B O R E A L I S
An International Journal of Hispanic Linguistics

JO DIC, DIGO YO: ESTUDIO COMPARATIVO DE CONSTRUCCIONES CON EL VERBO *DECIR* EN ESPAÑOL Y *DIR* EN CATALÁN HABLADOS*

Peter Herbeck[a] & Pekka Posio[b]
Universidad de Viena[a], Universidad de Helsinki[b]

RESUMEN. En este artículo se investiga la expresión de sujeto en dos lenguas de sujeto nulo canónicas – castellano y catalán. Analizamos diferentes usos del verbo *decir/dir* en dos corpus orales: el corpus *PRESEEA* (Madrid y Alcalá) y el *Corpus Oral de Conversa Col·loquial* (*COC*). De especial importancia es el uso del *verbum dicendi* en contextos de estilo directo, ya que observamos un grado de pragmaticalización hacia partículas citativas en las dos lenguas, lo que contribuye a la alta frecuencia de sujetos nulos. Se mostrará que las tendencias en los dos corpus son parecidas y que las diferencias observadas se deben, en su mayoría, a factores discursivos y diafásicos, y no a factores morfosintácticos. Comparamos los resultados con estudios previos sobre otras variedades románicas, como el portugués europeo y el valenciano. Aunque en estas variedades se observan patrones diferentes con respecto a la expresión y/o posición del sujeto, argumentamos que en todas ellas es vital considerar la existencia de secuencias rutinizadas y formulaicas para entender la variación en la expresión de los sujetos.

Palabras clave: expresión del sujeto, verbos de comunicación, estilo directo, estudio de corpus, castellano, catalán, pragmaticalización

ABSTRACT. This article investigates subject expression in two canonical null-subject languages – Spanish and Catalan. We analyze different uses of the verb *decir/dir* in two oral corpora: the PRESEEA corpus (Madrid and Alcalá) and the Corpus Oral de Conversa Col·loquial (COC). Particularly important is the use of the *verbum dicendi* in direct-speech contexts, since we observe a degree of pragmaticalization toward citative particles in both languages, which contributes to the high frequency of null subjects. It will be shown that the tendencies in the two corpora are similar and that the differences observed are mostly due to discursive and diaphasic factors rather than morphosyntactic ones. We compare the results with previous studies on other Romance varieties, such as European Portuguese and Valencian. Although these varieties display different patterns regarding the expression and/or position of the subject, we argue that in all of them it is crucial to consider the existence of routinized and formulaic sequences in order to understand variation in subject expression.

Keywords: subject expression, communication verbs, direct speech, corpus study, Spanish, Catalan, pragmaticalization

1. Introducción y estado de la cuestión

Las expresiones formulaicas o gramaticalizadas basadas en los verbos más frecuentemente utilizados en el habla han sido objeto de no pocos estudios en los últimos años. En particular, el verbo *decir* en español ha llamado la atención de algunos investigadores (véanse, por ejemplo, Díaz & Labarca 2010, Posio & Pešková 2020, Herbeck & Posio 2022, Martínez Vázquez 2023). Estos estudios han demostrado que los verbos más frecuentes tienden a presentar características idiosincráticas en cuanto a

* Queremos agradecer a Suvi Tynnilä e Ignasi Planas Villalba su trabajo minucioso en la codificación de los datos del castellano y catalán, respectivamente. En el caso de Peter Herbeck, esta investigación fue financiada en parte por el *Österreichischer Wissenschaftsfonds* (FWF) 10.55776/F1003 (P08).

© Peter Herbeck & Pekka Posio. *Borealis: An International Journal of Hispanic Linguistics*, 2025, 14 / 2. pp. 211-238. https://doi.org/10.7557/1.14.2.8250

This is an Open Access Article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/legalcode) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

la expresión y colocación de sujetos. Por ejemplo, el verbo *decir* se caracteriza por una alta frecuencia de sujetos posverbales en el español peninsular (Posio & Pešková 2020), aunque no pertenece a ninguna de las clases de verbos que se suelen considerar como favorecedores de sujetos posverbales, como los verbos inacusativos o las construcciones existenciales (ver, entre otros, Torrego, 1984, Fernández Soriano 1993, Cuartero Sánchez 2007 para el castellano; Picallo 1984; Bel 2003; Cabré Sans & Gavarró 2007 para el catalán,). También presenta algunas diferencias con respecto a la expresión de sujetos: en español peninsular, *decir* se caracteriza por una frecuencia relativamente baja de sujetos expresados, mientras que en el español argentino, no difiere de la tendencia general (Posio & Pešková 2020) y en el español colombiano, presenta frecuencias relativamente altas de expresión de sujetos pronominales (Orozco 2015).

La expresión y omisión de los sujetos pronominales es uno de los fenómenos más estudiados en la lingüística hispánica. Existe un consenso general sobre los factores que influyen en dicha variación (Silva-Corvalán 2001, 154–169). Los factores más importantes son la continuidad referencial (es decir, si el sujeto es correferencial con el sujeto anterior u otro antecedente, es menos probable que se exprese el sujeto) y la persona gramatical (la expresión es más frecuente en la primera y segunda persona del singular que en las demás personas). Los sujetos se expresan menos frecuentemente en las personas plurales y en los tiempos perfectivos, y más frecuentemente en las formas imperfectivas o irrealis. También influye la existencia de construcciones frecuentes de sujeto-verbo: por ejemplo, la secuencia *yo creo* [*que*] favorece la expresión del sujeto, mientras que la secuencia *no sé* la desfavorece. En las frecuencias generales de sujetos expresados y no expresados hay bastante variación entre los dialectos del español (Otheguy & Zentella 2012), así como entre las diferentes lenguas (Posio 2012), pero los factores que influyen en la expresión del sujeto suelen ser los mismos. Sin embargo, es justamente con los verbos más frecuentemente utilizados y las secuencias formulaicas con funciones pragmáticas donde se encuentran diferencias a veces inesperadas entre variedades y lenguas, como es el caso del verbo *decir* en español europeo y argentino (Posio & Pešková 2020).

Por falta de espacio, no profundizaremos aquí en la cuestión de la expresión/omisión del sujeto pronominal en general, ya que nuestro objetivo es describir las particularidades de un solo verbo. Para lectores interesados en la variación general en la expresión de sujetos, se recomiendan artículos de reseña y obras editadas recientes como Carvalho, Orozco & Shin (2015), Posio (2018), Silva-Corvalán (2001, 154–168) y Posio & Herbeck (2023).

Hasta ahora, la mayoría de los estudios sobre los usos pragmáticos y discursivos sobre el verbo *decir* se han concentrado en el español peninsular. Posio & Pešková (2020) comparan el uso de *decir* en la variedad peninsular y la argentina y revelan diferencias significativas en los usos gramaticalizados como marcadores de citas. Sabemos, pues, que los patrones morfosintácticos de los verbos frecuentemente utilizados tienden a diferenciarse en diferentes variedades del español (Orozco 2015) y entre el español y portugués europeo (Posio 2012). En comparación con el español, los estudios sobre el catalán han sido menos frecuentes. Casanova (1998) examina la expresión del sujeto en el catalán oral de forma cuantitativa a nivel global y observa que la frecuencia general de sujetos explícitos (incluyendo los pronominales y nominales) es de un 28 %. Bentivoglio (1987) da el 40 % como un porcentaje general de sujetos expresados en el español hablado de Caracas, mientras que, según los datos de Enríquez (1984, 348), en el español hablado de Madrid los porcentajes de sujetos pronominales expresados varían de un 10,40 % en la primera persona del plural a un

31,89 % en la primera persona del singular, y hasta un 72,70 % para el pronombre *usted*.[1]

De Prada Pérez (2015) ha analizado la expresión de los sujetos pronominales en la situación bilingüe del español y catalán menorquines en contacto. La autora no observa diferencias significativas entre las dos lenguas en la expresión del sujeto de 1ª persona del singular (un 21,3 % en los hablantes bilingües dominantes en catalán y un 20,7 % en el grupo de control catalán, comparado con un 19,8 % del grupo de control español y un 19,9 % del grupo de bilingües dominantes en español). Además, existen estudios enfocados en la adquisición de los pronombres (implícitos) de sujeto en catalán (Bel 2003; Cabré Sans & Gavarro 2006). En el estudio de Bel (2003), los niños producen sujetos explícitos con una frecuencia de un 32,3 % en catalán y de un 32,7 % en español (Bel 2003, 7). Según Cabré Sans y Gavarró (2006), los niños producen un 31,08 % de sujetos explícitos, comparado con un 34,17 % en el habla de los adultos, y no se observa ninguna diferencia significativa entre los dos grupos. No obstante, estudios de la expresión de sujeto con formas verbales frecuentes (como *(yo) creo*, *digo*, etc.), aunque abundantes en la literatura sobre el castellano, han sido más escasos en el caso del catalán. Herbeck (2022) investiga la expresión del sujeto de 1ª persona singular en el catalán valenciano, y llama la atención sobre la alta frecuencia de sujetos expresados (57 %) en la expresión *jo dic*, sobre todo cuando introduce un complemento de estilo directo. Sin embargo, mientras en castellano *digo* en esta función suele usarse sin sujeto expreso o con un sujeto posverbal, en el corpus del catalán de Valencia la forma *dic* suele ir acompañado por el sujeto *jo* en la posición preverbal. Surge, entonces, la pregunta sobre la existencia de diferencias interlingüísticas en la expresión de sujeto con expresiones formulaicas y verbos altamente frecuentes en la lengua oral entre el español y el catalán en sus variedades peninsulares.

El objetivo del presente estudio es analizar y comparar el uso del verbo *decir* en español y *dir* en catalán en dos corpus de lengua oral, el *PRESEEA* (los subcorpus de Madrid y Alcalá) y el *Corpus Oral de Conversa Col·loquial* (*COC*; Payrató & Alturo 2002), teniendo en cuenta factores lingüísticos y pragmáticos que pueden afectar la expresión y colocación de los sujetos. Trataremos los sujetos de todas las personas gramaticales, aunque el mayor enfoque será en la primera y segunda persona del singular, ya que estas personas gramaticales son las que más frecuentemente se usan en el discurso oral y las que también presentan usos pragmáticos especializados.

El artículo está estructurado de la siguiente manera: en el apartado 2 presentaremos el marco teórico y en el 3 los datos analizados. El apartado 4 presenta los resultados del análisis, el 5 la discusión y el 6 las conclusiones del estudio.

2. Marco teórico: frecuencia de uso y construccionalización

Como ya se adelantó en la introducción, en el presente estudio partimos del marco teórico de gramaticalización de construcciones o construccionalización, entendido como un proceso de cambio lingüístico que afecta a construcciones más amplias en vez de a solo palabras o morfemas individuales e implica cambios simultáneos en la forma y significado (véanse Hilpert 2013, Traugott y Trousdale 2013, Trousdale 2014, Enghels 2018). Este concepto enfatiza la existencia simultánea de varios grados de construccionalización, desde usos completamente productivos de los lexemas –en

[1] En general, resulta difícil establecer comparaciones entre las frecuencias mencionadas en diferentes fuentes, debido a las diferencias en los criterios de inclusión de diferentes tipos de sujetos y en la manera de cuantificarlos en cada estudio: además, dar un porcentaje general para todas las personas gramaticales es poco informativo, ya que la persona gramatical es uno de los factores que más influye en las tendencias de expresión y omisión de sujetos (ver, entre otros, Posio 2015).

nuestro caso, el verbo *decir/dir* con toda la flexión de persona, número, tiempo y modo que presenta, además de la posibilidad de expresión preverbal y posverbal u omisión del sujeto y de otros argumentos– hasta usos completamente fijados que encontramos en los marcadores discursivos deverbales *es decir* / *és a dir* (cf. Martín Zorraquino Portolés 1999; Cuenca 2003). Las partículas citativas *digo* y *dice* representarían un caso intermedio de construccionalización, porque siguen teniendo formas personales de primera y tercera persona, pero ya no presentan variabilidad en cuanto al tiempo o modo verbal o la expresión de argumentos (Posio & Pešková 2020 para el español; ver la sección 4.3 del presente estudio para *dic* y *diu* del catalán).

