

Estudio de las variables predictoras de la expresión del sujeto pronominal en el corpus PRESEEA Málaga. Nivel de instrucción bajo

Sociolinguistic Predictors in the Expression of the Pronominal Subject in the PRESEEA Malaga Corpus. Low Educational Level

Antonio Manuel Ávila

Universidad de Málaga

amavila@uma.es

ORCID: 0000-0002-5239-2670

Alba Segura Lores

Universidad de Málaga

albasegura18@gmail.com

ORCID: 0000-0002-0788-1950

Resumen

La ausencia o presencia del pronombre personal sujeto es un fenómeno lingüístico propio de las lenguas *pro-drop*. Este trabajo inicia una serie de estudios que tratan de encontrar los condicionantes lingüísticos y sociales que afectan a la aparición de este pronombre en la variedad del español hablado en Málaga (España). Presentamos resultados parciales referidos a hablantes con nivel de instrucción bajo obtenidos del estudio del corpus lingüístico del Proyecto para el Estudio Sociolingüístico del Español de España y América en Málaga (PRESEEA-Málaga) recogido en la década de los noventa. Para ello, asumimos la metodología empleada en otros estudios que se enmarcan dentro del proyecto panhispánico PRESEEA para así poder comparar los resultados obtenidos. Empleamos la *Guía de codificación de la variable “expresión del sujeto pronominal”* (Bentivoglio, Ortiz y Silva-Corvalán, 2011), aunque la hemos actualizado y adaptado. El análisis exploratorio nos ayuda a comparar a grandes rasgos los porcentajes de presencia del pronombre con estudios previos realizados en distintas zonas geográficas de habla hispana. El análisis de correlaciones nos permite observar qué variables predictoras favorecen su aparición u omisión y apreciar si hay similitudes con los datos de los corpus de Málaga (2015) y otras ciudades como Granada (2016), Sevilla (2019) y Valencia (2021) para, finalmente, avanzar en la propuesta de un posible patrón lingüístico.

PALABRAS CLAVE: pronombre personal sujeto, PRESEEA, Málaga, variables predictoras, sociolingüística.

Abstract

The absence or presence of the subject personal pronoun is a linguistic phenomenon typical of pro-drop languages. This paper initiates a series of studies that try to find the linguistic and social determinants that affect the occurrence of this pronoun in the variety of Spanish spoken in Malaga (Spain). We show partial results concerning speakers with a low educational level obtained from the study of the first linguistic corpus of the Project for the Sociolinguistic Study of Spanish from Spain and American in Malaga (PRESEEA-Malaga) collected in the 1990s. For this purpose, we assume the methodology employed in other studies within the framework of the pan-Hispanic PRESEEA project to compare the results obtained. We use the Coding Guide for the variable “expression of the pronominal subject” (Bentivoglio, Ortiz and Silva-Corvalán, 2011), although we have updated and adapted it. The exploratory analysis helps us to broadly compare the percentages of pronoun presence with previous studies carried out in different Spanish-speaking geographical areas. The correlation analysis allows us to observe which predictor variables favour its presence or omission and to assess whether there are similarities with the data from the corpora of Málaga (2015) and other cities such as Granada (2016), Seville (2019) and Valencia (2021), to finally establish a possible linguistic pattern.

KEYWORDS: subject personal pronoun, PRESEEA, Malaga, predictor variables, sociolinguistics.

FECHA DE RECEPCIÓN: 19/08/2021

FECHA DE ACEPTACIÓN: 24/02/2022



1. Introducción

La gramática universal ha asumido que las lenguas están formadas por elementos explícitos y elementos implícitos. Chomsky (1981, p. 34) denomina *categorías vacías* (*empty categories*) a aquellos elementos que no tienen contenido fonético intrínseco, pero forman parte de la oración; por lo que el contenido fonético del sujeto y su presencia en la oración nos permiten distinguir entre sujetos *expresos* y *tácitos* (NGLE, 2009, p. 638). Al igual que otras lenguas romances, como el italiano, la lengua española admite la alternancia en la formación de oraciones con sujeto explícito o sin él; es decir, el español es una lengua *pro-drop* (Chomsky, 1981). Por lo tanto, aunque nos encontremos con un sujeto omitido, que no está presente en la estructura superficial de la lengua, sí se encuentra en la estructura profunda. Además, los sujetos presentan marcas gramaticales en la oración, y

se pueden distinguir esencialmente tres tipos: la concordancia entre el sujeto y el verbo en número y persona; el caso, que es una característica relativa a la función sintáctica; y la posición sintáctica en la que aparece el sujeto en la oración, que puede ser preverbal o posverbal (NGLE, 2009, p. 638).

Este fenómeno despertó el interés de los investigadores en la década de los ochenta y ha sido estudiado en diferentes zonas geográficas de habla hispana. Con el fin de aunar metodologías y poder comparar resultados, el Proyecto para el Estudio Sociolingüístico del Español de España y de América (PRESEEA) proporciona una *Guía de codificación de la variable “expresión del sujeto pronominal”* (Bentivoglio, Ortiz y Silva-Corvalán, 2011). Gracias a la comparabilidad de los datos, se pretende llegar a establecer determinados patrones lingüísticos que afecten a la omisión del pronombre personal sujeto en español.

El trabajo aquí presentado analiza los datos procedentes de informantes con nivel de estudios bajos como primer paso para obtener una aproximación al estudio de las variables predictoras —tanto lingüísticas como sociales— que propician la ausencia o presencia del pronombre personal sujeto en el habla de la ciudad de Málaga. En esta ocasión, los datos analizados proceden del corpus oral de la ciudad de Málaga recogido en la década de los noventa (Vida, 2007), aunque establecemos comparaciones diacrónicas con la base de datos lingüística recogida en 2015 en la misma ciudad. El motivo del sesgo educativo se justifica porque en nuestro trabajo, además, ofrecemos de modo experimental la actualización de la guía de codificación original incorporando nuevas variables predictoras que han resultado ser significativas en estudios posteriores —como la perífrasis verbal— y otras que se plantean como novedosas en el presente trabajo: *el referente del sujeto, la ambigüedad fonética potencial, la longitud de la forma verbal, el tipo de discurso, el imperativo (modo), orden imperativa afirmativa y orden imperativa negativa (tipo de cláusula), auxiliar de perífrasis (clase semántica del verbo)*. Consideramos que, si la actualización de la guía de codificación resultara adecuada en nuestro estudio parcial, estaríamos en condiciones de, *en primer lugar*, extender la metodología a toda la muestra de hablantes de la ciudad de Málaga, y, *en segundo lugar*, de proponer

la actualización señalada como futura guía en el ámbito general del PRESEEA.

Una vez analizados los resultados, los hemos comparado con otros, procedentes de estudios similares, realizados en otras ciudades como Granada, Sevilla, Valencia y, como hemos señalado, también con corpus posteriores recogidos en la ciudad de Málaga. Aunque los resultados de dicha comparación no pueden considerarse concluyentes, puesto que en otras comunidades de habla se incluyeron hablantes de diversos niveles de instrucción, hemos querido presentar una comparación tanto diatópica como diacrónica, basada esta última en los dos cortes temporales obtenidos en Málaga (1995-2015). El fin de tales comparaciones es la búsqueda de un patrón lingüístico que justifique la presencia o ausencia del pronombre personal sujeto en todos los trabajos considerados, con independencia de diferencias diacrónicas y estratificacionales.

Creemos que, por todo lo anterior, estamos contribuyendo al desarrollo de los estudios contrastivos que se desarrollan en el contexto del Proyecto PRESEEA. Como es lógico, para poder establecer las oportunas comparaciones, hemos asumido los fundamentos metodológicos previos y hemos replicado los análisis que, sin excepción, se han empleado en trabajos anteriores. Sin embargo, este procedimiento excesivamente mecanicista nos ha llevado a entender que, en futuras fases de nuestra investigación, deberíamos explorar nuevos procedimientos de análisis que faciliten la obtención del objetivo final.

