,7)	88 (43,3)	203	134 (91,2)	13 (8,8)	147
3	109 (50,2)	108 (49,8)	217	11 (68,8)	5 (31,3)	16
4	81 (50,9)	78 (49,1)	159	77 (61,6)	48 (38,4)	125
Total	409 (53,8)	351 (46,2)	760	354 (73,3)	129 (26,7)	483
	$\chi^2=3,3$, $p=0,35$, $df=3$ > ninguna diferencia			$\chi^2=36$, $p=0$, $df=3$ > diferencia significativa		

Tabla 2. Ejemplos (N=1.243) recogidos con Startpage (23- 29-1-22).

1 ESP se puede decir que – puede decirse que - CAT es pot dir que – pot dir-se que

2 ESP se puede concluir que – puede concluirse que - CAT es pot concloure que – pot concloure's que

3 ESP se debe interpretar como – debe interpretarse como - CAT es deu/deuen interpretar com – deu/deuen interpretar-se com

4 ESP se puede explicar por – puede explicarse por - CAT es pot explicar per – pot explicar-se per

La AP es predominante en ESP, pero la diferencia no es estadísticamente significativa. Por otro lado, el CAT sí presenta una diferencia significativa ($\chi^2=36$, $p=0$, $df=3$), y podemos concluir que en CAT se prefiere la posición preverbal en estos casos. Además, la diferencia entre el ESP y el CAT también es significativa ($\chi^2=46$, 4; $p=0$, $df=1$;). En consecuencia, al menos sobre la base de este resultado, el CAT tiende a la AP del CL, mientras que el ESP presenta una distribución uniforme. Los siguientes ejemplos ilustran los resultados:

(11a) Desde la óptica del autor, *se puede decir que* la educación ambiental es el proceso

(11b) Pero sí *debe interpretarse como* la respuesta reparativa a un daño

(11c) Per tant, *es pot dir que* les anomenades xarxes socials han multiplicat les...

(11d) De l'experiència viscuda fins ara *pot concloure's que*...

Los resultados de la búsqueda relacionada con los predicados en pretérito perfecto compuesto (PPC) y en *perfet* se ilustran en la Tabla 3.

Caso	Español (n=1.450)			Catalán (n=288)		
	Procl.	Encl.	Total	Procl.	Encl.	Total
1	198 (42,6)	267 (57,4)	465	73 (50)	73 (50)	146
2	242 (45,6)	289 (54,4)	531	52 (49,1)	54 (50,9)	106
3	203 (44,7)	251 (55,3)	454	27 (75)	9 (25)	36
Total	643 (44,3)	807 (55,7)	1450	152 (52,8)	136 (47,2)	288
	$\chi^2=0,9$, $p=0,6$, $df=2 >$ diferencia no significativa			$\chi^2=8,1$, $p=0,02$, $df=2 >$ diferencia significativa		

Tabla 3. Los Vs 1= querer/voler, 2=poder/poder, 3=tener que/haver de en PPC/*perfet* (N=1.738).

“Nos hemos querido basar” – “Hemos querido basar**nos**” (etc.)

“**Ens** hem volgut basar” – ”Hem volgut basar-**nos**” (etc.)

En primer lugar, el ESP prefiere la PP, a diferencia de la Tabla 2. Sin embargo, tampoco este resultado es estadísticamente significativo. Al contrario, la prueba de χ^2 pone de manifiesto una diferencia significativa *entre* las Tablas 2 y 3 ($\chi^2=17,8$, $p=0$, $df=1$). El resultado sugiere que la diferencia impersonal-personal podría explicar la diferencia sistemática observada.

En la Tabla 3, el catalán presenta otra vez una diferencia significativa ($\chi^2=8,1$, $p=,02$, $df=2$), al igual que lo hacen ambas lenguas entre sí ($\chi^2=6,5$, $p=,01$, $df=1$). Sin embargo, la diferencia en CAT se explica únicamente por la perífrasis *haver de*. Sin esta, *el CAT no presentaría diferencia significativa*. Los resultados para *haver de* también difieren estadísticamente de los otros dos, de una distribución más bien uniforme. Este resultado es interesante también ya que el V *poder* favorece claramente la AP del CL en la Tabla 2 (67,7%). Los casos de la Tabla 3 incluyen un sujeto personal, la forma de la primera persona de plural. A continuación, se presentan cuatro ejemplos del corpus:

(12a) **Ens** hem hagut de limitar a fer una observació dels resultats sense poder...