La alta frecuencia de uso de *decir* en la lengua hablada parece ser un factor clave en este proceso. En castellano, el verbo *decir* y sobre todo las formas *dice* y *digo* se encuentran entre las formas verbales más frecuentemente repetidas en el discurso hablado, lo que, según la teoría de los efectos de frecuencia (Bybee y Thompson 1997), favorece la pérdida de variabilidad característica de los procesos de gramaticalización, ya que los hablantes tienden a memorizar y repetir las formas frecuentes. Dicho de otro modo, la alta frecuencia de casos (*token frequency*) conlleva una baja frecuencia de tipos (*type frequency*). Se sabe que los verbos más frecuentemente utilizados, como los verbos copulativos o auxiliares, pero también los que desempeñan diferentes funciones pragmáticas como *decir, saber* o *creer,* son los que se desvían de los patrones generales de expresión del sujeto (Erker & Guy 2012, Posio 2015). Los verbos mentales o comunicativos frecuentes como *creer* o *decir* muestran cierto grado de construccionalización (ver, por ejemplo, Wiemer & Bisang 2004, Traugott & Trousdale 2013) o pragmaticalización (véase Aijmer 1997, Detges & Waltereit 2016). Es decir, han desarrollado patrones fijos y significados (o funciones) propios que los distinguen de los usos de los mismos lexemas verbales como formas productivas (ver Enghels 2018).

Por ejemplo, la forma *digo* puede servir como marcador de autoenmienda iniciada por el hablante (ejemplo 1a) o como marcador de cita directa (ejemplo 1b; véanse Martín Zorraquino & Portolés 1999, Benavent 2003, Posio 2015, Posio & Pešková 2020). Aunque estos usos obviamente guardan relación con el significado semántico del verbo *decir,* el hecho de que se trate de un marcador gramaticalizado y no de un uso productivo del verbo se refleja en la fijación de la forma y significado específico. El carácter gramaticalizado de la construcción citativa se ve claramente en ejemplos como (1c), donde la forma *dice* se usa ante la cita a pesar de la presencia de un verbo comunicativo conjugado en la primera persona del singular en la misma oración. Por último, la secuencia *digo yo* con el sujeto pronominal posverbal puede servir para expresar atenuación de la perspectiva epistémica del hablante, como en el ejemplo (1d), aunque también puede aparecer como marcador de cita, como en el ejemplo 1c, si bien en esta función es menos frecuente (Martínez Vázquez 2023).

(1) Ejemplos de construcciones con *decir* en español peninsular (todos los ejemplos son del corpus CORLEC):
   a. *– No, ese color no me gusta. –¿Larga?,* ***digo*** *¿blanca?* (Posio & Pešková 2020, 399)
   b. *Es que he dicho,* ***digo****: "Es que no puede ser, no lo entiendo."* (Posio & Pešková 2020, 399)
   c. *Me preguntó Marce el otro día,* ***dice****: "¿Y las palomas? Digo yo: "¿Qué palomas?"* (Posio & Pešková 2020, 406)
   d. *Será que ahora colecciona bastones.* ***Digo yo****, no, no tengo ni idea, vamos.* (Posio 2012, 173)

El hecho de que los hablantes asocien las diferentes construcciones con *decir* con funciones pragmáticas como autoenmienda, atenuación o marcación de citas se conecta con la frecuencia de uso del verbo en la lengua hablada. Por un lado, la necesidad de estas funciones en la comunicación oral hace que las formas de *decir* aparezcan con mucha frecuencia en el discurso. Por otro lado, la frecuente repetición de las construcciones –combinaciones de formas y funciones– hace que las formas repetidas se afiancen y empiecen a tener significados especiales que no se derivan solo del significado léxico del verbo. Según la teoría de los efectos de frecuencia (Bybee y Thompson 1997), la alta frecuencia de uso resultará en la fijación formal y reducción en la variabilidad de las construcciones. En la última instancia, esto puede llevar a la gramaticalización de formas como *digo* y *dice* en su función citativa (Posio & Pešková 2020).

3. Datos

El objetivo de esta investigación es realizar un análisis empírico contrastivo del uso de los sujetos implícitos y explícitos con el verbo *decir*/*dir* en español y catalán oral. Por un lado, se intentan descubrir los factores que determinan el uso del sujeto en las respectivas lenguas, ya que se ha observado un comportamiento especial de la expresión del sujeto de *decir* en estudios previos. Por otro lado, queremos ofrecer una visión contrastiva de dos lenguas románicas de sujeto nulo "canónicas" (en el sentido de Holmberg 2005, Holmberg et al. 2009) que no muestran diferencias respecto a la "riqueza" del sistema de concordancia de verbo-sujeto.

Para este fin, hemos analizado los datos de los subcorpus de Alcalá y Madrid de *PRESEEA* para el castellano y las 10 entrevistas disponibles en línea del *Corpus Oral de Conversa Col·loquial* (*COC*) del catalán oral (Payrató & Alturo 2002). En el caso de los dos corpus se trata de conversaciones orales. No obstante, existen diferencias en cuanto al tipo de discurso que se deberían tener en cuenta a la hora de interpretar los datos. El corpus *PRESEEA* es un corpus sociolingüístico y contiene conversaciones semidirigidas, es decir, se trata de una conversación entre un entrevistador y un informante. Los temas de todas las conversaciones se guían según nueve categorías: saludos, el tiempo, lugar donde vive el informante, familia y amistad, costumbres, peligro de muerte, anécdotas importantes de la vida, deseo de mejora económica, final (véase Moreno Fernández 2021, 18–19). Los participantes son de tres grupos etarios y tres niveles socioeconómicos diferentes.

En el caso del *COC*, se trata de conversaciones coloquiales libres entre los informantes (entre 3 y 11 participantes) en situaciones comunicativas cotidianas, por ejemplo, durante la cena, la comida, delante de la televisión, durante el café, etc. (véase Payrató & Arturo 2002, 13). Los hablantes son de la clase social media, y el 70 % de ellos tiene una edad de entre 20 y 40 años (Payrató & Arturo 2002, 9). Los hablantes del *COC* representan el catalán oriental (central), sobre todo de la región de Barcelona, que contrasta con el estudio de Herbeck (2022) sobre el catalán occidental (valenciano).

Los datos de los dos corpus se han extraído buscando por todas las formas conjugadas de *decir/dir*. Después, se han revisado y codificado de forma manual. Todos los datos del *PRESEEA* se han clasificado de forma manual por un anotador y han sido revisados por los autores del presente estudio. Los datos del catalán han sido anotados por el primer autor del estudio y los casos difíciles (como el tipo de complemento, la distinción entre estilo directo y no directo en datos orales) se han comprobado teniendo en cuenta la codificación por otro anotador. Para la anotación de los datos del castellano

y catalán, se han seguido las mismas categorías que expondremos en el apartado siguiente.

3.1. Categorías de anotación

Investigamos la variable dependiente "expresión del sujeto" con los valores "implícito" vs. "explícito" con relación a las siguientes variables independientes: (i) persona/número gramatical, (ii) tiempo verbal, (iii) tipo de complemento, (iv) argumento dativo, (v) tipo de oración:

Tabla 1: Categorías de anotación

variable/factor	**valores**
expresión del sujeto	implícito vs. explícito
persona/número gramatical	1sg, 2sg, 3sg, 1pl, 2pl, 3pl
tiempo verbal	presente de indicativo, pasado perfectivo, pasado imperfectivo, subjuntivo (presente e imperfecto), condicional, futuro, formas infinitivas, perífrasis
tipo de complemento	complemento de estilo directo, estilo indirecto, Sintagma Nominal, clítico, pronombre (personal, indefinido, demostrativo, …), sin complemento
expresión del argumento dativo	sí/no
tipo de oración	principal declarativa, subordinada, principal interrogativa, coordinada
elementos preverbales	adverbios, marcadores discursivos, etc.

Cuando un sujeto es explícito, también se ha anotado su posición preverbal vs. postverbal. No obstante, como el enfoque del presente estudio es la expresión del sujeto, presentaremos estos resultados solamente de forma cualitativa en algunos puntos relevantes. También se han anotado algunas características de forma cualitativa, como el uso de formas verbales como partícula citativa (véanse los ejemplos 1b y 1c) .

El factor de la persona gramatical es primordial, ya que se ha observado en varios estudios previos que la 1ª y la 2ª persona de algunos verbos tienen comportamientos idiosincráticos en la expresión de sujeto (por ejemplo, con los verbos de cognición *creer* y *saber*; véase Posio 2015). Por otro lado, se esperaría que los factores que determinan la expresión de sujeto de tercera persona fuesen diferentes de los de la 1ª y 2ª persona, ya que aquella tiene referencia anafórica en el discurso mientras estas son deícticas.[2] Además, el análisis de los sujetos de 3ª persona incluye los sintagmas nominales plenos y nombres propios mientras que la 1ª y 2ª persona singular incluyen solamente sujetos pronominales (nulos y fuertes). En cuanto al tiempo y aspecto verbal, se ha observado en algunos estudios que las formas imperfectivas tienen una frecuencia de expresión de

[2] En este artículo, investigamos la expresión variable del sujeto con el verbo *decir* y algunos usos semifijados. Por lo tanto, excluimos algunos datos en los que la alternancia entre un sujeto implícito y explícito es imposible. Por ejemplo, no forman parte de nuestro estudio las formas verbales infinitivas, ya que no permiten sujetos explícitos de forma libre. Tampoco analizamos secuencias fijadas como *és a dir* o *vol dir* del catalán y *es decir* o *quiere decir* del castellano. En el corpus del catalán, las formas de *voler dir* y sus variantes son frecuentes (110 casos). Es bien sabido que tienen rasgos de marcadores discursivos fijos con funciones de conectores, como la reformulación y/o ejemplificación (véase, por ejemplo, Cuenca 2003, 2017). No obstante, como estos casos se comportan de forma diferente a las formas flexionadas de *decir*/*dir* y muestran un grado más avanzado de gramaticalización, dejamos el estudio de estas formas y sus diferentres funciones pragmáticas para investigaciones futuras.

sujeto mayor en comparación con las formas perfectivas de pasado, lo que se ha atribuido al sincretismo de la 1ª y 3ª persona del singular (véase, por ejemplo, Hochberg 1986 para el español del Caribe) o a las diferentes funciones discursivas de los tiempos verbales y la focalización en el evento o en el sujeto (véase Silva-Corvalán 1997, 127). Además, se ha argumentado (véase, por ejemplo, Becker 2021 y referencias allí citadas) que el aspecto imperfectivo favorece una perspectiva narrativa interna, lo que podría afectar la frecuencia de expresión del sujeto de 1ª persona singular.

En cuanto al tipo de complemento, ya se ha mencionado que *decir* puede utilizarse como partícula citativa cuando introduce complementos de estilo directo. Como este uso del verbo de comunicación tiene un grado de fijación, podría tener consecuencias para la expresión del sujeto, así que se deriva la necesidad de examinar los sujetos implícitos y explícitos de *decir*/*dir* con diferentes tipos de complemento oracionales (directos e indirectos) y no oracionales (sintagmas nominales, pronombres demostrativos y relativos, etc.). La investigación de la omisión o presencia del argumento dativo está relacionada con el punto anterior: como partícula discursiva, se esperaría la omisión frecuente del argumento dativo, así que es relevante controlar este factor. Por último, examinamos si la expresión del sujeto con *decir*/*dir* está afectada por el tipo de oración en el que aparece (oración principal vs. subordinada; declarativa vs. interrogativa, etc.). Se podría esperar que la expresión fuese más frecuente en oraciones principales, ya que los sujetos de las subordinadas son a menudo correferenciales con los de las principales, o al menos se refieren a referentes ya mencionados en el discurso.