2. Marco teórico

La *NGLE* (2009, p. 2548) considera que “existe una relación estrecha entre la capacidad de una lengua para admitir sujetos táticos y la riqueza morfológica de su flexión verbal”. Añade que, en algunas zonas de España —como en las islas Canarias o en el occidente de Andalucía— y en partes del Caribe, se observa una mayor presencia de sujetos expresos que se relaciona con la pérdida del carácter distintivo en la flexión verbal. En la misma línea se expresa Fernández Ramírez (1951, pp. 23-24), quien

considera que los morfemas personales del verbo “componen en español un sistema que puede parangonarse con el de los pronombres personales” y coinciden con ellos “en la dirección del señalamiento tanto en la deixis textual como extratextual”.

La gramática tradicional consideraba al pronombre personal sujeto como enfático, redundante e innecesario. Sin embargo, puede ser necesario en casos donde las desinencias verbales o la referencia a una tercera persona no dejan al sujeto lo suficientemente explícito (*Esbozo*, 1973, p. 421). Además, el sujeto puede hacerse expreso por motivos morfológicos, cuando se encuentran dos formas verbales iguales, como ocurre con la primera y la tercera personas del singular del imperfecto de indicativo o del condicional simple; o por motivos de naturaleza léxica, como el pronombre *usted* que no se interpreta de manera contrastiva.

Como podemos observar, la ausencia o presencia del pronombre personal sujeto no se produce de manera aleatoria, sino que ciertas variables predictoras provocan su aparición o su omisión. Esta variación ha despertado el interés por conocer de qué depende su aparición en la actuación lingüística. Si revisamos estudios anteriores realizados sobre la ausencia o presencia del pronombre personal sujeto, observamos coincidencia al señalar los siguientes parámetros como pertinentes:

- La correferencialidad (Silva-Corvalán, 1982) ha resultado ser la variable predictora estadísticamente más significativa en los estudios realizados sobre el tema. Los pronombres personales tienen relaciones de correferencialidad, debido a que pueden hacer referencia a entidades ya mencionadas en el discurso. Para Lastra y Martín (2015, p. 45): “rest on the idea that it will be more probable to mention the subject when it has had a change in the referent, *i.e.*, in order to establish the new topic”.
- La inespecificidad del referente (Cameron, 1993). Hay ocasiones en las que los sujetos pueden tener una interpretación indefinida (equivalen a alguien no identificado) o una interpretación general (similar a ‘la gente en general’). En el Manual de *NGLE* (2010, p. 644) se afirma que: “la segunda persona del singular puede interpretarse en ciertos contextos con el sen-

tido de ‘uno, cualquiera’, tanto en su forma expresa como en la tácita”. Lo mismo ocurre con las oraciones impersonales de pasiva refleja, donde el sujeto omitido se puede sustituir por ‘todo el mundo’, dándole así una interpretación general.

- La clase semántica del verbo. Tienen mayor influencia aquellos verbos que poseen un carácter subjetivo y que pertenecen al campo semántico de la opinión, del conocimiento o de la percepción intelectual (Enríquez, 1984, p. 118). En los primeros estudios realizados sobre el fenómeno, los verbos se clasificaban según su semántica de la siguiente forma: verbos cognitivos, de percepción, enunciativos, desiderativos y de manipulación, y otros verbos, reservando este último tipo para aquellos que no corresponden con los grupos anteriores y que, por lo tanto, tienen un carácter objetivo (Bentivoglio, 1987). Sin embargo, en estudios posteriores se ha modificado esta clasificación: verbos referentes a los procesos mentales, verbos de estado, verbos *dicendi*, verbos de actividad y otra clase. Gracias a esta última clasificación, se ordenan más detalladamente los verbos de carácter objetivo y subjetivo.
- La ambigüedad morfológica y fonética se produce en aquellos casos en los que la flexión verbal no es capaz de reproducir totalmente los rasgos del sujeto. La ambigüedad morfológica se observa en algunos tiempos verbales como el pretérito imperfecto o el condicional de la primera y la tercera personas del singular, donde la forma verbal es idéntica; por lo que, si el sujeto se presenta como tácito, se produce ambigüedad respecto a la persona que realiza la acción. Mientras que la ambigüedad fonética se documenta solo en algunas zonas donde se produce un debilitamiento o pérdida de las consonantes finales dando lugar a la convergencia de las tres formas del singular (Fernández Soriano, 1999, p. 1236) y la tercera del plural. La hipótesis defendida por Hochberg (1986) en sus estudios sobre el español caribeño propone que la alta frecuencia de aparición del pronombre personal sujeto se debe a la necesidad de compensar la información que se pierde de persona y número al elidir la /-s/ de las desinencias verbales. Sin embargo, estudios posteriores, como los de Ranson (1991) y Cameron (1993), contradicen esta afirmación y consideran que no hay

una correlación entre la elisión final de /-s/ y la presencia del pronombre.

Por lo tanto, como afirma Silva-Corvalán (2003, p. 850), “la expresión llamada ‘opcional’ del sujeto parece válida en un número cada vez más limitado de contextos discursivos”. En el mismo sentido apuntan las palabras de Manjón-Cabeza y otros (2016, p. 185) “la posibilidad de alternancia del pronombre sujeto en español, más que responder a una cuestión estilística, que no modifica el contenido de lo dicho, obedece a condicionamientos léxico-pragmático-morfofonológicos”.

2.1 *Precedentes*

Los estudios que existen sobre la variación del pronombre sujeto en español son muy numerosos y comenzaron en los años ochenta con los estudios de Barrenechea y Alonso (1977) sobre el español de Buenos Aires; Bentivoglio (1980), que se ocupó del español hablado de Venezuela; Silva-Corvalán (1982), que estudió el español hablado mexicano-americano, o Enríquez (1984), que trabajó sobre la lengua hablada en Madrid. El tema ha seguido suscitando el interés de los investigadores, y encontramos trabajos más actuales, como los de Orozco y Guy (2008), Erker y Guy (2012), Shin y Otheguy (2013), Carvalho, Orozco y Shin (2015), Silva-Corvalán (2015) o Peskova (2015). Todos ellos abordan este asunto desde perspectivas diferentes y “con diversos y no siempre adecuados *corpora*” (Manjón-Cabeza y otros, 2016, p. 185).

Los estudios realizados sobre la omisión/presencia del pronombre personal sujeto enmarcados en el Proyecto PRESEEA utilizan un corpus homologable y unos parámetros similares de análisis que permiten establecer comparaciones entre los diferentes resultados obtenidos. En el contexto de este proyecto panhispánico, Manjón-Cabeza y otros (2016) analizan el habla de Granada, Repede (2019) estudia el habla de Sevilla y Roselló (2021) el habla de Valencia. En cuanto al español de América, cabe mencionar, entre otras, las investigaciones de Martínez Lara en Caracas (2013), Lastra y Martín en México D. F. (2015), Guerrero en Santiago de Chile (2019) y Orozco y Hurtado en la ciudad de Medellín (2019).

3. Metodología y corpus

3.1 *Diseño de hipótesis*

En opinión de Silva-Corvalán (2003, p. 850), “La expresión llamada ‘opcional’ del sujeto parece válida en un número cada vez más limitado de contextos discursivos”. A partir de esta afirmación, hemos considerado las siguientes preguntas de investigación iniciales: ¿podríamos corroborar, a tenor de los datos analizados, que la expresión del pronombre personal sujeto depende de variables estilísticas, lingüísticas o sociales, como ha quedado de manifiesto en estudios previos?; ¿qué variables propician su aparición en nuestro corpus?; ¿han cambiado estas variables dentro de la ciudad de Málaga a lo largo del tiempo?; ¿coinciden las variables que afectan al cambio en Málaga con los descritos en otras ciudades?; ¿se puede establecer un patrón que explique el condicionamiento sociolingüístico del fenómeno?

La hipótesis general planteada en nuestra investigación sugiere que los condicionantes lingüístico-sociales que resultaron significativos en la explicación del fenómeno estudiado en investigaciones previas, como las realizadas en la ciudad de Granada (Manjón-Cabeza y otros, 2016), Sevilla (Repede, 2019) o Valencia (Roselló, 2021), afectarán igualmente a los resultados obtenidos al analizar los corpus de la ciudad de Málaga. A partir de esta hipótesis general de corroboración, se proponen otras más específicas:

1. *Hipótesis del peso de las variables predictoras (lingüístico-sociales)*. Las variables lingüísticas serán más influyentes que las sociales en la aparición del pronombre personal sujeto (Martín-Butragueño, 2020).
2. *Hipótesis de la especificidad dialectal*. Las características propias del habla malagueña pueden provocar cierto grado de ambigüedad fonética y, por lo tanto, que el pronombre cumpla una función de desambiguación (Hochberg, 1986).
3. *Hipótesis de la estrategia discursiva*. Es de esperar que las secuencias próximas al cambio de turno conversacional serán más proclives al mantenimiento del sujeto pronominal (Arroyo, 2005).