(12b) ...**ens** hem hagut de limitar a analitzar només aquest període de temps.

(12c) En aquest estudi **ens** hem volgut centrar en el paper de la dieta

(12d) Fent aquest treball **hem** pogut adonar-**nos** de la falta de referents sobre...

Luego seguimos estudiando *haver de*. En la Tabla 4 figuran los resultados sobre otras tres frases exactas.

	Procl.	Encl.	Total
1	5	-	5
2	66	11	77
3	63	9	72
Total	134 (87)	20 (13)	154 (100%)

Tabla 4. <S'ha de + INF> frente a <ha de + INF-se> (StartPage, 4-2-22).

1 = s'ha de interpretar com – ha de interpretar-se com

2 = s'ha de dir que – ha de dir-se que

3 = s'ha de concloure que – ha de concloure's que

Se confirma otra vez la preferencia por la AP del CL con esta perífrasis. Notemos, sin embargo, que otra vez los ejemplos incluyen una construcción impersonal, lo que puede fomentar la AP como segundo factor. Los siguientes ejemplos ilustran su uso en el corpus.

(13a) *S'ha de interpretar com* si estiguessin els...

(13b) Però és clar que *s'ha de dir que* els alemanys també ho havien fet

(13c) *s'ha de concloure que* o bé han estat plagiats o bé han estat escrits pel mateix autor

El resultado sobre *haver de* en CAT despertó nuestro interés por examinar la construcción española *haber de*, de la que recogimos todas las ocurrencias disponibles en una misma frase (tabla 5).

	Procl.	Encl.	Total
1	24 (17,5)	112 (82,4)	136
2	12 (8,7)	126 (91,3)	138
3	50 (27,9)	129 (72,1)	179
Total	86 (19)	367 (81)	453
$\chi^2=19$; $p=0$; $df=2$ > Diferencia significativa			

Tabla 5. La PM hemos de (Startpage, 5-8-22).

1. **Nos** hemos de limitar – Hemos de limitarnos

2. **Nos** hemos de dar cuenta – Hemos de darnos cuenta

3. **Nos** hemos de basar – Hemos de basarnos

A diferencia de las Tablas 2 y 3, esta vez también en ESP se registra una diferencia significativa ($\chi^2=19$, $p=0$, $df=2$). Muy curiosamente, este resultado es además del todo contrario al que se obtuvo en CAT. Presentamos los siguientes ejemplos de los resultados:

(14a) *Hemos de limitarnos* pues a recomendar aquello que tiene alguna base científica

(14b) *hemos de darnos cuenta de que* cuanto menor sea la varianza residual...

(14c) para entender el texto bíblico no *hemos de basarnos* únicamente en...

Lo observado con respecto a *hemos de* encuentra apoyo en Schwenter y Cacoullos (2014). Al analizar un corpus oral del ESP hablado en Ciudad de México, los autores concluyeron que las construcciones menos frecuentes como *andar*, *ir + GER*, así como *deber (de)*, ***haber de***, *saber*, *tratar de*, *venir a*, *volver a + INF* favorecen la PP, en contraste con las más frecuentes, como *ir a*, *poder*, *querer*, *tener que + INF*. El resultado de Davies (1995, p. 377) difiere de forma decisiva del nuestro (Tabla 5) y del presentado por Schwenter y Cacoullos (2014). *Haber de* ocupa en la lista de Davies la tercera posición entre los auxiliares asociados con la mayor frecuencia de la AP –incluso sobrepasando a *tener que*–. La discrepancia puede explicarse por diferencias metodológicas, lo que subraya la importancia de examinar los resultados de varios estudios.