En la siguiente sección, presentamos los resultados de forma cuantitativa descriptiva. Después, compararemos los resultados de los corpus del castellano y catalán y discutiremos algunas observaciones de forma cualitativa.

4. Resultados

4.1. Persona y número gramatical

Si consideramos la persona y el número gramatical, llama la atención que la primera persona del singular no se relacione con una expresión de sujeto más elevada que la 2ª o 3ª. En los datos del castellano (*PRESEEA*), la 1ª y 2ª persona del singular tienen frecuencias parecidas de sujetos explícitos (20 % y 15 %, respectivamente). En la 3ª persona del singular, observamos una expresión de sujeto más elevada de 37 %, pero este dato se debe a la inclusión de los sintagmas nominales plenos en el análisis (véase Tabla 2), ya que la 1ª y 2ª persona singular solamente permiten una alternancia entre pronombre implícitos y pronombres fuertes.[3]

[3] Si se excluyen los sujetos léxicos plenos y nombres propios del análisis de la tercera persona, la frecuencia de pronombres explícitos baja a un 7 % (singular) y un 4 % (plural) en los datos del español.

Tabla 2: Frecuencia de pronombres implícitos y explícitos con decir *en los datos del PRESEEA*

	implícito	explícito	total	%-explícito
1sg	632	158	790	20 %
2sg	148	26	174	15 %
3sg	277	162 (140 SSNN + 22 pronombres)	439	37 %
1/2pl	94	0	94	0 %
3pl	182	46 (38 SSNN + 8 pronombres)	228	20 %
Total	1333	392	1725	23 %

En términos de frecuencia de sujetos explícitos, el verbo *decir* se comporta de forma diferente al verbo de cognición *creer* que tiene una frecuencia de sujetos explícitos de 1ª persona singular más elevada en castellano (véanse, por ejemplo, Enríquez 1984, Bentivoglio 1987, Morales 1997, Hurtado 2005, Posio 2011, 2013).

En los datos del catalán central (*COC*) se da una situación comparable a la del castellano con la diferencia de que la frecuencia de los sujetos expresados de la 3ª persona no es más alta que en la 1ª y 2ª (23 % y 25 %), incluso si se incluyen los sujetos léxicos.[4] La tabla 3 muestra las frecuencias de los sujetos implícitos y explícitos con *dir* en los datos del *COC*.

Tabla 3: Frecuencia de pronombres implícitos y explícitos con dir (excluyendo los sujetos léxicos) en los datos del COC

	implícito	explícito	total	%-explícito
1sg	105	32	137	23 %
2sg	38	13	51	25 %
3sg	245	79 (64 SSNN + 15 pronombres)	324	24 %
1/2pl	32	2	34	6 %
3pl	45	8 (7 SSNN + 1 pronombre)	53	15 %
total	465	134	599	22 %

La diferencia de la frecuencia de sujetos explícito léxicos plenos de 3ª persona entre los dos corpus probablemente se debe a los tipos de discurso: en *PRESEEA*, se introducen más referentes nuevos con sujetos léxicos debido a la estructura semidirigida de las entrevistas y las preguntas de los entrevistadores, mientras que en las conversaciones espontáneas, los referentes de la 3ª persona suelen ser más conocidos o pertenecer al terreno común de los hablantes.

[4] Si se excluyen los sujetos léxicos plenos y nombres propios del análisis de la tercera persona, la frecuencia de pronombres explícitos baja a un 6% (singular) y 2% (plural) en los datos del catalán.

En cuanto al número gramatical de las formas verbales, se observa en las dos lenguas que las formas plurales desfavorecen los sujetos explícitos en todas las personas.[5] En general, se observa también que el uso de *decir/dir* es más frecuente en las personas singulares, cosa que se explicará por el uso del verbo para introducir citas que típicamente se atribuyen a una sola persona, con la excepción de la construcción impersonal *dicen* (véase Siewierska y Papastathi 2011).

4.2. Tiempo y aspecto

En cuanto al factor de tiempo y aspecto, hay algunas tendencias paralelas en los datos del castellano y catalán: en las dos lenguas, *decir*/*dir* tiene una frecuencia de expresión de sujeto baja en las formas verbales de subjuntivo y una frecuencia media en las formas verbales del presente de indicativo. Nótese que esta tendencia es evidencia en contra de la relevancia del sincretismo *per se* entre la 1ª y 3ª persona gramatical para la expresión del sujeto: las formas de subjuntivo son sincréticas en castellano y catalán, igual que las del pasado imperfectivo. No obstante, no tienen una frecuencia de expresión del sujeto más elevada en comparación con las formas no sincréticas del presente de indicativo. Por otro lado, las formas del condicional, que también presentan el sincretismo entre la 1ª y la 3ª persona, sí se asocian con alta presencia del sujeto expresado en las dos lenguas (véase el Gráfico 3). Así pues, se puede concluir que la presencia de un sincretismo en sí no favorece la expresión del sujeto, sino que se trata de las funciones pragmáticas de las formas verbales y también la probabilidad de que el referente del sujeto sea información dada en el discurso: esta probabilidad sería más alta en las formas del subjuntivo que rara vez se usan para introducir referentes nuevos al discurso.[6]

[5] Como menciona un revisor, es muy interesante que en la primera y segunda persona del plural se den omisiones casi totales aunque existe la posibilidad de sujetos léxicos plenos a diferencia de la 1ª y 2ª persona del singular.

[6] Surge la pregunta por qué no existe el mismo efecto del subjuntivo en el condicional, ya que ambos introducen formas de irrealis. Una posible solución sería que el subjuntivo se encuentra más fácilmente en oraciones subordinadas, como observa un revisor anónimo. Otra posibilidad sería que las formas condicionales del verbo *decir*/*dir* a menudo introducen pensamientos u opiniones más subjetivas, parecido a *yo creo/pienso que*:

(i) MMM (... 1.44) *vols dir que no l'he sentit i t'he do*[*nat un co:p/* 3]
[...]
PPP *no ho cr*[*ec*\ *no*
PPP *no*
[...]
MMM [*ah*\]
MMM ***jo diria que sí***\
PPP (.. 0.41) *no t'he pas sentit jo a tu*\ (COC07_líneas 502-510)

(ii) I: *no **yo diría que** no que es un barrio seguro ¿no?*
(PRESEEA Madrid_H22-026_166-1)

No obstante, las formas condicionales del verbo *decir*/*dir* son poco frecuentes en los corpus examinados, así que se deberían comprobar estas tendencias con una base de datos más amplia.

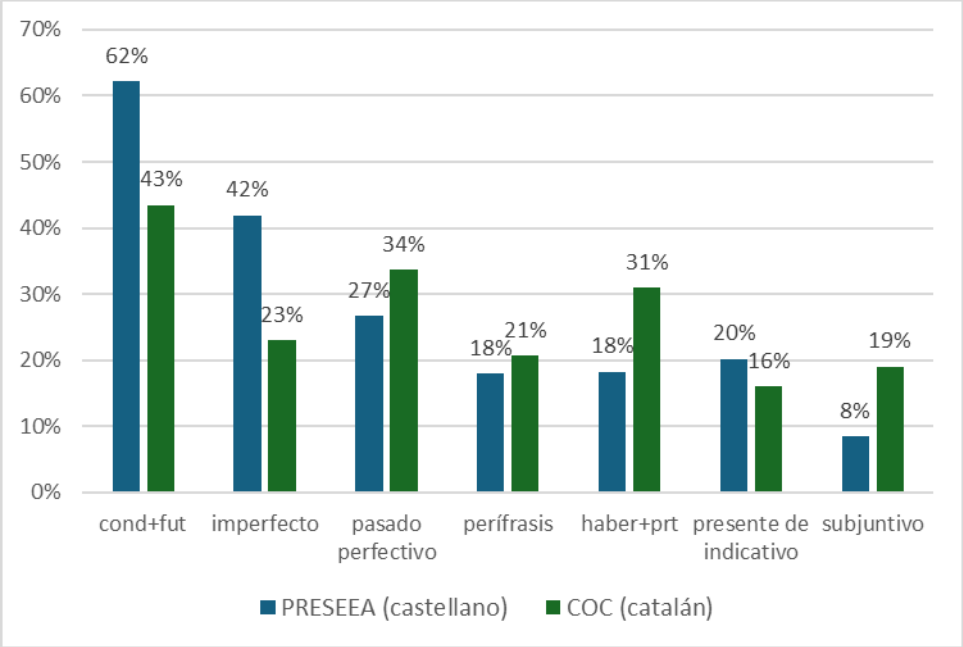


Gráfico 3: Sujetos explícitos (pronombres y SSNN plenos) y tiempo/aspecto en los datos del castellano (PRESEEA) y catalán (COC)

En las formas del pasado es donde surgen diferencias interesantes entre los datos del castellano y catalán: en castellano, las formas imperfectivas tienen una frecuencia alta de sujetos explícitos en comparación con las formas perfectivas del pretérito indefinido. En catalán, por el contrario, son las formas de pasado perfectivo (con el auxiliar *anar* 'ir') las que tienen una frecuencia elevada de sujetos explícitos. Parece ser relevante que las formas de pasado de *decir* en los datos del *PRESEEA* y *dir* en el *COC* se usan en contextos diferentes: mientras en el *PRESEEA* el pasado imperfectivo de *decir* aparece con frecuencia en narraciones de pasado para introducir citas directas (en total 81 casos de imperfecto y 111 de indefinido), como en los ejemplos (2–4), este uso es marginal en los datos del *COC* (solamente 7 casos del imperfecto, comparado con 29 casos de pasado perfectivo).

(2) *o sea fueron a verme todos mis amigos menos él* ***y yo ya decía*** <cita> *qué cosa más rara ¿no? que no venga a verme* </cita> (*PRESEEA*, Alcalá, M12-023)

(3) *buah allí en la pastelería que no tenemos calefacción / con la rebeca el jersey de cuello vuelto debajo* ***y yo decía*** <cita> *por Dios nos falta aquí el esquimal* </cita> / <entre_risas> (*PRESEEA*, Madrid, M11-004)

(4) *no me acuerdo vamos* <ruido = "chasquido boca"/> *no* ***te puedo decir yo*** *pues / yo decía / y* ***digo*** <cita> *pues esto sucedió* <ininteligible/> (*PRESEEA*, Alcalá, M33-018)

En (2), la forma imperfectiva *decía* introduce un complemento oracional de estilo directo. En este caso, no se trata de un evento habitual ni ilimitado. No obstante, se expresa la cita desde una perspectiva interna en las narraciones del pasado. Se ha discutido en diferentes estudios la relación entre una perspectiva narrativa interna y el uso del aspecto imperfectivo en las lenguas románicas (véase, por ejemplo, Giorgi 2010, Becker 2021). En otros estudios (véase, por ejemplo, Silva Corvalán 2001), se ha argumentado que el aspecto imperfectivo coloca el evento en el trasfondo (en el

presente caso, el evento de *decir*/*dir*), lo que permite poner más énfasis en el sujeto. En (2), adicionalmente se da un cambio referencial del referente de tercera persona *él* en la perspectiva del hablante, lo que favorece el uso del pronombre explícito. En los ejemplos (3) y (4), al contrario, no se trata de un cambio de referente, así que parece ser la subjetividad el factor primordial en estos casos.