4. *Hipótesis semántica*. Del mismo modo, los verbos de procesos mentales, de percepción sensorial y de entendimiento serán los más proclives a ir acompañados de sujetos pronominales (Enríquez, 1984).
5. *Hipótesis de la persona verbal*. Igualmente, la primera persona del singular debería ser la que más aparición concentre por el conocido egocentrismo intrínseco a la naturaleza de las lenguas (Givón, 1970).

3.2. *Corpus PRESEEA. Selección de casos, criterios de análisis y procesamiento de los datos*

En general, los trabajos sobre la ausencia y presencia del pronombre personal sujeto se han realizado a partir de la exploración de corpus orales obtenidos gracias a la investigación de comunidades de habla concretas. El presente estudio utiliza los datos contenidos en la base electrónica del corpus PRESEEA en la ciudad de Málaga (Vida, 2007). A estos datos se les añaden, con fines comparativos generales, los obtenidos en la edición parcial de 2015 del corpus de la misma ciudad. Ambos subcorpus están formados por entrevistas orales semidirigidas y respetan las cuotas de afijación uniforme establecidas en PRESEEA que pueden consultarse en su página web (<https://preseea.linguas.net>). Estas cuotas clasifican a los hablantes en diversas subpoblaciones al atender las variables sociales de género, edad y nivel de instrucción y se asigna el mismo número de informantes para cada uno de esos grupos como podemos observar en el cuadro 1. Por lo tanto, respecto al tamaño de la muestra, se propone una asignación uniforme de cuatro informantes por cada casilla (N=72). Ello supone una representatividad de 1/25.000 para una ciudad de casi dos millones de habitantes como Málaga.

Como paso previo al estudio general del fenómeno en la ciudad de Málaga, en este trabajo vamos a analizar únicamente los datos procedentes de los hablantes con nivel de instrucción bajo de las tres generaciones y de los dos géneros considerados, lo que supone una muestra de 24 hablantes para el corpus de 1995 y otros 18 para el corpus —aún en desarrollo— comenzado en 2015; sin embargo, como uno de los informantes del corpus

de 1995 (hombre con nivel de instrucción primario y mayor de 55 años) no realizaba ninguna intervención entre los minutos 10 y 15 de la entrevista, fue eliminado del análisis.

	Generación 1 (20-34 años)		Generación 2 (35-54 años)		Generación 3 (más de 55)	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
Nivel de estudios bajo	4	4	4	4	4	4
Nivel de estudios medio	4	4	4	4	4	4
Nivel de estudios alto	4	4	4	4	4	4

Cuadro 1. Muestra-tipo por cuotas propuesta para el Proyecto PRESEEA

En cada entrevista se ha seleccionado un fragmento representativo de cinco minutos, los comprendidos entre el minuto 10 y el 15. El motivo es que, a partir de ese momento aproximado de la entrevista, el informante está más relajado y se puede acceder de manera más cómoda a su interacción comunicativa más natural, con lo cual el fenómeno estudiado se manifiesta de manera más espontánea. Una vez seleccionados los informantes, se procedió a la anotación de los casos en los que es posible la alternancia estudiada, como muestra el siguiente ejemplo 1:

- (1) Ø era chiquita / <énfasis> y yo no sé cómo </énfasis> pero el abuelo pasó / y yo creía que no Ø m<[e]> iba a ver <risas = E> pero me Ø vio / y <simultáneo> cuando Ø llegué / <cláusula no completa> [MA_M21_105].¹

¹ La matrícula de identificación de los hablantes sigue las pautas de PRESEEA comunes para todos los equipos de investigación. Su interpretación es la siguiente: MA (Málaga), H (Hombre) o M (Mujer), Grupo etario 1 (20 a 34), 2 (35 a 55) y 3 (>55), Nivel de

Se han excluido del recuento algunos casos que pasamos a enumerar:

- Casos que no permiten la alternancia como 2:

(2) **mira** / en la peña antes **se** organizaba una cosa **que** era / cultural / [MA_M31_004].

- Las frases idiomáticas que no presentan variación como: “Yo qué sé, Tú sabes, ...”
- Expresiones deslexicalizadas (muletillas): “Vamos..., Mira...”
- Usos enfáticos con *mismo/misma*: “Yo mismo lo hice”
- Los sujetos focales, tanto en los casos en los que el sujeto es información nueva que responde a una pregunta, como en casos en los que el sujeto es indiscutible por ser foco de contraste.
- Las realizaciones con verbos impersonales como *hacer* y *haber*.
- Cláusulas con *se* impersonal.
- Cláusulas relativas con sujeto relativizado.

3.3 *Guía de codificación*

Para analizar los ejemplos que hemos hallado en nuestras entrevistas, hemos partido de la guía de codificación propuesta por Bentivoglio, Ortiz y Silva-Corvalán (2011). En ella se presenta la codificación de la variable dependiente (ausencia o presencia del sujeto pronominal), las variables independientes de tipo social, como la edad, el género y el nivel de estudios, y 11 variables independientes de tipo lingüístico: persona del sujeto, especificidad del sujeto, modo, tiempo verbal, progresividad, perfectividad, ambigüedad, clase semántica del verbo, tipo de cláusula, correferencialidad y turno de habla.

Como novedad, en nuestro estudio se han añadido nuevas variables predictoras que hemos considerado influyentes en la presencia o ausencia del pronombre personal sujeto, algunas recogidas

instrucción: 1 (nivel bajo), 2 (nivel intermedio), 3 (nivel alto). Los tres últimos dígitos representan el número de identificación del informante.

das en estudios como el de Manjón-Cabeza y otros en la ciudad de Granada (2016) y otras que se proponen aquí de manera original:

- *Referente del sujeto*. Se refiere al tipo de entidad que expresa el pronombre: humano, animado (no humano) e inanimado (cosas y neutros).
- *Ambigüedad fonética potencial*. Se analiza aquí la ambigüedad fonética que se produce por la elisión o pérdida de las consonantes finales de la morfología flexiva que pudiera confluir en la segunda persona del singular, la tercera persona del singular y del plural. Mientras que la guía original solo contempla la ambigüedad producida entre la primera y la tercera personas del singular del imperfecto, condicional y tiempos del subjuntivo, nosotros hemos añadido esta nueva posibilidad, al renombrar la propuesta en la guía original como *ambigüedad morfológica*.
- La variable *modo del verbo* en la guía original se divide únicamente en indicativo y subjuntivo. Sin embargo, en esta investigación hemos encontrado formas verbales en imperativo, por lo que hemos decidido añadir este modo.
- Al tener en cuenta el modo imperativo, en la variable *tipo de cláusula* hemos recogido también las posibilidades *orden imperativa afirmativa* y *orden imperativa negativa*.
- *Perífrasis verbal*. Esta variable ya fue analizada por Manjón-Cabeza y otros (2016), por lo que hemos considerado relevante incorporarla también a nuestro estudio. Las opciones que presenta esta variable son: *perífrasis modales*, *perífrasis aspectuales* y *forma no perifrástica*.
- En cuanto a la *clase semántica del verbo*, en la guía original los verbos se clasifican en mentales, de estado, de *dicendi*, de actividad y otra clase. Al añadir la variable *perífrasis verbal*, hemos considerado pertinente agregar a la clasificación la categoría *verbos auxiliares de perífrasis verbal*.
- Además, hemos tenido en cuenta la pronominalidad de los verbos. Esta variable, ya recogida en el estudio de Manjón-Cabeza y otros (2016), los divide en *verbos pronominales* y *verbos no pronominales*.
- El *tipo de discurso* discrimina entre diálogo rápido, descripción, narración y argumentación. En la guía de codificación de Ben-

tivoglio, Ortiz y Silva-Corvalán (2011) se considera un requisito que aparezcan todos estos tipos de discurso, por lo que hemos decidido contabilizarlo y convertirlo en una variable.