Requena (2020, p. 16) pone de manifiesto que la frecuencia de *haber de* ha disminuido significativamente desde el siglo XVIII hasta el siglo XX y que la de *tener que* ha aumentado en la misma proporción. Concluye en su estudio sobre el ESP oral argentino que con *tener que* el CL está pospuesto muy frecuentemente (72%, N=39) por la analogía con *haber que*, que requiere su PPEl autor no menciona la sustitución del futuro sintético por <*haber de + INF*> en (algunas) variedades hispanoamericanas, aparte de *ir a* (p. ej., Sánchez Lobato 1994, p. 563; *Diccionario panhispánico de dudas* 2005, s. v. *haber*, § 3a), que necesitaría tenerse en cuenta. Es evidente que el carácter particular de <*haber de + INF*> reserva sorpresas para los estudios empíricos. Blas Arroyo y Vellón Lahoz (2015) también tratan la sustitución histórica de *haber de* por *tener que*, pero lamentablemente no incluyen la posición de los CLs.

Por último, estudiamos el posible efecto condicionante de la longitud de la parte de la frase que precede al V auxiliar, esto es, si se refleja en la elección de una u otra variante. Calculamos el número de las palabras precedentes en una de las frases usadas y analizamos los resultados con la prueba U de Mann-Whitney (Tabla 6).

	a. media, b. mediana (N de palabras)	N de ocurrencias
[N palabras] ... <i>es pot explicar per</i>	a. 5,34 b. 3	77
[N palabras] ... <i>pot explicar-se per</i>	a. 6,15 b. 4,5	46
Prueba U de Mann-Whitney: U=1569,5; Z=-1,4, p=,16 → Diferencia no significativa.		

Tabla 6. *Catalán: longitud de la secuencia precedente (como número de palabras, de 0 a 24, Startpage, 8 a 9-2-2022).*

Las ocurrencias del segundo caso, la posición posverbal del CL, van precedidas de un ligeramente mayor número de palabras, pero la diferencia no es significativa, y podemos concluir que la longitud de la secuencia precedente no condiciona la elección. Los resultados sobre el ESP son similares (Tabla 7).

	a. media, b. mediana (N de palabras)	N de ocurrencias
[N palabras] ... <i>se puede explicar por</i>	a. 4,1 b. 2	81
[N palabras] ... <i>puede explicarse por</i>	a. 5 b. 2	78
Prueba U de Mann-Whitney: U=2872; Z=-0,99, p=0,3 → Diferencia no significativa.		

Tabla 7. *Español: Longitud de la secuencia precedente (como número de palabras, de 0 a 24, Startpage 7-2-22).*

Comparado con el ESP, el valor de p (0,16) obtenido para el CAT está más cerca del umbral de significancia que el valor correspondiente al ESP (0,3). Por ello, un nuevo análisis podría producir un resultado contrario. Los resultados para ambas lenguas por lo menos *sugieren* que un mayor número de palabras antes del sintagma verbal podría fomentar la PP del CL. Esto sería lógico teniendo en cuenta la posibilidad de que un incremento en el número de palabras pueda hacer que el usuario del idioma desee colocar el CL más lejos, para no “congestionar” demasiado la parte precedente de la oración. De todos modos, habría que estudiar este factor en un mayor número de ejemplos; una muestra de un tamaño reducido no pone de manifiesto tendencias estadísticamente constatables de óptima manera, aunque permita el uso de alguna prueba estadística.

5. Discusión y conclusiones

Para el ESP, los resultados confirman el fuerte arraigo de la AP del CL pronominal también en los textos científicos, pese a su asociación con el lenguaje oral. Además, en su mayoría los resultados demuestran mucha homogeneidad entre ambas variantes, aunque la comparación *entre* los distintos casos pone de manifiesto diferencias, incluso estadísticamente significativas.