En los datos del catalán central del *COC*, son las formas del pasado perfectivo las que se asocian a una frecuencia elevada de expresión de sujeto. En esta lengua, como es bien sabido, es la forma perifrástica *anar* + infinitivo la que se usa para expresar el pasado perfectivo en la lengua oral de la mayoría de las variedades (excepto el valenciano central, algunas zonas del mallorquín, ibicenco y de Elx; véase Perea 2007; Beltran & Segura 2019). En los datos del *COC*, la alta frecuencia de sujetos explícitos en pasado perfectivo parece estar relacionada con el hecho de que esta forma temporal se usa con cierta frecuencia para introducir referentes nuevos al discurso con el verbo *dir*. Así, de los 30 sujetos explícitos con *dir* en pasado perfectivo, 22 (= 73 %) son sintagmas nominales plenos o nombres propios:[7]

(5) *que [quan el] seu fill **va dir**_ bueno\ ara no ens trauran d'aquí quatre dies oi/* (*COC* 08, 1099-1100)
‘que cuando su hijo dijo “bueno…ahora no nos sacaran de aquí durante cuatro días, ¿no?’

(6) *saps què ens va dir ahir **la meva cunyada**/* (*COC* 09, 123)
‘¿Sabes lo que nos dijo ayer mi cuñada?’

Por lo tanto, en cuanto a la expresión de sujeto con las formas de pasado, las diferencias entre los datos del castellano y catalán parecen explicarse por la influencia que tiene el tipo y contexto discursivo y no necesariamente por diferencias interlingüísticas.

4.3. Tipo de complemento y construcciones con el verbo decir/dir

La siguiente tabla muestra la frecuencia de la expresión del sujeto a nivel global en relación con el factor del tipo de complemento de *decir*/*dir* en los datos del castellano y catalán:

Tabla 4: Frecuencia de sujetos explícitos con decir/dir según el tipo de complemento

	PRESEEA (castellano)			*COC* (catalán)[8]		
tipo de complemento	sujeto explícito	no. total	%-explícito	sujeto explícito	no. total	%-explícito
estilo directo	164	749	22 %	42	255	16 %
estilo indirecto	93	381	24 %	39	140	28 %
sin complemento	53	314	17 %	0	16	0 %
clítico acusativo	15	78	19 %	17	67	25 %
SN/PRN/SP	14	68	21 %	23	58	40 %
pronombre relativo	27	80	34 %	5	15	33 %
pronombre interrogativo	26	55	47 %	7	30	23 %
total	392	1725	23 %	133	581	23 %

[7] De los 8 sujetos pronominales explícitos, 5 son pronombres fuertes o demonstrativos de 3ª persona.
[8] Se han excluido 18 casos de anotación dudosa.

Si consideramos los datos del catalán, la frecuencia de expresión de sujeto es más baja cuando el complemento es una oración de estilo directo en comparación con los complementos oracionales de estilo indirecto (introducidos por *que*) – 16 % vs. 28 %, respectivamente. Esto se debe a que *dir* se utiliza a menudo como partícula citativa en los datos del catalán del *COC* y en este uso *dir* se asemeja a un marcador discursivo (véase Benavent 2003; Posio & Peskova 2020 para una discusión del castellano), aunque el proceso de gramaticalización o pragmaticalización no esté tan avanzado como en el caso de los marcadores deverbal prototípicos como *venga*, *vale*, *vaya*, etc.

El uso del *verbum dicendi* como partícula citativa es bien documentado en castellano oral, pero no tenemos constancia de descripciones detalladas del catalán oral. En los datos del *COC*, observamos que las formas verbales de *dir* con sujeto nulo y en combinación con un complemento de una cita directa tiene al menos las siguientes funciones discursivas interrelacionadas:

A) Introducción de una cita:
El uso de *dir* para marcar una oración como cita directa es evidente en catalán oral cuando aparece en presente de indicativo por defecto y coincide con la presencia de un *verbum dicendi* con especificación temporal de pasado en la misma oración (cf. Benavent (2003) y Posio & Pešková (2020) para el castellano). Véanse los ejemplos (7) y (8).

(7) *I jo li* ***he dit_ dic***_ *ai* […] (*COC* 07, líneas 99-102 [adaptado])
‘y yo le he dicho, digo: “ay […]”’

(8) *em* ***va dir diu*** *juga bé* (*COC* 08, líneas 576-578 [adaptado])
‘Me dijo, dice “juega bien”’

En estos casos, la carga semántica del verbo de comunicación la lleva la primera mención del verbo flexionado (*he dit* y *va dir* ‘dijo’) mientras que la segunda mención tendría la mera función de marcar el complemento oracional como cita. En este último uso, *dir* aparece sin sujeto explícito. Evidencia adicional para este razonamiento se encuentra en ejemplos en los que el verbo principal tiene una semántica diferente de la de *dir*:

(9) FAN: *Jo l’altre dia* ***pen[sava_ dic***\]
[HIL: […]]
FAN: ***dic*** *en lloc de gastà'ns-e cent mil peles pel* {(L2) *traje*} *de nu- de nuvis_*
[…] (*COC* 02, líneas 398-402 [adaptado])
‘Hablante FAN: Yo, el otro día, **pensaba, digo** [Hablante HIL: […]] FAN: digo: “en vez de gastarnos cien mil pesetas por el traje de novios […]’

En (9), el verbo *pensar* lleva la carga semántica de la oración principal y *dic* marca el complemento oracional como estilo directo que introduce pensamientos.

B) Señalar la continuidad de una cita:
Aparte de introducir una cita, *dir* aparece frecuentemente con la función de señalar la continuidad del turno de una cita:

(10) NAA:***diu*** *ara*\
diu_
*ara que em començo a imaginar*_
diu\
*que seré jo*_
diu\
{(F) *lo r:idícul*_
*i inútil que em sentiré quan entri per la porta i tothom m'aplaudeixi*_}
(*COC* 02, líneas 930-937 [adaptado])
Hablante NAA: 'Dice "ahora" dice "ahora que empiezo a imaginar" dice "que seré yo" dice "lo ridículo que me sentiré cuando entre por la puerta y todo el mundo me aplauda"'

En la primera línea en el ejemplo (10), el primer *diu* 'dice' introduce la cita directa. En las líneas siguientes, la repetición de *diu* tiene la función de señalar la continuación de la cita a la vez que se mantiene el turno del hablante que está reproduciendo la cita.

C) Cambiar el turno de una cita:
Aparte de introducir y continuar una cita, *dir* puede indicar el cambio del emisor de una cita directa. Lo interesante en este caso, y lo relevante para el estudio de la expresión de sujeto, es que el cambio de referente no causa el uso de un sujeto explícito, aunque la cadena referencial es discontinua:

(11) RRR:***diu***_
p(e)rò tothom diu que bueno\
que fa uns passes fantàstics\
dic *ah sí:/* → cambio de hablante en el discurso referido
diu_ → cambio de hablante en el discurso referido
no no\
diu_ → continuidad de la cita del mismo hablante
jo no vull pas que es dediqui an això\
(*COC* 08, líneas 582-589 [adaptado])
Hablante RRR: '**Dice** pero todo el mundo dice que bueno que hace unos pases fantásticos **digo** "ah sí" **dice** "no no" **dice** "yo no quiero que se dedique a eso"'

El cambio de turno como factor para la expresión del sujeto ha sido el objeto de investigación de muchos estudios previos que postulan que favorece el uso de un sujeto explícito en español (véase Cameron 1992). No obstante, el cambio de turno entre citas directas parece comportarse de forma diferente a los casos de cambio de turnos canónicos en catalán. Lo mismo se puede observar en varios ejemplos del castellano:

(12) *y* ***yo*** *le* ***dije*** *"es que yo no tengo un duro" "bueno"* ***dice***<alargamiento/> *"es que / yo no sé" "bueno" y siempre* ***me decía*** *"bueno"* (*PRESEEA*; Alcalá H33-051)

En estos casos, parece ser suficiente que los dos referentes que emiten la cita formen parte del terreno común (*common ground*) para legitimar un sujeto nulo.

En todos estos casos, *decir* y *dir* parecen haber alcanzado un grado de pragmaticalización desde un verbo pleno hacia un marcador citativo. Lo interesante es que este proceso de gramaticalización parece haber afectado a la forma verbal junto con

su sujeto nulo, a diferencia de otras formas rutinizadas como *yo creo que* en español, que tiene un grado de fijación junto con su sujeto explícito (véase Posio 2011; Herbeck 2021).

La hipótesis de que la combinación de *dir* y su sujeto nulo es una forma semifijada en estos casos recibe evidencia adicional de la semántica reducida que tiene cuando introduce citas directas. En estos casos, *decir*/*dir* puede aparecer en contextos de complementos declarativos (equivalente a *decir*), interrogativos (equivalente a *preguntar*) y puede introducir incluso pensamientos (equivalente a *pensar*; véase (9) arriba), lo que indica que en vez de expresar la semántica de *dir*, tiene la mera función de marcar estilo directo (de algo dicho, pensado o interrogado):

(13) ***Jo li vaig dir_ dic*** *com és que no te n'has comprat* (*COC* 09, 2198-2199)
'Yo le dije, digo "¿Cómo es que no te has comprado una?"

En los datos del castellano (*PRESEEA*), estos usos de *decir* se documentan también, pero existen además usos más subjetivos en las narraciones del pasado (véanse los ejemplos (2)–(4)), lo que explica la frecuencia más alta de sujetos explícitos en los datos de *PRESEEA* en comparación con los del *COC*. Por lo tanto, más que a una diferencia interlingüística, los datos apuntan a que el tipo discursivo, el uso de ciertas formas construccionalizadas y el grado de subjetividad son decisivas para la frecuencia de la expresión del sujeto con el verbo *decir*/*dir* (véase también Enríquez 1984; Aijón Oliva & Serrano 2010, Posio 2011, Herbeck 2021 para los verbos *creer* y *saber*).

Alguna diferencia en secuencias formulaicas entre *decir* en castellano (*PRESEEA*) y *dir* en catalán (*COC*) se puede observar en la construcción con *como*. En castellano, aparece con frecuencia la secuencia <*como* DECIR (sujeto)> – en total, 63 casos. En esta secuencia, predominan los sujetos explícitos (44/63 = 70 %) y la gran mayoría de ellos aparece en posición postverbal (33/44 = 75 %; véase también Herbeck & Posio 2022):

(14) *ni madrugo ni trasnocho* ***como digo yo*** (*PRESEEA*; Alcalá; H12-043)

(15) *aunque también / yo conozco a gente / conozco a gente que bueno / eeh a lo mejor son* ***como digo yo*** <cita> *individuos* </cita> (*PRESEEA*; Alcalá_H12-019)

(16) *pero yo no me enteré de la mitad de las cosas se me pasó el día*<alargamiento/> / *¡fum*<alargamiento/>*! volado / llegó* ***como digo yo*** *de pronto / las nueve de la noche ya*<alargamiento/> *ha* <vacilación/> / *ha terminado el baile* (*PRESEEA*; Alcalá_H11-037)

En algunos casos, se podría argumentar que <*como* DECIR (sujeto)> introduce una cita directa (véase (15), donde la secuencia *individuos* se ha codificado como cita en el corpus). No obstante, la comparación con otros ejemplos revela que *como digo yo* no suele introducir una simple secuencia de discurso directo reproducido, sino más bien indica que se trata de una expresión que el hablante considera típica para sí mismo o un uso irónico o marcado de una palabra o concepto común. Independientemente de este análisis, la secuencia rutinizada aparece en la mayoría de los casos con un sujeto explícito en castellano. En los datos del catalán, al contrario, no hemos podido documentar esta construcción. Es posible que su frecuencia en los datos del castellano

se deba al género discursivo, ya que las entrevistas pueden favorecer el uso de este tipo de fórmulas en comparación con conversaciones coloquiales.