- *Longitud de la forma verbal.* Manjón-Cabeza y otros (2016) realizan un recuento *grosso modo* de las formas verbales perifrásticas y no perifrásticas sin tener en cuenta los fenómenos de la fonética sintáctica. Hemos decidido analizarlos teniendo en cuenta los fenómenos de *shandi*, es decir, los procesos de modificación fónica que sufren los fonemas al unirse entre sí.

En definitiva, nuestra guía de codificación consta de una variable dependiente (ausencia o presencia del pronombre sujeto), dos variables predictoras sociales (edad y género) y 17 de naturaleza puramente lingüística, de las cuales seis son originales o con algunas modificaciones parciales respecto a la guía de codificación original de Bentivoglio, Ortiz y Silva-Corvalán.

Para optimizar el proceso de codificación, ciertas variables se han analizado mediante un programa escrito en *Python*. Para ello, se ha usado la biblioteca de datos *Verbecc*, que recoge las conjugaciones completas de numerosos verbos en español. Se han podido organizar con este programa algunas variables relacionadas con las formas verbales: la persona del sujeto, el tiempo del verbo, el modo verbal, la progresividad, la perfectividad y la ambigüedad morfológica. El programa, como refleja la figura 1, reconoce el verbo conjugado de la casilla D y, según sus características, completa de manera automática las casillas G, K, L, M, N y O que aparecen marcadas con una pantalla. Este procedimiento analítico puede ser considerado también como una novedad en nuestra propuesta respecto a estudios anteriores del fenómeno y supone una simplificación a tener en cuenta en futuros análisis.

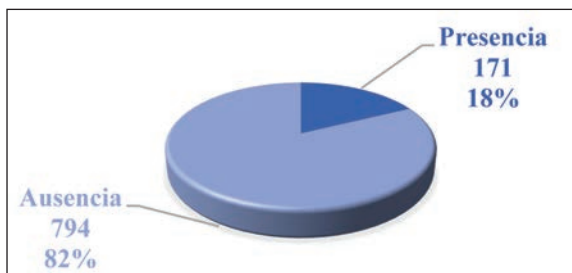
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	INFORMANTE	MCONTEXTO	mas verbal ana	1	infinitivo	Presencia	1	del sujeto	Referen	Ambigüa	Especifick	Modo	ipo verbal	gresividad	fectividad	arfológica	ca del vé.	Tipo de el	Cometer	Turno de	Perifrástis	Pronomi	Tipo de	Longitud
SHE_MA_M03_004	Miyo	<	allic	cantamos	cantar		1	4	1	2	1	0	0	0	0	0	1	3	0	0	0	0	2	3

Figura 1. Ejemplo de codificación en Excel con el programa creado en *Python*

4. Análisis de resultados

4.1 Análisis exploratorio

Se han recogido un total de 965 ejemplos procedentes del corpus oral de la ciudad de Málaga de la década de los noventa (gráfica 1), de los cuales 171 son de sujeto explícito (17.7%) y 794 de sujeto omitido (82.3%). Cabe destacar que el elevado porcentaje de sujetos tácitos se debe a que el español, como hemos dicho, es una lengua *pro-drop* (Chomsky, 1981).



Gráfica 1. Representación en gráfica de la presencia y ausencia del pronombre personal sujeto

Si comparamos los porcentajes que hemos obtenido con los que ofrecen estudios previos, observamos que los resultados varían considerablemente. A partir de la información ya recopilada por Roselló (2021) obtenemos el cuadro 2, donde hemos incorporado nuestros propios resultados. No obstante, en la interpretación del cuadro 2 hay que tener en cuenta que existen diferencias entre las muestras de los estudios comparados. La ciudad de Málaga solo recoge los hablantes de nivel bajo, en ambos corpus (1995/2015), y la de Sevilla los de nivel alto. El resto de los trabajos comprenden los tres niveles de estudios.

4.2 Análisis de correlaciones

Hemos procedido al análisis de los datos obtenidos tras la codificación con el programa estadístico IBM SPSS 20 para comprobar cuáles de las 19 variables independientes resultan pre-

Localidad	Tasa pronominal
Barranquilla, Colombia (Pérez Córdoba y Gomes Camacho, 2019)	45.6%
Cartagena, Colombia (Pérez Córdoba y Gomes Camacho, 2019)	45.3%
San Juan, Puerto Rico (Cameron, 1993)	44.8%
Valledupar, Colombia (Pérez Córdoba y Gomes Camacho, 2019)	41.8%
Barranquilla, Colombia (Orozco y Guy, 2008)	35.7%
Nueva York, EEUU (Otheguy y Zentella, 2012)	34.0%
Málaga, España (Corpus de 2015)	29.0%
Los Ángeles, EEUU (Silva-Corvalán, 1994)	28.9%
Santiago de Chile (Martínez, Guerrero y González, 2020)	25.3%
Sevilla, España (Repede, 2019)	24.8%
New Jersey, EEUU (Flores Ferrán, 2007)	24.0%
Valencia, España (Rosselló, 2021)	21.8%
México DF (Lastra y Martín, 2015)	21.7%
Madrid, España (Cameron, 1993)	20.9%
Málaga, España (Corpus de 1995)	17.7%
Granada, España (Manjón-Cabeza y otros, 2016)	17.5%
Huancayo, Perú (Cerrón-Palomino, 2017)	16.2%

Cuadro 2. Porcentaje de presencia de sujeto pronominal en distintas zonas de habla hispana

dictoras de la presencia o ausencia del pronombre sujeto. Como podemos observar en el cuadro 3, los predictores que han resultado ser significativos son: *persona del sujeto*, *referente del sujeto*, *especificidad*, *progresividad*, *perfectividad*, *clase semántica del verbo*, *turno de habla y edad*. Por la proximidad a *p*-valor, hemos considerado también relevantes las variables de *ambigüedad fonética*, *correferencialidad* y *perífrasis*. Para la observación de la ambigüedad fonética potencial se han estudiado la segunda persona del singular de los verbos, la tercera personal del singular y la tercera del plural, ya que se podía producir ambigüedad fonética entre ellas al elidir su consonante final.

Variable predictorora	X ²	Sig.	V de Cramer	Sig.
Persona del sujeto	120.235	.000	.353	.000
Referente del sujeto	23.846	.000	.157	.000
Especificidad	11.647	.000	.110	.000
Clase semántica del verbo	20.434	.002	.146	.002
Progresividad	5.820	.016	.078	.016
Perfectividad	5.602	.018	.076	.018
Turno de habla	4.674	.031	.070	.031
Edad	6.035	.049	.079	.049
Ambigüedad fonética	3.777	.052	.063	.052
Correferencialidad	5.558	.062	.076	.062
Género	3.087	.079	.057	.079
Perífrasis	4.994	.082	.072	.082
Tiempo verbal	7.704	.173	.089	.173
Ambigüedad morfológica	1.436	.231	.039	.231
Tipo de discurso	1.419	.701	.038	.701
Longitud de la forma verbal	6.196	.720	.080	.720
Tipo de cláusula	2.000	.736	.046	.736
Pronominal	0.090	.764	.010	.764
Modo	0.284	.868	.017	.868

Cuadro 3. Análisis de significación de variables predictoras en la expresión del pronombre personal sujeto mediante SPSS corpus de la ciudad de Málaga 1995

Como vemos, la variación lingüística no se produce por cuestiones estilísticas, sino que existen unos condicionantes lingüísticos que propician su aparición. Puesto que uno de los objetivos que persigue este trabajo es determinar las variables predictoras que afectan a la ausencia y presencia del pronombre, hemos realizado un análisis binomial de subida y bajada (*Binomial Up & Down*) mediante el programa *Goldvarb X* (Sankoff, Tagliamonte y Smith, 2005). Este análisis —más exhaustivo— señala si existe relación entre las variables independientes y la dependiente. Para ello, los pesos probabilísticos que superen 0.500 mostrarán influencia en la variable dependiente, es decir, propiciarán la aparición del pronombre personal sujeto.