La prueba de χ^2 confirma que en las Tablas 2 y 3 no hay diferencia. Las tablas sí presentan una diferencia muy significativa *entre sí*, interpretable por una tendencia del *se* despronominalizado a la

AP en las pasivas reflejas con perífrasis (Tabla 2). Esta observación debería confirmarse a través del análisis de nuevos datos, pero a estas alturas nos permitimos proponer una explicación. Como se ha visto, las impersonales sintácticas (ejs. 4 y 5) no aprueban la PP del CL. Proponemos que las pasivas reflejas con perífrasis (3a, 3b) favorecen en cierta medida la AP por analogía con las impersonales sintácticas. Aunque los dos tipos difieren en cierta medida en el plano semántico, sintácticamente son idénticos. Además, dicha diferencia semántica entre ambos puede parecerle inexistente al hablante, ya que en ambos casos se trata de una construcción impersonal y pasiva de tercera persona. En consecuencia, el hecho de que las impersonales sintácticas requieran la AP de *se* hace que las pasivas reflejas se comporten igual.

En el apartado 2 se discute la animacidad como factor que fomenta la AP del CL. Los resultados de la Tabla 2, con el *se* despronominalizado, no permiten examinarlo, pero los de la Tabla 3 ofrecen un buen punto de comparación con los de Davies (2015), mencionados en el apartado 2. La prueba de χ^2 pone de manifiesto una diferencia significativa entre ambos ($\chi^2=8,6$; $p=0,003$; $df=1$); Davies: AP 38,6%, PP 61,4% frente a AP 44,3%, PP 55,7% (Tabla 3). En consecuencia, nuestros resultados tienden en algo menor medida a la PP –aunque se componen de CLs reflexivos, que deberían favorecer la posición enclítica–. Lo mismo se repite al contrastar los resultados de Davies con los de la Tabla 8 (ver más abajo): AP 47,1%, PP 52,9% ($\chi^2=14,2$; $p=0$; $df=1$). Esto podría explicarse por diferencias metodológicas. Evidentemente, Davies no incluyó los Vs reflexivos en la categoría \pm animacidad (lengua escrita), ya que incluye 1.883 ejemplos, mientras que la de \pm reflexividad cuenta con 1.634. Por consiguiente, el efecto combinado de varios factores puede producir resultados que difieren de los registrados para un único factor.

En general, los resultados sobre el ESP confirman un importante equilibrio entre la AP y la PP del CL, a pesar de las diferencias entre los casos ilustrados por las Tablas 2 y 3. No obstante, la Tabla 5 con *<haber de + INF>* difiere notablemente del patrón general, y, curiosamente, por su fuerte tendencia a la PP, produce además un resultado opuesto al CAT.

Con *tener que*, la PP (55,3%) es más frecuente en la Tabla 3, y la AP corresponde al 44,7% de los casos (frente al 19% para *haber de*). No obstante, para confirmarlo una vez más, creamos una muestra con *tener que* (Tabla 8) con los elementos idénticos a la Tabla 5 (*haber de*).

	Procl.	Encl.	Total
1	144 (49)	150 (51)	294
2	102 (44)	130 (56)	232
3	130 (47,8)	142 (52,2)	272

Total	376 (47,1)	422 (52,9)	798
$\chi^2=1,4$; $p=0,5$; $df=2 >$ Diferencia no significativa			

Tabla 8. La PM tenemos que (*Startpage*, 12-1-23).

1. **Nos** tenemos que limitar – Tenemos que limitarnos
2. **Nos** tenemos que dar cuenta – Tenemos que darnos cuenta
3. **Nos** tenemos que basar – Tenemos que basarnos

La PP del CL resulta más común que su AP, pero la prueba del χ^2 permite otra vez excluir una diferencia significativa ($p=0,5$). *Haber de* se comporta de una manera distinta de las demás PMs. Comparado con *tener que*, es menos frecuente en ESP, tanto en la lengua oral como en la escrita (cf. Blas Arroyo y Vellón Lahoz, 2015). Esto también puede reflejarse en la colocación del CL pronominal. Hemos visto (apartado 2) que la colocación general de los CLs ha experimentado cambios históricos, de tal manera que la AP disminuyó del siglo XV hasta el XIX (Iglesias, 2012, p. 188). Durante esa época, la frecuencia de *haber de* disminuyó a partir del siglo XVIII al XX (Requena, 2020, p. 16). La AP ha estado en aumento desde el XIX, pero durante este periodo la disminución de la frecuencia de *haber de* parece haber continuado al menos hasta el XX. Es posible que la PP de los CLs con este V sea una especie de “arcaísmo”: como *hemos de* ha quedado en un segundo plano, no habría participado en la reintroducción de la AP desde el XIX de la misma manera que los Vs más frecuentes. De todas formas, la colocación de los CLs con *haber de* requeriría un nuevo estudio, que también tome en cuenta la observación de Davies (1994, p. 377) de su frecuencia sorprendentemente elevada, así como su uso para expresar el futuro (Sánchez Lobato, 1994, p. 563) en el ESP de América.