En cuanto a los complementos de pronombre interrogativo, *decir* tiene una frecuencia de expresión de sujeto elevada en castellano (47 %) pero no en catalán (23 %). De nuevo, parece influir el tipo discursivo y la situación comunicativa: en el caso de los datos del español de *PRESEEA*, hay turnos entre un/a entrevistador/a y un/a informante. En los turnos de preguntas, se introduce con frecuencia un tema nuevo o se contrastan perspectivas sobre un tema ya dado (véase (17)–(19)). Es justo en estos contextos en los que la expresión del sujeto es frecuente:

(17) *claro / pues **yo** creo que ya hemos terminado ¿qué dices **tú**?* (*PRESEEA*; Alcalá, H32-033)

(18) *¿pero **tú** qué dirías?* (*PRESEEA*; Madrid, M32-047)

(19) *¿y **la gente** qué decía de / eso?* (*PRESEEA*; Madrid, H21-020)

Además, encontramos algunas frases rutinizadas a nivel idiolectal, como por ejemplo la secuencia *qué te diré yo* que se puede considerar como un recurso de atenuación ya que expresa "fingimiento de incertidumbre, de incompetencia o ignorancia" (Cestero y Albelda 2022, 98–99):

(20) a. *y me gusta así* <ininteligible/> */ me gusta/ mmm **¿qué te diré yo?** / tengo*<alargamiento/> */ la Pardo Bazán* […]
b. *me* <vacilación/> *mmm me gusta / mmm **qué te diré yo** / eeh bueno / uno que ahora no se lee que es Ricardo León* / […]
c. *Unamuno me gusta algo / **qué diré yo** / mmm me* <vacilación/> *me gusta Zorrilla las nov*<palabra_cortada/> *las*<alargamiento/> <vacilación/> / *las Leyendas de Zorrilla y es oct*<palabra_cortada/> *y un verso menor /* (*PRESEEA*; Alcalá, M33-018)

4.4. Otros factores: Tipo de oración, presencia y ausencia del dativo

En el análisis del tipo de oración, hemos distinguido entre oraciones principales declarativas, interrogativas, coordinadas y oraciones subordinadas. En cuanto a las interrogativas, los resultados confirman las observaciones de la sección anterior: se relacionan con una frecuencia de expresión de sujeto elevada en los datos del castellano (41 %), pero no en los del catalán (23 %). Lo interesante es que, tanto en castellano como en catalán, la frecuencia de los sujetos explícitos es elevada en las oraciones subordinadas (34 % en los datos del castellano y 28 % en el catalán). Al contrario, en las oraciones principales declarativas y coordinadas no se observa una frecuencia alta de sujetos explícitos en ninguna de las lenguas.

Una razón para estas tendencias podría encontrarse en la estructura informativa. Es bien sabido que los sujetos que representan tópicos continuos o información dada tienden a omitirse, mientras que los que representan información nueva o contrastiva tienden a expresarse explícitamente. Ahora bien, el contenido de la oración subordinada es, en muchos casos, información remática (véase NGLE 44.10 para las oraciones relativas). Además, hemos clasificado muchas construcciones hendidas como oraciones subordinadas. En ellas, la oración a menudo expresa información focal o remática (véase Gutiérrez Bravo 2021 para un análisis de las construcciones con *lo que* en castellano):

(21) *és lo que]* ***diu el Francesc***\ (*COC* 02, 1431)
'Es lo que dice Francesc'

Es interesante mencionar también que las oraciones introducidas por la conjunción *y*/*i*, clasificadas como coordinadas, no tienen una frecuencia de expresión de sujeto más baja que las oraciones principales declarativas no coordinadas, ni en castellano ni en catalán. *A priori* se podría pensar que la coordinación favorece una continuidad discursiva y, por tanto, se esperaría una tendencia a la omisión del sujeto. En los datos de los dos corpus, no obstante, esta conjunción en combinación con un verbo de comunicación introduce con frecuencia temas y turnos nuevos o cambiantes:

(22) […] *su hermano siguió* ***y nosotros*** *veníamos por detrás o lo que fuera /* ***y el chico*** *se volvió para ver /* ***y el taxista*** */ que lo vio en una plazuela en una parada de taxis allí* ***le dijo*** <cita> *chaval ¿qué buscas?* </cita> (*PRESEEA*; Alcalá H33-051)

(23) *saqué buena nota /* ***pero me dijo*** */ que me podía haber extendido un poco más /* ***y entonces yo le dije*** <cita> *pero bueno ¿no se trata de un resumen?* […] (*PRESEEA*; Alcalá M21-047)

El último factor que examinamos es la ausencia o presencia del argumento dativo. Aquí observamos una diferencia interesante entre los datos del castellano y catalán. Mientras en los primeros no hay diferencia en la expresión del sujeto, la presencia del argumento dativo correlaciona con un uso elevado de sujetos explícitos en los datos del catalán, como se puede observar en el Gráfico 4.

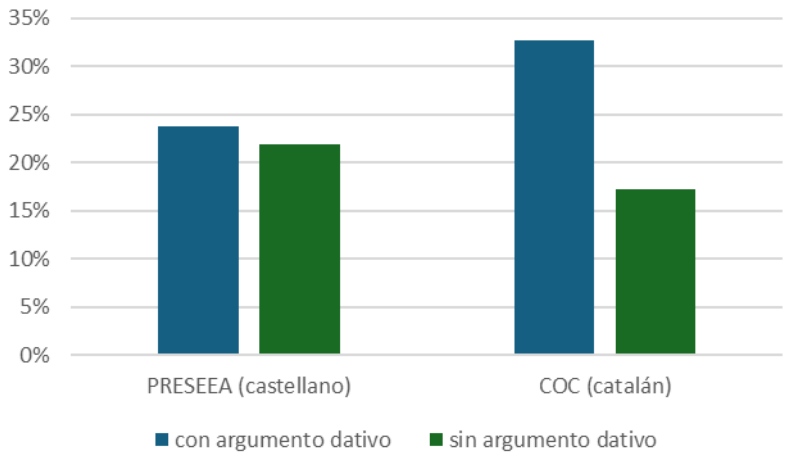


Gráfico 4: Frecuencias relativas de sujetos explícitos y la presencia o ausencia del dativo

En los datos del castellano, la expresión del sujeto es de 24 % y 22 % cuando *decir* va con y sin el argumento dativo, respectivamente. En catalán, al contrario, la frecuencia de los sujetos explícitos aumenta (33 %) cuando *dir* tiene el argumento dativo expreso, en comparación con los usos de *dir* sin dativo (17 % de sujetos explícitos).[9]

[9] Un revisor anónimo propone que sería interesante examinar el papel de las diferencias entre los sistemas de clíticos entre las dos lenguas en este contexto. Es bien sabido que el castellano y el catalán muestran diferencias a la hora de formar combinaciones de clíticos de objeto directo e indirecto – por ejemplo, la

Argumentamos que esta diferencia se debe a la mayor presencia de usos de *dir* como partícula discursiva en los datos del catalán oral. Si el uso del verbo de comunicación como partícula citativa se debe a un proceso de gramaticalización o pragmaticalización, no es sorprendente que tenga una estructura argumental reducida. Evidencia adicional para este razonamiento se encuentra en los datos del castellano, en los que se observa el mismo efecto si consideramos solamente aquellos contextos de *decir* en los que introducen una cita directa (véase la Tabla 4).

	PRESEEA (castellano)				*COC* (catalán)			
	suj. nulo	suj. explícito	total	%-expl.	suj. nulo	suj. explícito	total	%-expl.
con dativo	119	68	187	36 %	33	21	54	39 %
sin dativo	466	96	562	17 %	180	21	201	10 %
total	585	164	749	22 %	213	42	255	17 %

Tabla 4: Sujetos implícitos y explícitos con usos citativos decir/dir con y sin argumento dativo

Por lo tanto, tenemos evidencia adicional de que el uso citativo de *decir*/*dir* favorece sujetos implícitos en los datos del castellano y catalán aquí examinados. En cuanto al catalán, esto podría ser una diferencia importante con los datos de las variedades valencianas investigadas en Herbeck (2022), donde el uso citativo de *dir* y *fer* 'hacer' correlaciona con frecuencias altas de sujetos explícitos.

En la siguiente sección, discutiremos los resultados con más detalle a nivel cualitativo. También ofreceremos algunas consideraciones comparativas con estudios previos sobre otras lenguas (véase Posio 2013, 2014 para el portugués europeo) y variedades (véase Herbeck 2022 para el catalán de Valencia).

5. Discusión

Los resultados indican que la expresión del sujeto con el verbo *decir*/*dir* sigue patrones principalmente parecidos en los datos del castellano y catalán. Las diferencias observadas se pueden derivar de (i) diferencias en el tipo discursivo y la situación comunicativa y (ii) el uso de secuencias formulaicas y gramaticalizadas incluyendo el verbo de comunicación.

5.1. La expresión del sujeto y el tipo discursivo

En cuanto a (i), los datos del *PRESEEA* figuran, entre otros, temas de experiencias personales (pasadas) y opiniones. Por lo tanto, se observa un mayor uso de formas imperfectivas de *decir* en los datos del castellano que en los del catalán. Las narraciones de experiencias personales también explican por qué la frecuencia de sujetos explícitos es mayor en el corpus del castellano *PRESEEA* que en *COC* del catalán. Por otro lado, las formas de *dir* en pasado perfectivo en el corpus del catalán tienen una mayor

sustitución del clítico dativo *le* por *se* en castellano cuando se combina con un clítico acusativo o la sustitución del pronombre dativo *li* por *hi* cuando se combina con clíticos acusativos de 3ª persona en variedades del catalán orientales y nordoccidentales (véase GEIEC 13.5.1), entre otros fenómenos. No obstante, no tenemos datos suficientes de combinaciones de clíticos con *dir*/*decir* en la muestra de datos para ofrecer un estudio contrastivo de un posible efecto para la expresión del sujeto. El hecho de que no hay diferencias de expresión de sujeto cuando miramos solamente la ausencia o presencia del clítico dativo con los complementos oracionales de estilo directo (véase Tabla 4) nos hace pensar que la expresión del dativo en sí no tiene efectos para la expresión de sujeto. No obstante, se necesitan estudios futuros con un número mayor de estructuras de *decri*/*dir* con combinaciones de clíticos para investigar más este punto.

frecuencia de sujetos explícitos que las del indefinido en español. Eso se debe a que en estos contextos, *va dir* frecuentemente introduce referentes nuevos en los discursos correspondientes de los datos del *COC*. Por lo tanto, una conclusión principal del estudio contrastivo del castellano y catalán es que el sincretismo entre la 1ª y 3ª persona del singular no juega el papel más decisivo en la expresión del sujeto en las dos lenguas.

5.2. Secuencias semifijadas y rutinizadas

En cuanto a (ii), los datos indican que el uso de algunas secuencias semifijadas que incluyen sujetos nulos o explícitos influye la frecuencia de expresión del sujeto. Por un lado, algunas formas del verbo *decir*/*dir* se utilizan como partículas citativas en las dos lenguas. De ahí que los sujetos nulos se favorezcan con *decir*/*dir* cuando introduce complementos de estilo directo, pero no son obligatorios. Por lo tanto, una alta frecuencia general de *decir*/*dir* en contextos citativos conlleva una menor frecuencia de sujetos explícitos en los datos. Aunque se podría argumentar que no se trata de una situación de alternancia entre sujetos nulos/explícitos en estos casos y no se deberían computar en el cálculo de las frecuencias de la expresión del sujeto, creemos que sí son relevantes para el estudio de tal fenómeno. De hecho, no podemos hablar de formas completamente gramaticalizadas en un sentido estricto: aunque muestren signos de reducción (semántica y temporal), mantienen propiedades de verbos plenos, como por ejemplo la flexión de persona y número. Por lo tanto, al introducir una cita directa, *decir*/*dir* es *a priori* ambiguo entre un verbo de dicción y una partícula citativa, y los usos citativos se hacen visibles en los casos donde coinciden con otro verbo comunicativo (véanse los ejemplos (1c) y (9)).