Variable predictora	Pesos probabilísticos	%	Frecuencias absolutas
Persona del sujeto (p-valor: 0.000)			
1s	0.777	33.5	107/319
2s	0.711	26.2	21/80
3s	0.451	10.6	27/254
1p	0.294	5.7	10/176
3p	0.255	4.7	6/127
Especificidad (p-valor: 0.000)			
Específico	0.622	20.1	163/810
No específico	0.378	8.5	13/153
Progresividad (p-valor: 0.008)			
No progresivo	0.689	18.9	174/919
Progresivo	0.311	4.5	2/44
Perfectividad (p-valor: 0.027)			
Tiempo compuesto	0.577	28.2	22/78
Tiempo simple	0.423	17.4	154/885
Clase semántica (p-valor: 0.004)			
Verbos sensoriales	0.642	27.5	44/160
Otra clase	0.572	22.1	43/195
Verbos de movimiento	0.539	19.8	22/111
Verbos de <i>dicendi</i>	0.515	18.4	9/49
Verbos de estado	0.457	15.1	24/159
Verbos de actividad	0.387	11.8	24/104
Verbo auxiliar de perífrasis	0.387	11.8	10/85
Turno de habla (p-valor: 0.036)			
Cambio de turno	0.546	22.0	73/332
Mismo turno	0.454	16.3	103/631
Correferencialidad (p-valor: 0.065)			
No correferencial	0.578	21.7	84/387
Con sujeto	0.493	16.4	82/500
Con otro complemento	0.429	13.2	10/76

Variable predictorora	Pesos probabilísticos	%	Frecuencias absolutas
Perífrasis (p-valor: 0.048)			
Modal	0.609	20.7	6/29
No perífrasis	0.582	18.9	160/878
Aspectual	0.315	7.1	4/56
Log likelihood: -457.961 Input: 0.183			

Cuadro 4. Análisis de regresión logística binomial para la presencia de sujeto pronominal en corpus de 1995

En los resultados que muestra el cuadro 4, procedentes como hemos señalado de un análisis más exigente, ha dejado de ser significativa la variable de género, que sí lo era en el cuadro 3. Además, el predictor referente del sujeto es especialmente relevante porque solo se tienen muestras en ausencia y presencia cuando el referente es humano, mientras que, cuando es inanimado o neutro, solo encontramos ejemplos en ausencia. La razón que se esconde tras esta inexistencia de muestras radica en que su presencia se corresponde en una inmensa mayoría de casos con un sintagma nominal. Se han mantenido en el cuadro 4 la correferencialidad y la perífrasis para poder establecer comparaciones más adelante.

4.2.1 Condicionantes lingüísticos

4.2.1.1 Persona del sujeto

La persona del sujeto es una de las variables lingüísticas que presenta mayor significación en la presencia o ausencia del pronombre personal sujeto. La primera persona del singular es la que posee el peso probabilístico más elevado (0.777), y se aprecia una tendencia general donde los pronombres de singular tienen mayor fuerza que los pronombres del plural. Al encontrarnos ante el análisis de entrevistas semidirigidas en las que interactúan el informante y el entrevistado, estaríamos en una situación que propicia la aparición de los pronombres *yo* y *tú* o pronombres “de diálogo” (Enríquez, 1984, p. 191).

Mientras que la primera y la segunda personas hacen referencia al hablante y al oyente, que forman parte de la situación comunicativa, la tercera persona responde a otras posibilidades diferentes a la dicotomía de la ausencia o presencia del pronombre, puesto que ofrece otras opciones, como un sintagma nominal, pronombres demostrativos, pronombres relativos. Por lo tanto, la alta aparición del pronombre de tercera persona del singular (0.451) puede responder a un recurso puramente lingüístico como es la economía expresiva al hacer referencia a un elemento ya mencionado en el discurso (Miró Vera y De Pineda, 1990, p. 38).

4.2.1.2 Ambigüedad fonética potencial

Es llamativa la diferencia entre la significación de la ambigüedad fonética potencial (0.052) y la ambigüedad morfológica (0.231), donde podemos observar claramente que, dentro de la variable ambigüedad que recogía ambas formas, es la ambigüedad fonética la que propicia la aparición del sujeto.

No podemos obviar que nos encontramos analizando el habla de una ciudad donde la pérdida de la /-s/ final se da de manera generalizada. Gracias a los resultados obtenidos, podemos refutar la afirmación de Alarcos (1980) y Hochberg (1986), quienes defienden que la alta aparición del pronombre se debe a un papel desambiguador producido por la pérdida de la consonante final. De esta manera, nuestros resultados coinciden con los estudios de Ranson (1991) y Cameron (1993) al demostrar que hay una mayor presencia del pronombre cuando va acompañado de una forma verbal no ambigua (0.541). La posible inexactitud de los datos se debe a no haber podido realizar la escucha de manera completa en el estado actual de nuestra investigación.

4.2.1.3 Especificidad

Este predictor ha resultado ser influyente en la aparición del pronombre personal sujeto. Si seguimos la afirmación que realiza la *NGLE* (2009, p. 1134): “un argumento es específico cuando hace referencia a un ser, real o imaginario, que se considera identificable, al menos para el emisor. En caso contrario, se dice

que es inespecífico”. Como podemos observar en los ejemplos 3 y 4, hemos encontrado casos de inespecificidad con pronombres de segunda persona del singular y tercera persona del plural, respectivamente:

- (3) I.: lo mismo con unos que con otros / porque <(:)> si hicieron **unos**<(:)> daño los otros también lo hicieron / quizás los otros más por llevar más / <corrección> por tener más tiempo / ¿comprendes? <ruido = moto> [MA_M31_009].
- (4) bueno tenía un novio / <vacilación> la madre / <estilo directo> ¡ay niña! tú no <(:)> / **tú** no <(:)> te no <(:)> no te vaya <(:)> s a liar con él o no sé cuánto que ese viene aprovecharse de ti [MA_M21_906].

Los resultados indican una mayor presencia del pronombre cuando el referente es específico, mientras que con un referente inespecífico se tiende a rechazar la aparición del pronombre personal sujeto. Además, cabe destacar que el uso del pronombre *uno/a* está directamente relacionado con la inespecificidad, ya que hace referencia a cualquier persona (Manjón-Cabeza y otros, 2016).

4.2.1.4 Progresividad

La progresividad verbal ha resultado ser significativa para este corpus, aunque cabe destacar que no es muy frecuente la significación de esta variable en estudios anteriores. El predictor progresividad se compone de formas progresivas y formas no progresivas. Observamos que esta última incita considerablemente la presencia del pronombre personal sujeto con un peso probabilístico de 0.689.

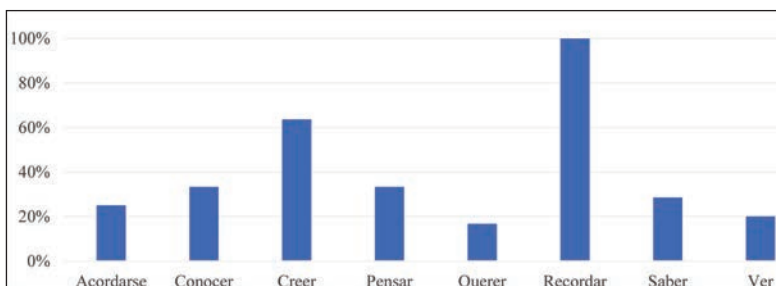
4.2.1.5 Perfectividad

La variable lingüística de la perfectividad afecta a la aparición del pronombre personal sujeto. Este predictor se divide en formas simples (*come, compras, etc.*) y formas compuestas (*ha comido, habías comprado*). Gracias a este análisis, podemos afirmar que la presencia de una forma verbal compuesta fomenta la presencia del pronombre (0.577).

4.2.1.6 Clase semántica del verbo

La clase semántica del verbo tiene influencia en la dicotomía ausencia y presencia del pronombre personal sujeto. Partíamos de la hipótesis de que los verbos de procesos mentales, opinión, creencia, entendimiento y verbos de percepción sensorial son más proclives a la aparición del pronombre (Enríquez, 1984; Bentivoglio, 1987), y los resultados mostrados en el cuadro 3, lo confirman. Son estos verbos los que acumulan el mayor peso probabilístico (0.642), seguidos de los verbos de otra clase (0.572), verbos de movimiento (0.539), verbos de *dicendi* (0.515) y verbos de estado (0.457).