El análisis de la supuesta influencia de la longitud de la secuencia precedente a la perífrasis de INF en la colocación del CL pronominal (Tabla 7) produjo un resultado negativo. Este factor no ha sido antes objeto de estudios. Aunque con los datos usados, la prueba U de Mann-Whitney dio un resultado negativo, no se puede excluir la posibilidad de que su aplicación a un corpus más grande pueda producir un resultado diferente.

Los resultados sobre el CAT presentan diferencias estadísticamente significativas tanto entre sí como respecto al ESP. Se trata siempre de una tendencia a la AP del CL. En las Tablas 2 y 3, esta tendencia constituye una diferencia estadística respecto al ESP. No obstante, en los casos en que el V modal aparece *con un sujeto personal*, la distribución es mucho más pareja (Tabla 3), y dicha diferencia a favor de la AP no puede considerarse como absoluta. Hemos visto que, en la Tabla 3, la

diferencia se explica por una única PM, *haber de*, que demostradamente favorece la AP del CL –al revés de *haber de* en ESP (Tabla 5)–.

La construcción impersonal, con el *se* despronominalizado, resulta favorecer la AP del CL en CAT, al igual que en ESP. Precisamente, el foco del presente estudio en el lenguaje científico hace que no dispongamos de un número representativo de ejemplos usados con una forma personal; esto es un tema para un estudio futuro.

La prueba estadística realizada para la Tabla 6, relacionada con el supuesto efecto condicionante de la secuencia inmediatamente precedente a la perífrasis con INF, excluye su papel. No obstante, un valor de p de ,16 basado en solo 123 ejemplos *puede* dar un resultado distinto con una muestra más grande. Podría plantearse la hipótesis de que el incremento en el número de palabras fomenta la PP del CL, para que lo que precede no se vuelva demasiado pesado:

(15a) D'altra banda, la relació observada entre els símptomes de TDAH i el diagnòstic de TOC subclínic, també pot explicar-se per ser dos trastorns que...

(15b) El fet de trobar-se la forma sevici en el text de la Sibilla publicat per Moliné i Brasés, tant pot explicar-se per...

frente a

(16a) Això es pot explicar per una ponderació major de la...

(16b) Aquest fet es pot explicar per diversos motius

Un futuro estudio, con más ejemplos, permitirá verificar o falsar esta última hipótesis. De todos modos, la tendencia del CAT a la AP de los CLs, que constituye una diferencia respecto al ESP, es evidente. Como hemos dicho, el corpus del presente estudio no es suficiente para describir y explicar esta diferencia de forma satisfactoria. Hacen falta resultados sobre una gama más amplia de casos. La Tabla 3, con un sujeto personal, proporciona resultados que difieren en cierta medida del patrón observado. Aunque se necesitan más resultados empíricos, se nos ocurre la posibilidad de que, en CAT, la AP frecuente del CL en la lengua oral influya en la lengua escrita en mayor medida que lo hace en ESP (cf. apartado 3). Inversamente, la lengua escrita en español se habría mantenido más conservadora por haber resistido más a la subida del CL.

Referencias

Bermúdez, F. (2006). La “subida de clíticos”: modalidad, prominencia y evidencialidad. *Lexis* XXX(1), 83-115.