Otras secuencias semifijadas se han documentado en el corpus del *PRESEEA*, pero no en el del *COC*: en los datos de *PRESEEA*, aparece con frecuencia la combinación *como*+DECIR(+sujeto), a veces como expresión parentética. Esta combinación incluye sujetos explícitos en posición postverbal con una frecuencia alta. No hemos documentado esta construcción en los datos del catalán, lo que influye en la frecuencia general de sujetos explícitos. Es probable que el uso frecuente de *como*+DECIR(+sujeto) sea característico de las entrevistas semidirigidas más que de las conversaciones espontáneas.

5.3.¿Diferencias interlingüísticas?

Las diferencias observadas entre los verbos *dir* y *decir* del catalán y castellano nos llevan a una discusión de las razones que explican estas diferencias en la expresión del sujeto. Como se trata de lenguas *pro-drop* consistentes por su riqueza de morfología verbal parecida, no se esperarían diferencias interlingüísticas a la hora de investigar las frecuencias de expresión de sujeto en español y catalán. No obstante, estudios anteriores han señalado tendencias opuestas con formas verbales determinadas que tienen una alta frecuencia de uso, por ejemplo, entre el español peninsular y portugués europeo (Posio 2013, 2014): la expresión del pronombre de sujeto de 1ª persona singular es frecuente en castellano con el verbo de cognición *creer*, pero baja en el portugués europeo con el verbo *achar* ‘creer’. Por otro lado, la tasa de expresión del sujeto es baja en español peninsular con *decir,* pero alta con el verbo correspondiente *dizer* en portugués europeo (Posio 2013, 2014). En su estudio sobre el catalán de Valencia, Herbeck (2022) observa algo parecido: *(jo) crec* no tiene frecuencias de expresión más altas que otros verbos (por ejemplo, agentivos), pero el verbo de dicción *dir* ‘decir’ o *fer* ‘hacer’ en la 1ª persona singular sí las tiene. En concreto, *(jo) dic* en 1ª persona singular del presente de indicativo tiene un sujeto explícito en un 57 % de los casos y *fer* ‘hacer’ cuando se

usa como verbo de dicción tiene un sujeto explícito en un 65 % de los casos (véase Herbeck 2022, 149).

Los datos de *dir* en el catalán oral del corpus *COC* no confirman esta observación, sino que se asemejan a los datos de *PRESEEA*. Entonces, surge la pregunta de si estamos, de hecho, observando diferencias dialectales (ya que los hablantes del *COC* son del catalán oriental y los del estudio de Herbeck 2022 del corpus *Parlars* son del catalán occidental), sociolingüísticas y/o diafásicas relacionadas con el tipo de discurso y el tipo de datos.

Aunque no podemos responder a esta pregunta del todo dentro de este estudio, algunas observaciones nos llevan a las siguientes conclusiones tentativas. Por un lado, la subjetividad de la perspectiva de los hablantes parece ser decisiva: la frecuencia alta de sujetos explícitos con *verbum dicendi* se observa en la 1ª persona singular en el estudio de Herbeck (2022). Aparte de *(jo) dic* '(yo) digo', la forma verbal *(jo) fea* '(yo) hacía' se usa con una tasa alta de sujetos explícitos en contextos citativos (véase ibid. 149):[10]

(24) […] *però perquè jo volia eh* ***jo fea*** *"Jo vuic…"* […] (Corpus *Parlars* (Sollana); apud. Herbeck 2022, 146)

Nótese que el verbo *fer* 'hacer' se usa con aspecto imperfectivo en el ejemplo (24) aunque no se trata de un evento habitual o ilimitado, sino más bien de una perspectiva subjetiva del hablante. Lo mismo se observa con algunos ejemplos del presente narrativo (véase (25)):

(25) *i aní al metge i li* ***diguí*** *a a A. /* ***jo dic*** *… me passa açò* […] (Corpus *Parlars* (Catarroja); Herbeck 2022, 148)
'y fui al médico i le dije a a <nombre> yo digo "me pasa eso […]"

Entonces, podría ser la perspectiva interna/subjetiva en las narraciones del pasado que causa el uso frecuente del pronombre de 1ª persona singular explícito con *(jo) dic* o *(jo) fea* (véase Herbeck 2022). Por lo tanto, se podría argumentar que la combinación *jo + DIR/FER* ha alcanzado una cierta rutinización para introducir citas desde una perspectiva interna a la narración.

En los datos del presente estudio del *COC*, al contrario, no encontramos este uso subjetivo de *dir*/*fer* con la misma frecuencia, lo que nos lleva a la conclusión de que la perspectiva que adoptan los hablantes en la narración de eventos es vital para explicar la expresión del sujeto con los verbos de dicción, sobre todo en 1ª persona singular. Hasta qué punto podemos hablar de un proceso de gramaticalización o pragmaticalización en estos casos concretos lo dejamos para investigaciones futuras. Asimismo queda por resolver hasta qué punto se tienen que tener en cuenta propiedades dialectales, diafásicas o más bien sociolingüísticas (los hablantes del corpus *Parlars* son hablantes mayores de 60 años; véase Montserrat & Segura 2020) para explicar las diferencias observadas.

En el portugués europeo hablado, la frecuencia de expresión del pronombre de sujeto con los verbos de comunicación (por su mayoría, el verbo *dizer*) es de un 78 % en los datos analizados por Posio (2013). En un estudio del mismo autor de 2012 donde se analizan todas las ocurrencias de verbos con sujetos posverbales en un corpus oral, son

[10] Para un estudio de *fer* 'hacer' como verbo de dicción en catalán, véase Cuenca & Herbeck (2024) y referencias allí citadas.

también los verbos comunicativos, en primer lugar *dizer* y en segundo lugar *chamar,* los que más frecuentemente ocurren con sujetos posverbales, representando un 49 % de todas las formas verbales analizadas. Con pocas excepciones, la construcción *dizer/chamar* + sujeto posverbal tiene función citativa (Posio 2012, 23–24). Lo que más llama la atención es la alta frecuencia de las secuencias *digo eu assim* y *diz ele/ela assim,* donde además del sujeto posverbal, el adverbio *assim* 'así' forma parte de la construcción citativa (ver el ejemplo (26)):

(26) ***Digo eu assim****: "Ora, vá para o caraças!"*
'Digo yo así: "Mira, vete al carajo".' (CORDIAL-SIN: COV35; Posio 2012, 23)

En el caso del portugués europeo, al menos en los datos analizados por Posio (2012) que constan de entrevistas dialectales con hablantes generalmente mayores, se da un alto grado de rutinización de la construcción citativa con el sujeto posverbal y, opcionalmente, el adverbio *assim.* El autor asocia la colocación posverbal en estos casos con la alta frecuencia de sujetos expresados en el portugués europeo, en comparación con el español peninsular. En una lengua de alta frecuencia de expresión de sujetos, la colocación posverbal del sujeto puede funcionar como una manera de reducir la agentividad del sujeto y darle más prominencia a la cita, mientras que en una lengua *pro-drop* más prototípica, como el español peninsular, el mismo efecto se alcanza con la omisión del sujeto. Así pues, los sujetos posverbales del portugués europeo serían pragmáticamente neutros, mientras que en español peninsular tienden a asociarse con funciones pragmáticas específicas (Posio 2012, 25).

5.4. Consecuencias para un análisis sintáctico

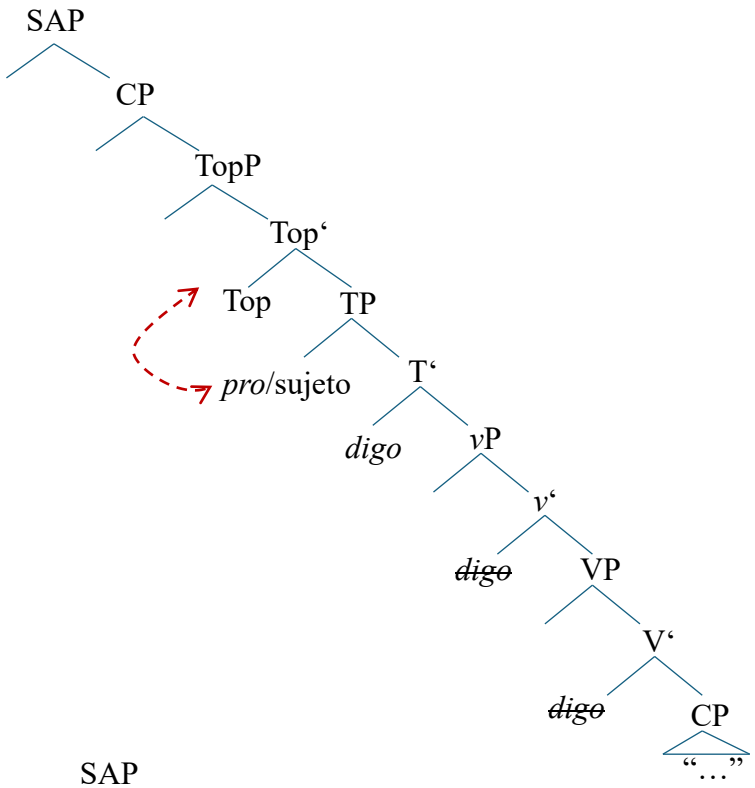
Desde perspectivas funcionales, se han analizado casos como usos rutinizados del tipo *digo*/*dice* en diferentes trabajos como usos formuláicos, secuencias formuláicas o resultados de un proceso de construccionalización (véase el apartado 2). Es decir, una frecuencia alta de uso de ciertas combinaciones lleva a una fijación de la secuencia y una asociación a una función más o menos especializada, sin haber llegado a una fijación o gramaticalización completa. Por ejemplo, *digo* citativo conserva la posibilidad de concordancia de verbo-sujeto y, por lo tanto, tiene una variabilidad parecida a un verbo léxico pero, no obstante, se ven algunos indicios de un proceso de reducción y fijación, como la especificación temporal y semántica reducida en ciertos contextos de uso.

Además, estas construcciones tienden a aparecer con o sin sujeto explícito, dependiendo del verbo y la construcción específica: por ejemplo, *no sé* tiende a aparecer sin sujeto explícito, pero *yo creo* con sujeto explícito en castellano (Posio 2015). Además, la polifuncionalidad de la misma forma verbal puede influir sobre la expresión del sujeto (*creo* con función argumentativa tiende más a aparecer con sujeto explícito que *creo* epistémico [Aijón Oliva & Serrano 2010] *no sé* tiende a aparecer sin sujeto, pero en el caso de *yo qué sé*, el sujeto explícito es obligatorio).