En cuanto a los verbos de otra clase, hemos agrupado aquí los verbos que no encajaban en ninguna de las categorías anteriores, como: *costar*, *tener* o *poder*. Como novedad, en esta variable hemos incluido verbos auxiliares de perífrasis que, aunque no han resultado ser pertinentes para el cambio lingüístico en este corpus, puede ser que sí lo sean en otros. En el estudio de Manjón-Cabeza y otros en Granada (2016) y en el de Repede en Sevilla (2019) aparece una gráfica que muestra el número de apariciones de los verbos sensoriales, de entendimiento, opinión, creencia y procesos mentales. Hemos decidido representarlo con nuestros datos para así poder comparar los verbos más usados.



Gráfica 2. Presencia pronombre personal sujeto con verbos de entendimiento en el corpus 1995 de la ciudad de Málaga

Como podemos observar en la gráfica 2, los verbos que más aparecen con el sujeto pronominal en forma expresa son *creer*, *conocer* y *pensar*. Téngase en cuenta, en este sentido, que

el alto porcentaje del verbo *recordar* en presencia del pronombre pronominal se debe a las pocas apariciones de este. Si comparamos nuestros resultados con los de Granada y Sevilla (Manjón-Cabeza y otros, 2016 y Repede, 2019), ambos coinciden en que uno de los verbos que más aparece es *creer*.

4.2.1.7. Turno de habla

El análisis del cuadro 3 muestra que el turno de habla también es un predictor estadísticamente significativo para el fenómeno analizado. El porcentaje de aparición del sujeto en el mismo turno de habla es similar a la aparición general del pronombre (16.3%-18%). No obstante, cuando se produce un cambio de turno aumenta significativamente la aparición (22%) y obtiene un peso probabilístico del 0.546 siendo, por tanto, significativa. Estos resultados pueden responder a las características de la entrevista semidirigida que afecta a la naturalidad de la toma de turnos de habla, ya que se sigue el esquema de pregunta-respuesta.

4.2.2 Condicionantes sociales

En cuanto a los condicionantes sociales, hemos tenido en cuenta el género y el grupo de edad en el que se sitúa cada hablante. Recordemos que en este trabajo solo hemos analizado las realizaciones de hablantes con un nivel bajo de instrucción. Como podemos observar en el cuadro 3, el único predictor social que ha resultado ser significativo para este corpus ha sido la edad y podemos afirmar que el grupo 2 (0.568), es decir, los hablantes con una edad comprendida entre 35 y 55 años, son los que más utilizan el pronombre personal sujeto de manera tácita. Sin embargo, los grupos 1 y 3 tienden a un mayor empleo del pronombre en su forma omitida (0.466). La variable social del género de los informantes no ha resultado ser un predictor que afecte a la aparición del pronombre, ya que tiene una significación de 0.79.

5. Discusión

A continuación, vamos a comparar distintos corpus homólogos entre sí. Primero, de manera diacrónica para descubrir

si las variables predictoras que propician la presencia del pronombre personal sujeto han cambiado a lo largo de 20 años en la ciudad de Málaga. Después, vamos a comparar los resultados de estudios similares de distintas ciudades para poder establecer patrones similares.

5.1 *Comparación de resultados en la ciudad de Málaga (1995 y 2015)*

Gracias a los trabajos realizados por el grupo de investigación VUM (Vernáculo Urbano Malagueño) de la Universidad de Málaga, contamos con dos corpus de entrevistas semidirigidas homologables que emplean la metodología de PRESEEA en esta ciudad andaluza. El primer corpus es el que recoge el habla de 1995 en la ciudad de Málaga (Vida, 2007) y sobre este hemos basado gran parte del presente estudio. Además, hemos tenido acceso al corpus del 2015 (actualmente en desarrollo) y hemos decidido emplear la misma metodología de codificación y análisis para poder comparar así los resultados obtenidos dentro de la ciudad de Málaga de manera diacrónica. La muestra de habla de este segundo corpus se basa en las entrevistas semidirigidas de los hablantes de nivel de instrucción bajo (N=18). Se han analizado las realizaciones lingüísticas aparecidas entre los minutos 10 y 15.

En el análisis exploratorio obtenemos que en este corpus se han codificado un total de 1073 ejemplos donde la alternancia de presencia y ausencia del pronombre personal sujeto es posible, de los cuales 311 son de sujeto expreso (29%) y 762 (71%) presentan un sujeto omitido. Por lo tanto, se observa un aumento considerable de la presencia con respecto al corpus malagueño anterior (18%).

De igual manera que hicimos con el corpus de 1995, realizamos el análisis de correlación entre las variables para obtener los predictores que son significativos. Los resultados son los que se presentan en el cuadro 5:

Variable predictorora	χ^2	Sig.
Persona del sujeto	157.630	.000
Ambigüedad fonética potencial	28.431	.000
Especificidad	93.790	.000
Clase semántica del verbo	69.541	.000
Correferencialidad	19.660	.000
Perfectividad	9.961	.002
Perífrasis	12.437	.002
Turno de habla	7.978	.005
Referente del sujeto	7.165	.007
Género	7.048	.008
Modo	7.287	.026
Tipo de cláusula	9.792	.044
Pronominal	3.883	.049
Tipo de discurso	5.778	.123
Progresividad	1.954	.162
Tiempo verbal	7.079	.215
Edad	1.300	.522
Longitud de la forma verbal	4.903	.556
Ambigüedad morfológica	0.004	.950

Cuadro 5. Análisis de significación de variables predictoras en la expresión del pronombre personal sujeto mediante SPSS del corpus de la ciudad de Málaga 2015

Si comparamos el cuadro 3 y el cuadro 5, incluso tratándose de datos procedentes de la misma ciudad con 20 años de diferencia de un corpus respecto al otro, encontramos diferentes variables que propician la aparición del pronombre personal sujeto. En cuanto a las variables lingüísticas, observamos que existen algunas coincidencias, como la persona del sujeto, el referente humano, la especificidad, la perfectividad, la clase semántica del verbo y el turno de habla donde los valores de significación son aproximadamente semejantes. No obstante, algunas variables significativas resultan novedosas en la actualidad, como el modo (Sig.: 0.026), el tipo de cláusula (Sig.: 0.044), los verbos pronominales (Sig.: 0.049) y, en cuanto a los predictores

sociales, el género parece afectar a la presencia del pronombre (Sig.: 0.008). Cabe destacar que la ambigüedad fonética potencial, la perífrasis, la correferencialidad y el género se quedaban bastante cerca del p valor 0.05 en el primer corpus y en el segundo corpus consiguen ser significativas.

Una vez conocidos los predictores significativos, presentamos a continuación un análisis más detallado, *Binomial Up & Down* con el programa *Goldvarb*, que nos permitirá observar qué variables aportan mayor peso probabilístico en la aparición del pronombre personal sujeto en el corpus de 2015. Los resultados se muestran en el cuadro 6, donde llama la atención, especialmente, que han dejado de ser significativas las variables *modo* y *tipo de cláusula*, que sí lo eran en los resultados presentados en el cuadro 5.