- Blas Arroyo, J. L. & Vellón Lahoz, J. (2015). The refuge of a dying variant within the grammar: Patterns of change and continuity in the Spanish verbal periphrasis *haber de* + infinitive over the past two centuries. *Language Variation and Change*, 27, 89–116.
- Butt, J., Benjamin, C. y Moreira Rodríguez, A. (2019). *A New Reference Grammar of Modern Spanish* (6th ed). Routledge.
- Davies, M. (1995). Analyzing Syntactic Variation with Computer-Based Corpora: The Case of Modern Spanish Clitic Climbing. *Hispania*, 78(2), 370-380.
- Diccionario panhispánico de dudas* (2005). Real Academia Española. <https://www.rae.es/dpd/haber>
- Fabra, P. (1956). *Gramàtica catalana*. Teide.
- Gábor, K. (2002). *Subida de clítics en corpus electrònics: anàlisi automàtic de textos en espanyol peninsular modern* [Tesis doctoral, Universidad de Szeged].
- Gavarró, A. y Laca, B. (2008). Les perífrasis temporals, aspectuals i modals. En J. Solà, M.-R. Lloret, J. Mascarò, M. Pérez Saldanya (dirs.) amb la col·laboració de G. Rigau, *Gramàtica del català contemporani* (Quarta ed.). Volum 3. Editorial Empúries.
- Gómez Torrego, L. (1999). Los Vs auxiliares. Las perífrasis verbales de infinitivo. I. Bosque y V. Demonte (eds.), *Gramática descriptiva de la lengua española, Vol. II*, (pp. 3323-3390). Espasa Calpe.
- Gramàtica de la llengua catalana* (2023). Institut d'Estudis Catalans. DOI: 10.2436/10.2500.08.1
- Gran Diccionari de la llengua catalana* (s.f.). Grup Enciclopèdia. <https://www.diccionari.cat/gran-diccionari-de-la-llengua-catalana>
- Gudmestad, A. (2005). Clitic climbing in Caracas Spanish: A sociolinguistic study of *ir* and *querer*. *Indiana University Linguistics Club Working Papers*, 6, 1–14.
- Iglesias, O. (2012). *Le placement des clitiques dans les complexes verbaux en espagnol : une nouvelle approche de la question* [tesis doctoral, Université Paris 8 VINCENNES-SAINT DENIS].
- Illamola, C. y Vila, F. X. (2015). Análisis de la posición de los pronombres átonos en construcciones perifrásticas entre escolares de la Región Metropolitana de Barcelona. *Diálogo de la Lengua*, VII, 36-57.
- Koch, P. & Oesterreicher, W. (2007). *Lengua hablada en la Romania: español, francés, italiano (Versión española A. López)*. Gredos.
- Llibre d'estil de la Universitat Pompeu Fabra*. <https://www.upf.edu/web/llibre-estil/pronoms>
- Myhill, J. (1988a). Variation in Spanish Clitic Climbing. *Georgetown University Round Table on Language and Linguistics*, 227–250.
- Myhill, John (1988b). The Grammaticalization of Auxiliaries: Spanish Clitic Climbing. *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society* (1988), 352–363
- Nueva gramática de la lengua española (NGLE)* (2010). Sintaxis II. Real Academia Española. Espasa Libros.
- Paradís Pérez, Anna (2019). *L'ascens dels clítics: reestructuració i control*. Universitat Autònoma de Barcelona. Departament de Filologia Catalana. <http://hdl.handle.net/10803/668058>
- Requena, P. E. (2020). A Usage-Based Perspective on Spanish Variable Clitic Placement. *Languages*, 5(3), 33. <https://doi.org/10.3390/languages5030033>.
- Sánchez Lobato, J. (1994). El español en América. En J. Sánchez Lobato e I. Santos Gargallo (coords.), *Problemas y métodos en la enseñanza del español como lengua extranjera: actas del IV Congreso Internacional de ASELE* (Asociación para la Enseñanza del Español como Lengua Extranjera), 553-570.
- Schwenter, S. A. y Torres Cacoullós, R. (2014). Competing constraints on the variable placement of direct object clitics in Mexico City Spanish. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 27(2), 514–536.

Sinnott, S. y Smith, E. (2007). *¿Subir o no subir? A look at clitic climbing in Spanish*. MS., The Ohio State University and Pikeville College.
Startpage. (s.f.). <https://www.startpage.com/>