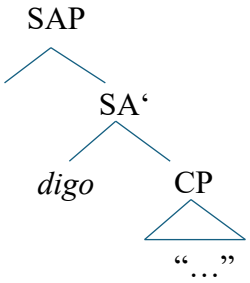
En la sintaxis, surge el problema de cómo codificar estos usos intermedios entre verbo léxico (completo) y elemento discursivo-funcional derivado de un verbo (o sea, marcador discursivo deverbal). En (27) y (28) representamos las estructuras (de forma simplificada) para (i) un verbo léxico y (ii) un marcador discursivo deverbal pragmaticalizado. Adoptamos un marco teórico en el que el interfaz con la pragmática y el discurso, sobre todo relacionado con la perspectiva del hablante y oyente, se codifican en proyecciones funcionales por encima de la proyección máxima de la

oración (véase Speas & Tenny 2003, Giorgi 2010; Haegeman & Hill 2013; Cardinaletti 2015; Cruschina 2015; Wiltschko 2021):[11]

(27)



(28)



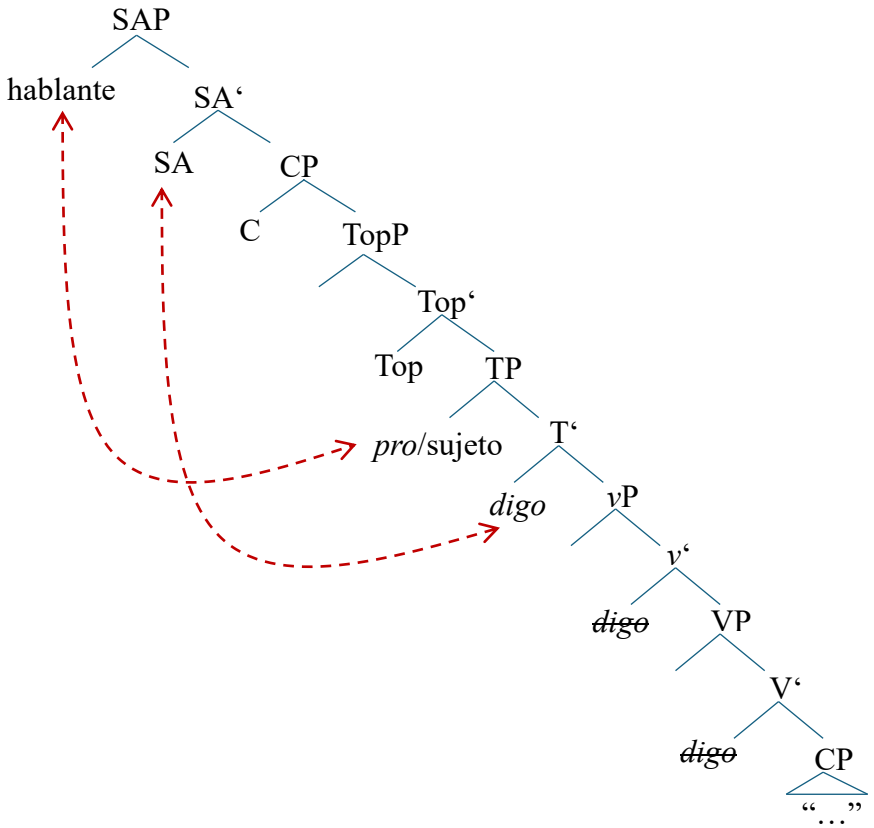
Los marcadores discursivos deverbales reciben un análisis adecuado si asumimos que la periferia de la oración contiene proyecciones funcionales relacionadas con la perspectiva del hablante y oyente y con la interaccionalidad del discurso (por ejemplo, SAP en el análisis de Speas & Tenny 2003). Así, un verbo léxico y su argumento externo se generan en el sintagma verbal y su proyección extendida (S*v*). Un marcador discursivo deverbal (funcional), por el contrario, se puede generar directamente en la proyección discursiva periférica (véase, por ejemplo, Haegeman & Hill 2013; Cardinaletti 2015).

La pregunta es qué estructura sintáctica tienen las expresiones que están en un estadio intermedio entre verbo y marcador discursivo. Se podría argumentar que un paso intermedio entre las estructuras (27) y (28) se encuentra en una relación de *Agree*

[11] Presentamos aquí una versión simplificada de la estructura sintáctica y la posición del sujeto (en Spec,TP). Estudios previos han argumentado a favor de una estructura más detallada y una posición de sujeto diferente a la posición Spec,TP, como por ejemplo, Spec,SubjP (Cardinaletti 2004), Spec,FP (Uriagereka 1995) o Spec,TopP (Ordóñez & Treviño 1999; Frascarelli 2007). El objetivo aquí es solamente representar la relación entre el sujeto oracional (dentro del SComp) y el argumento de hablante de la proyección periférica por encima del SComp – Speech Act Phrase (SAP) en Speas & Tenny (2003).

y movimiento del verbo con la proyección discursiva periférica (véase también Herbeck 2021 para *yo creo (que)*; cf. Demonte & Fernández Soriano 2013 para un análisis de *dizque*). Véase, al respecto, el ejemplo (29).

(29)



Siguiendo este análisis, el desarrollo de un marcador discursivo incorpora procesos parecidos a los de la gramaticalización de los auxiliares – movimiento y pérdida de movimiento (véase Roberts & Roussou 2003) o *Agree* y pérdida de *Agree*. La diferencia reside en la proyección que desencadena el proceso de concordancia y movimiento – una categoría morfosintáctica o una funcional-discursiva. En el caso del marcador de citas *digo*/*dic* o *dice*/*diu* se trata de una relación de concordancia, no necesariamente entre el sujeto y un tópico, sino entre el sujeto/tópico y el argumento de ‘hablante’ de la proyección de SAP (en el sentido de Speas & Tenny 2003).

Una línea de investigación que se podría seguir para explicar la relación de concordancia y movimiento entre las proyecciones periféricas de SAP y el sujeto en Spec,TP es examinar con más detalle la relevancia de la defectividad de la flexión de *dice*/*diu* citativo: en estudios minimalistas del marco teórico de las fases (Chomsky 2000, 2001) se podría argumentar que la defectividad flectiva parcial de *dir* citativo tiene como consecuencia que el SComp es solamente una fase defectiva (y no fuerte), haciendo posible la relación sintáctica de *Agree* entre el argumento de hablante / SAP y el sujeto / T dentro del SComp.

Aunque muchos aspectos de esta línea de investigación quedan por resolver (por ejemplo, la posición exacta del sujeto en las diferentes proyecciones por debajo del SComp - Spec,SubjP (Cardinaletti 2004), Spec,TopP (Ordóñez & Treviño 1999; Frascarelli 2007), Spec,FP (Uriagereka 1995) – y su relevancia para el sujeto en la construcción citativa aquí investigada), creemos que supone una perspectiva fructífera para resolver el proceso de reducción de un verbo con estructura argumental completa hacia un marcador reducido de oraciones de estilo directo.

6. Conclusiones

En el presente artículo, hemos comparado las construcciones con el verbo de dicción *decir/dir* en español y catalán, atendiendo especialmente a la expresión del sujeto del verbo. En términos de frecuencia, observamos tendencias parecidas entre el castellano y catalán: los sujetos se expresan con una frecuencia de un 23 % y un 22 % en castellano y catalán, respectivamente, y si excluimos los SSNN plenos, con una frecuencia de un 14 % y un 12 %, respectivamente. No obstante, observamos algunas diferencias interesantes en cuanto a la expresión en relación con el tiempo y aspecto, con la presencia del argumento dativo y con el tipo de complemento. Hemos argumentado que las diferencias se deben o al tipo de discurso (narrativo vs. no narrativo; entrevista semiestructurada vs. libre) o a la presencia de ciertas construcciones semifijadas, como por ejemplo el uso de *decir*/*dir* como partícula citativa. Estas conclusiones ponen de relieve la relevancia de dichas construcciones para entender la variación en la expresión de sujeto en lenguas *pro-drop*. Esta característica de una lengua puede ser explotada para funciones discursivas específicas llevando a un proceso de pragmaticalización más o menos avanzado.

Comparando los resultados del presente estudio con estudios previos de otras variedades de lenguas románicas, específicamente del valenciano y el portugués europeo, se ha visto que un proceso de pragmaticalización puede tener como base o nulo+verbo, pronombre+verbo o verbo+pronombre. No obstante, para comprobar con seguridad la relevancia de diferencias interlingüísticas para estos procesos, se necesitan estudios futuros con una base de datos que controla el efecto de las variables sociolingüísticas y diafásicas.

Peter Herbeck
Universidad de Viena,
Institut für Romanistik
peter.herbeck@univie.ac.at

Pekka Posio
Universidad de Helsinki,
Departamento de Lenguas
PL 24 / Unioninkatu 40
00014 Helsinki, Finlandia
pekka.posio@helsinki.fi

Referencias

Aijmer, Karin (1997). *I think*: An English modal particle, in Toril Swan and Olaf Jansen Westvik (eds), *Modality in Germanic Languages – Historical and Comparative Perspective.* Berlin: Mouton de Gruyter, pp. 1–47. https://doi.org/10.1515/9783110889932.1

Aijón Oliva, Miguel Ángel & Serrano, María José (2010). El hablante en su discurso: expresión y omisión del sujeto de *creo*. *Oralia* 13, 7–38. https://doi.org/10.25115/oralia.v13i.8100

Becker, Martin (2021). Free indirect discourse, quotative readings and the Romance Imperfect. *Italian Journal of Linguistics* 33(2), 23–58.

Bel, Aurora (2003). The syntax of subjects in the acquisition of Spanish and Catalan. *Probus* 15, 1–26. https://doi.org/10.1515/prbs.2003.003

Beltran, Vicent & Carles Segura (2019). *Els parlars valencians*. València: Publicacions de la Universitat de València.

Benavent Payá, Elisa (2003). ¿Por qué contamos nuestras historias cotidianas en estilo directo? In Nicole Delbecque (ed.), *Aproximaciones cognoscitivo-funcionales al español, Foro Hispánico* 23, 11–20. https://doi.org/10.1163/9789004334403_003

Bentivoglio, Paola (1987). *Los sujetos pronominales de primera persona en el habla de Caracas*. Universidad Central de Venezuela, Consejo de Desarrollo Científico y Humanístico, Caracas.

Brenes Peña, E. (2017). *Lo que yo te diga*: funciones discursivas y proceso de gramaticalización. *Círculo de Lingüística Aplicada a la Comunicación* 71, 63–82. https://doi.org/10.5209/CLAC.57303

Bybee, Joan & Thompson, Sandra A. (1997). Three frequency effects in syntax. *Berkeley Linguistics Society* 23, 378–388. https://doi.org/10.3765/bls.v23i1.1293

Cabré Sans, Yolanda & Anna Gavarró (2007). Subject Distribution and Verb Classes in Child Catalan. En Alyona Belikova, Luisa Meroni & Mari Umeda (eds.), *Proceedings of the 2nd Conference of GALANA*, 51–60. Somerville: Cascadilla Press.

Cameron, Richard (1992). *Pronominal and Null Subject Variation in Spanish: Constraints, Dialects, and Functional Compensation*. Doctoral dissertation, University of Pennsylvania at Philadelphia.

Cardinaletti, Anna (2004). Toward a cartography of subject positions. En Luigi Rizzi (ed), *The Structure of CP and IP*. *The Cartography of Syntactic Structures* (Volume 2). New York: Oxford University Press, pp. 115-165. https://doi.org/10.1093/oso/9780195159486.003.0005

Cardinaletti, Anna (2015). Italian verb-based discourse particles in a comparative perspective, in Josef Bayer, Roland Hinterhölzl & Andreas Trotzke (eds.), *Discourse-oriented Syntax*. Amsterdam: John Benjamins, pp. 71–92. DOI: https://doi.org/10.1075/la.226.04car

Carvalho, Ana M., Rafael Orozco and Naomi Shin (eds.) (2015). *Subject Pronoun Expression in Spanish*. *A Cross-Dialectal Perspective*. Washington D.C: Georgetown University Press.

Casanova, Lourdes (1998). El sujeto en el catalán coloquial. *Revista Española de Lingüística* 29(1), 105–131.