Variable predictora	Pesos probabilísticos	%	Frecuencias absolutas
Persona del sujeto (p-valor: 0.000)			
Uno/una	0.807	60%	3/5
1s	0.748	51.7%	184/356
2s	0.503	26.7%	36/135
1p	0.417	20.4%	28/137
3s	0.404	19.6%	50/255
3p	0.141	5.6%	10/179
Ambigüedad fonética potencial (p-valor: 0.000)			
No ambiguo	0.590	36.2%	198/547
Ambiguo	0.410	21.5%	113/526
Especificidad (p-valor: 0.000)			
Específico	0.675	40.8%	247/606
No específico	0.325	13.7%	64/467
Clase semántica (p-valor: 0.00)			
Verbos sensoriales	0.760	53.9%	76/141
Verbo auxiliar de perífrasis	0.760	53.9%	76/141
Verbos de estado	0.617	37.3%	72/193
Verbos de movimiento	0.475	24.1%	32/133
Verbos de actividad	0.462	22.1%	32/145
Verbos de <i>dicendi</i>	0.381	18.5%	10/54
Otra clase	0.348	25%	66/264

Variable predictorora	Pesos probabilísticos	%	Frecuencias absolutas
Correferencialidad (p-valor: 0.000)			
No aplicable	0.575	36.6%	78/213
Con otro complemento	0.526	32.1%	52/162
No correferencial	0.519	31.5%	105/333
Con sujeto	0.382	20.8%	76/365
Turno de habla (p-valor: 0.008)			
Cambio de turno	0.554	36,2%	88/243
Mismo turno	0.446	26.9%	223/830
Perífrasis (p-valor: 0.001)			
No perífrasis	0.645	30.8%	289/938
Aspectual	0.473	18%	16/89
Modal	0.380	13%	6/46
Pronominal (p-valor: 0.040)			
Verbo no pronominal	0.563	29.9%	290/970
Verbo pronominal	0.437	20.4%	21/103
Género (p-valor: 0.009)			
Mujer	0.545	32.5%	373/553
Hombre	0.455	25.2%	131/520
Log likelihood: -645.956 Input: 0.290			

Cuadro 6. Análisis de regresión logística binomial para la presencia de sujeto pronominal en corpus de Málaga 2015

De los resultados mostrados en el cuadro 6 se pueden realizar las siguientes observaciones:

- Las personas del sujeto que más afectan a la aparición del pronombre son la primera persona del singular y los pronombres *uno* y *una*.
- La ambigüedad fonética potencial producida por la elisión de la consonante final del verbo, una vez más, vuelve a demostrar que no es necesario el pronombre para resolver tal ambigüedad.
- La especificidad del sujeto marca una mayor aparición del pronombre en el discurso.
- Las construcciones verbales con gerundio, es decir, progresivas, aumentan la frecuencia de la presencia del pronombre.

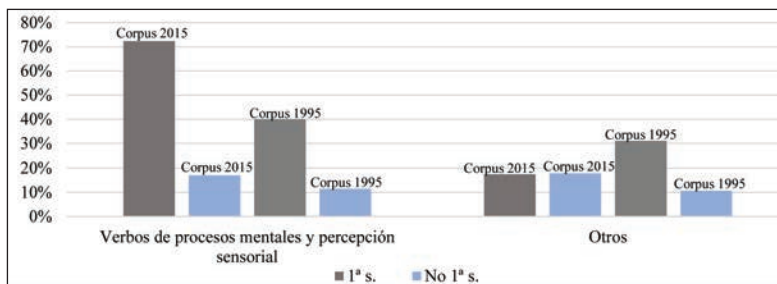
- La clase semántica del verbo, en concreto, si son verbos sensoriales y verbos auxiliares de perífrasis verbal, como novedad, manifiestan una mayor aparición del pronombre.
- La correferencialidad con otro complemento o que no sea correferencial propicia también el pronombre tácito.
- El cambio en el turno de habla también influye en la presencia.
- Las formas no perifrásticas tienen mayor peso probabilístico en la aparición del pronombre que las perifrásticas, sean del tipo que sean.
- Las construcciones no pronominales van acompañadas de una mayor presencia de pronombres expresos.
- En cuanto al predictor social significativo, el género, las mujeres tienden a hacer mayor uso del pronombre tácito que los hombres.

El cuadro 7 recoge, de manera comparada, los predictores que en los cuadros 4 y 6 demostraron ser significativos. Hemos identificado en una columna los predictores comunes en los corpus de 1995 y 2015, y los hemos separado de aquellos que mostraron significación exclusiva en 1995 y en 2015.

Comunes en 1995 y 2015	Exclusivos en el corpus Málaga 1995	Exclusivos en el corpus Málaga 2015
1. Persona del sujeto (1ª persona del singular)	A. Progresividad (No progresivo)	C. Ambigüedad fonética potencial (No ambiguo)
2. Especificidad (Específico)	B. Perfectividad (Tiempo compuesto)	D. Pronominal (No pronominal)
3. Clase semántica del verbo (Verbos de procesos mentales, de percepción sensorial y opinión)		E. Género (Mujer)
4. Turno de habla (Cambio de turno)		
5. Perífrasis verbal (No perífrasis verbal)		
6. Correferencialidad (No correferencial)		

Cuadro 7. Predictores significativos coincidentes o exclusivos en los análisis de regresión logística binomial para la presencia de sujeto pronominal en los corpus de 1995 y 2015

De los predictores significativos en ambos corpus, decidimos combinar los dos que mostraron mayor peso explicativo (persona del verbo y su clase semántica) para comprobar si existía algún tipo de aumento de la presencia del pronombre cuando se relacionan.



Gráfica 3. Porcentaje de presencia del sujeto pronominal combinado con los verbos mentales y la primera persona

La gráfica 3 muestra el porcentaje de la presencia del pronombre cuando se realizan las combinaciones señaladas. Como podemos observar, se produce un aumento considerable de la presencia del pronombre cuando un verbo mental se conjuga en primera persona del singular, y esto sucede en ambos corpus. Por lo tanto, podemos afirmar que existe interacción entre las variables lingüísticas de persona del sujeto (primera persona del singular) y clase semántica (verbos de procesos mentales), pues afectan conjuntamente a la aparición del sujeto pronominal.

5.2 Comparaciones adicionales (Sevilla, Granada, Valencia y Málaga)

Gracias a que, como hemos señalado, PRESEEA ha propiciado la aplicación de una metodología unificada, nuestro trabajo parte de un corpus homologable que nos permite establecer comparaciones de resultados parciales en trabajos realizados sobre este tema. Hemos seleccionado los estudios de la ciudad de Sevilla (Repede, 2019), Granada (Manjón-Cabeza y otros, 2016) y Valencia (Roselló, 2020). Todos ellos parten de la misma guía de codificación (Bentivoglio, Ortiz y Silva-Corvalán, 2011), pero,

además, han introducido nuevos predictores que han considerado relevantes para adaptarla a su población, como es el bilingüismo que se produce en Valencia o la elisión de las consonantes finales características de la variedad del español andaluz.

Para este estudio, en los dos corpus que hemos manejado de la ciudad de Málaga hemos limitado la observación a los informantes del nivel bajo de estudios. Lo mismo sucede con los estudios realizados hasta el momento en PRESEEA-Sevilla, que reduce su muestra a solo los hablantes con un sociolecto alto. Por el contrario, los trabajos de Granada y Valencia recogen el total de la muestra, 54 y 72 informantes respectivamente, de todos los niveles de instrucción. En cuanto a los minutos de la entrevista semidirigida, todos los trabajos, a excepción de Sevilla, seleccionan los cinco minutos comprendidos entre el 10 y el 15.

En cuanto al análisis de los datos obtenidos tras la codificación, observamos que, en las investigaciones de Sevilla y Valencia, al igual que en nuestro trabajo, se realiza en una primera aproximación estadística la prueba de χ^2 para extraer los predictores que han resultado ser significativos, mientras que en Granada se utiliza la diferencia de medias de todas sus variables. Con ambos análisis, se ha permitido eliminar las variables con un p -valor > 0.05 . En el cuadro 8 se muestra la significación de las variables, teniendo en cuenta solo las comunes a todos los estudios:

Variable predictora	Málaga (Corpus de 1995)	Málaga (Corpus de 2015)	Valencia (Roselló, 2021)	Sevilla (Repede, 2019)	Granada (Manjón-Cabeza y otros, 2016)
Persona del sujeto	.000	.000	.000	.000	.000
Especificidad	.000	.000	.022	.000	.000
Modo	.868	.026		.823	.260
Tiempo verbal	.173	.215	.032	.000	.067
Progresividad	.016	.162		.069	.461
Perfectividad	.018	.002		.067	.122
Ambigüedad	.231	.950	.000	.000	.000
Clase semántica del verbo	.002	.000	.000	.000	.000
Tipo de cláusula	.736	.044		.151	.102

Variable predictoras	Málaga (Corpus de 1995)	Málaga (Corpus de 2015)	Valencia (Roselló, 2021)	Sevilla (Repede, 2019)	Granada (Manjón-Cabeza y otros, 2016)
Correferencialidad	.062	.000	.000	.000	.000
Turno de habla	.031	.005	.000	.000	.008
Perífrasis verbal	.048	.002	.001	.002	.002
Edad	.049	.522	.359	.000	.000
Género	.079	.008	.133	.026	.457

Cuadro 8. Comparativa de resultados de variables predictoras significativas en diferentes corpus de PRESEEA España

Tras observar los resultados de los análisis procedentes de las diferentes sintopías, podemos afirmar que los predictores significativos comunes a los cinco corpus son: la persona del sujeto, la especificidad, la ambigüedad, la clase semántica, la co-referencialidad, el turno de habla y las perífrasis verbales. En cuanto a los predictores sociales, no hay una tendencia clara que nos permita afirmar que la edad o el género sean influyentes en todos los corpus analizados. Cabe destacar que la perífrasis verbal, que no aparece en la guía de codificación original, se muestra como un condicionante claro que afecta a la aparición del pronombre personal sujeto en los cinco estudios, con un alto grado de significación.