COC = Payrató, Lluís & Núria Alturo (eds.) (2002). *Corpus oral de conversa col·loquial*. *Materials de treball* [Multimèdia]. Barcelona: Publicacions i Edicions de la Universitat de Barcelona. [Consultado: 04/2021]

Cruschina, Silvio (2015). The expression of evidentiality and epistemicity: Cases of grammaticalization in Italian and Sicilian. *Probus* 27, 1–31. https://doi.org/10.1515/probus-2013-0006

Cuartero Sánchez, Juan Manuel (2007). Sobre la naturaleza y el alcance de los factores que influyen en el orden de palabras (con referencia a las oraciones declarativas en español actual). *Lingüística Aspañola Actual* 29, 101–118.

Cuenca, Maria-Josep (2003). Two ways to reformulate. A contrastive analysis of reformulation markers. *Journal of Pragmatics* 35, 1069–1093. https://doi.org/10.1016/S0378-2166(03)00004-3

Cuenca, Maria-Josep (2017). Connectors gramaticals i connectors lèxics en la construcció discursiva del debat parlamentari. *Zeitschrift für Katalanistik* 30, 99–121. https://doi.org/10.46586/ZfK.2017.99-121

Cuenca, Maria Josep & Peter Herbeck (2024). Quan fer és dir. Usos de *fer* com a verb de dicció. *Estudis Romànics* 46, 31–58.

Demonte, Violeta & Olga Fernández-Soriano (2013). Evidentials *dizque* and *que* in Spanish: Grammaticalization, parameters and the (fine) structure of Comp. *Revista de Estudos linguísticos da Universidade do Porto* 8, 211–234.

Detges, Ulrich & Richard Waltereit (2016). Grammaticalization and pragmaticalization, in Susann Fischer & Christoph Gabriel (eds.), *Manual of grammatical interfaces in Romance*. Berlin: de Gruyter, pp. 635–658. https://doi.org/10.1515/9783110311860-024

Díaz, Yarubi & María Labarca (2010). Las formas *digo*, *vamos a decir*, *dicen*, como marcadores discursivos (intensificadores, atenuadores y justificadores) en el habla de Mérida, Venezuela. *Lengua y Habla* 14, 12–24.

Enghels, Renata (2018). Towards a constructional approach to discourse-level phenomena: The case of the Spanish interpersonal epistemic stance construction. *Folia Linguistica* 52(1), 107-138. https://doi.org/10.1515/flin-2018-0002

Enríquez, Emilia V. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid.* Madrid: Instituto Miguel de Cervantes.

Erker, Daniel & Guy, Gregory R. (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3), 526–557. https://doi.org/10.1353/lan.2012.0050

Fernández Soriano, Olga (1993). Sobre el orden de palabras en español. Dicenda. *Cuadernos de filología hispánica* 11, 113–152.

Giorgi, Alessandra (2010). *About the Speaker*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199571895.001.0001

Gutiérrez Bravo, Rodrigo (2021). Las oraciones pseudohendidas en español: Sintaxis y propiedades informativas, in Manuel Leonetti & M. Victoria Escandell-Vidal (eds.), *Estructura informativa*. Madrid: Visor Libros, pp. 359-391

Haegeman, Liliane & Hill, Virginia (2013). The syntacticization of discourse, in Raffaella Folli, Robert Truswell & Christina Sevdali (eds), *Syntax and its Limits*. Oxford: Oxford University Press, pp. 370–390. https://doi.org/10.1093/acprof:oso/9780199683239.003.0018

Herbeck, Peter (2022). Variable First Person Singular Subject Expression in Spoken Valencian Catalan. *Catalan Journal of Linguistics* 21, 129–157. https://doi.org/10.5565/rev/catjl.380

Herbeck, Peter (2021). Perspectival factors and pro-drop: A corpus study of speaker/addressee pronouns with *creer* 'think/believe' and *saber* 'know' in spoken Spanish. *Glossa: a journal of general linguistics* 6(1), 122. doi: https://doi.org/10.16995/glossa.5873

Herbeck, Peter & Pekka Posio (2022). Word order variation with *decir* 'say' in spoken Peninsular Spanish: A case of residual V2? In *Residual Verb Second in Romance*, Silvio Cruschina, Antonio Fábregas & Christine Meklenborg Nilsen (eds.). Special issue of *Isogloss*. *Open Journal of Romance Linguistics*, 8(3)/7, 1–33. DOI: https://doi.org/10.5565/rev/isogloss.196

Hilpert, Martin (2013). *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139004206

Hochberg, Judith G. (1986). Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62(3), 609–621. https://doi.org/10.1353/lan.1986.0041

Holmberg, A. (2005). Is There a Little pro? Evidence from Finnish. *Linguistic Inquiry*

36(4), 533–564. https://doi.org/10.1162/002438905774464322
Holmberg, A., A. Nayudu & M. Sheehan (2009). Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica* 63(1), 59–97. https://doi.org/10.1111/j.1467-9582.2008.01154.x
Hurtado, Luz Marcela (2005). Condicionamientos sintáctico-semánticos de la expresión del sujeto en el español colombiano. *Hispania* 88(2), 335–348. https://doi.org/10.2307/20140954
Martín Zorraquino, María Antonia & José Lázaro Portolés (1999). Los marcadores del discurso, in I. Bosque & V. Demonte (eds.), *Gramática descriptiva de la lengua española*. Madrid: Espasa Calpe, pp. 4051–4214.
Martínez Vázquez, Montserrat (2023). A constructional analysis of *digo yo* in peninsular Spanish, in Inga Hennecke and Evelyn Wiesinger (eds.), *Constructions in Spanish*. Amsterdam: John Benjamins, pp. 255–278. https://doi.org/10.1075/cal.34.10mar
Morales, Amparo (1997). La hipótesis funcional y la aparición de sujeto no nominal: el español de Puerto Rico. *Hispania* 80, 153–165. https://doi.org/10.2307/345995
Montserrat, Sandra & Carles Segura (2020). Un corpus col·loquial i dialectal del valencià: PARLARS. *Zeitschrift für Katalanistik* 33, 9–44. https://doi.org/10.46586/ZfK.2020.9-44
Moreno Fernández (2021). Metodología del "proyecto para el estudio sociolingüístico del español de España y de América" (PRESEEA). Alcalá de Henares: Editorial de la Universidad de Alcalá. [https://preseea.uah.es/documentos-preseea-de-investigacion] https://doi.org/10.37536/PRESEEA.2021.doc1
NGLE = Real Academia Española y Asociación de Academias de la Lengua Española: *Nueva gramática de la lengua española* [en línea], https://www.rae.es/gramática/
Ordoñez, Francisco & Esthela Treviño (1999). Left dislocated subjects and the pro-drop parameter: A case study of Spanish. *Lingua* 107(1), 39–68. https://doi.org/10.1016/S0024-3841(98)00020-5
Orozco, Rafael (2015). Pronominal variation in Colombian Costeño Spanish. In Carvalho et al. (eds.), pp. 17–58.
Otheguy, Ricardo & Ana Celia Zentella (2012). *Spanish in New York: Language contact, dialect, levelling, and structural continuity*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199737406.001.0001
Parlars = Corpus oral del valencià col·loquial. València: Universitat de València. https://www.uv.es/corvalc/
Perea, Maria Pilar (2007). El perfet simple: extensió i ús a començament del segle XXI, in Sadurní Martí et al. (ed.), *Actes del tretzè Col·loqui internacional de llengua i literatura catalanes*, *Universitat de Girona, 9-12 de setembre de 2003*. Barcelona: Publicacions de l'Abadia de Montserrat, pp. 279–296.
de Prada Pérez, Ana (2015). First person singular subject pronoun expression in Spanish in contact with Catalan. En A.M. Carvalho, R. Orozco, N. Lapidus Shin (eds.), pp. 121–142.
Picallo, Carme (1984). The infl node and the null subject parameter. *Linguistic Inquiry* 15(1), 75-102.
Posio, Pekka. (2018). Properties of pronominal subjects, in K. Geeslin (ed.), *The Cambridge Handbook of Spanish Linguistics (Cambridge Handbooks in Language and Linguistics)*. Cambridge: Cambridge University Press, pp. 286–306. https://doi.org/10.1017/9781316779194.014

Posio, Pekka & Andrea Pešková (2020). *Le dije yo*, *digo*: Construccionalización de los introductores cuotativos con el verbo *decir* en español peninsular y argentino. *Spanish in Context* 17, 319–414. https://doi.org/10.1075/sic.18002.pos

Posio, Pekka & Peter Herbeck (eds.) (2023). *Referring to Discourse Participants in Ibero-Romance Languages*. Berlin: Language Science Press [Open Romance Linguistics]. https://doi.org/10.15460/repohh/sub.2024200013

Posio, Pekka (2011). Spanish subject pronoun usage and verb semantics revisited: First and second person singular subject pronouns and focusing of attention in spoken Peninsular Spanish. *Journal of Pragmatics* 43, 777–798. https://doi.org/10.1016/j.pragma.2010.10.012

Posio, Pekka (2012). The functions of postverbal pronominal subjects in spoken Peninsular Spanish and European Portuguese. *Studies in Hispanic and Lusophone Linguistics* 5(1), 149–190. https://doi.org/10.1515/shll-2012-1122

Posio, Pekka (2013). The expression of first-person-singular subjects in spoken Peninsular Spanish and European Portuguese: Semantic roles and formulaic sequences. *Folia Linguistica* 47(1), 253–291. https://doi.org/10.1515/flin.2013.010

Posio, Pekka (2015). Subject pronouns in formulaic sequences: Evidence from Peninsular Spanish, in A. M. Carvalho, R. Orozco & N. Shin (eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal perspective*. Washington: Georgetown University Press.

PRESEEA (2014-). *Corpus del proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. [http://*PRESEEA*.linguas.net].

Roberts, Ian & Anna Roussou (2003). *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486326

Siewierska, A., & Papastathi, M. (2011). Third person plurals in the languages of Europe: Typological and methodological issues. *Linguistics*, 43(2), 575–610.

Silva-Corvalán, Carmen (1997). Variación sintáctica en el discurso oral: problemas metodológicos. En F. Moreno Fernández (ed.), *Trabajos de sociolingüística hispánica.* Universidad de Alcalá, Alcalá de Henares, 115–135.

Silva-Corvalán, Carmen (2001). *Sociolingüística y pragmática del español*. Washington: Georgetown University Press.

Speas, Peggy & Carol Tenny (2003). Configurational properties of point of view roles, in Anna Maria Di Sciullo (ed.), *Asymmetry in grammar* 1. Amsterdam: John Benjamins, pp. 315–344. https://doi.org/10.1075/la.57.15spe

Torrego, Esther. (1984). On inversion in Spanish and some of its effects. *Linguistic Inquiry 15*, 103–129.

Traugott, Elizabeth Closs & Graeme Trousdale (2013). *Constructionalization and Constructional Changes*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679898.001.0001

Trousdale, Graeme (2014). On the relationship between grammaticalization and constructionalization. *Folia Linguistica* 48(2), pp. 557-578. DOI: https://doi.org/10.1515/flin.2014.018

Uriagereka, Juan (1995). An F Position in Western Romance. En K. Kiss (ed.), *Discourse configurational languages*. Oxford: Oxford University Press, 153–175. https://doi.org/10.1093/oso/9780195088335.003.0006

Wiemer, Björn & Walter Bisang (2004). What makes Grammaticalization? An appraisal of its components and its fringes, in W. Bisang, N.P. Himmelmann &

B. Wiemer, *What makes Grammaticalization?* Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197440.1.3

Wiltschko, Martina (2021). *The Grammar of Interactional Langua*ge. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108693707