Los trabajos mencionados, al igual que el que presentamos en este artículo, tras encontrar las variables predictoras que afectan al cambio, realizan un análisis más detallado para descubrir dentro de cada variable cuál de las opciones recogidas es la que alberga mayor peso probabilístico. Para ello, en el estudio de Valencia, Granada y en nuestro trabajo se ha utilizado el análisis de *Binomial Up & Down* con el programa de *Goldvarb*; en Sevilla, en cambio, se ha optado por un recuento de casos en presencia y ausencia sobre el desglose de cada predictor.

Al contrastar todos los resultados, tanto los obtenidos por medio del análisis de nuestros dos corpus como en estudios previos (Repede, 2019; Roselló, 2021 y Manjón-Cabeza y otros, 2016), podemos afirmar que la presencia o ausencia del pronombre personal no responde a cuestiones estilísticas, sino a

cuestiones lingüísticas que se repiten en los cinco corpus. Existe una mayor presencia del pronombre cuando:

- El sujeto de la oración es la primera persona del singular.
- El sujeto tiene un referente específico.
- El verbo que acompaña al pronombre es un verbo relacionado con los procesos mentales, opinión, creencia, entendimiento y verbos de percepción sensorial.
- No hay correlación con ninguna partícula previa del discurso.
- Se produce un cambio en el turno del habla.
- La construcción verbal no es perifrástica.

Por lo tanto, de esta aproximación parcial podríamos vislumbrar que la presencia del pronombre personal sujeto responde a un patrón lingüístico que se repite en los diversos corpus analizados.

6. Conclusiones

El objetivo principal de este trabajo ha sido contribuir a la investigación de los predictores que afectan a la ausencia o a la presencia del pronombre personal sujeto en las realizaciones de hablantes con un nivel de estudios bajo en la ciudad española de Málaga. Los análisis estadísticos nos han permitido obtener una serie de conclusiones parciales que nos conducen a identificar las variables predictoras que favorecen la aparición del sujeto pronominal. En líneas generales, en los corpus analizados de la ciudad de Málaga se aprecia un 17% de aparición del pronombre en 1995 y un 29% en 2015. La tasa pronominal más alta encontrada en 2015 podría asociarse con la tendencia evolucionaria que se ha propuesto para las lenguas romances basada, entre otros condicionantes, en un incremento progresivo de este indicador (Erker y Guy, 2012, Orozco, 2015, p. 32).

Hemos observado que, incluso tratándose de datos procedentes de la misma ciudad, los resultados cualitativos no coinciden completamente, ya que se trata de dos cortes temporales diferentes. En el corpus de Málaga de 1995 hemos obtenido que la presencia está condicionada por las variables predictoras *persona del sujeto, referente del sujeto, especificidad, progresividad,*

perfectividad, turno de habla, clase semántica del verbo y edad. Mientras que en la misma ciudad en 2015 los predictores que la propician son *persona del sujeto, referente del sujeto, ambigüedad fonética potencial, especificidad, modo, perfectividad, clase semántica del verbo, tipo de cláusula, correferencialidad, turno de habla, perífrasis, verbos pronominales y género.* Lo que nos ha llevado a corroborar que los predictores lingüísticos son más significativos que los sociales (hipótesis del peso de las variables predictoras). En una aproximación más exhaustiva hemos descubierto qué subvariables poseen mayor peso probabilístico, lo que nos ha permitido obtener conclusiones más específicas. Además, hemos comprobado que la primera persona del singular (hipótesis de la persona verbal) y los verbos de procesos mentales (hipótesis semántica) aumentan la presencia del pronombre cuando se combinan (gráfica 3). Sin embargo, la anunciada como *hipótesis de la especificidad dialectal* (hipótesis 2) ha sido rechazada, pues, tal y como hemos comprobado, la ambigüedad fonética, producida por la elisión de la /-s/ u otras consonantes en posición de distensión silábica propia de la variedad estudiada, no afecta significativamente a la elisión del pronombre en el corpus de 1995 y solo se muestra relevante en el corpus de 2015, cuando aparece una forma verbal no ambigua fonéticamente.

Una vez analizados e interpretados los resultados de la ciudad de Málaga, hemos procedido a compararlos con otros estudios similares (Repede, 2019; Roselló, 2021 y Manjón-Cabeza y otros, 2016) para poder establecer un posible patrón lingüístico dentro de la presencia del sujeto pronominal. De este modo, a pesar de los diferentes estratos educacionales considerados, podemos señalar que existe una mayor presencia del pronombre cuando se dan las siguientes circunstancias: el sujeto de la oración es la primera persona del singular; el sujeto tiene un referente específico; el verbo que acompaña al pronombre es un verbo relacionado con los procesos mentales, opinión, creencia, entendimiento y verbos de percepción sensorial; no hay correlación con ninguna partícula previa del discurso; se produce un cambio en el turno del habla (hipótesis de la estrategia comunicativa), y la construcción verbal no es perifrástica. Por lo tanto, podemos concluir que la hipótesis de corroboración general de

nuestro estudio se cumple, ya que los condicionantes lingüístico-sociales que condicionan la aparición u omisión del pronombre sujeto son similares a los que se obtuvieron en estudios previos que nos han servido de comparación.

En definitiva, este trabajo supone una contribución a la investigación sobre la presencia o ausencia del pronombre personal sujeto donde hemos desarrollado la misma metodología empleada en trabajos anteriores para establecer comparaciones, y hemos descubierto que la profundidad y riqueza del fenómeno analizado apunta hacia patrones de actualización sistemáticos. En investigaciones futuras está previsto el análisis de los datos a partir de otros procedimientos estadísticos que nos ofrezcan más información de este fenómeno para poder observar con mayor detalle otros aspectos interesantes, como la importancia del referente del sujeto o la posible relación entre diferentes predictores significativos. Puesto que la inclusión de nuevas variables predictoras ha resultado ser operativa en nuestro estudio parcial, consideramos que, a partir de ahora, la guía de codificación del proyecto PRESEEA podría incluirlas en sus recomendaciones para comprobar su eficacia en diferentes sintopías. Nuestra intención en posteriores etapas de esta investigación es añadirlas en la ampliación de la muestra a otros niveles de estudios para poder tener una visión real y completa de los condicionantes que afectan al fenómeno estudiado en la población general.

Referencias

- ALARCOS LLORACH, E. (1980). Los pronombres personales. En E. Alarcos Llorach (ed.), *Estudios de gramática funcional del español* (pp. 200-212). Madrid: Gredos.
- BARRENECHEA, A. M. y ALONSO, A. (1977). Los pronombres personales sujetos en el español hablado en Buenos Aires. En J. M. Lope Blanch (ed.), *Estudios sobre el español hablado en las principales ciudades de América* (pp. 333-349). México: UNAM.
- BENTIVOGLIO, P. (1980). *Why canto and not yo canto? The problem of first-person subject pronoun in spoken Venezuelan Spanish*. Tesis de Maestría. Universidad de California.

- BENTIVOGLIO, P. (1987). *Los sujetos pronominales de primera persona en el español hablado de Caracas*. Caracas: Universidad Central de Venezuela.
- BENTIVOGLIO, P., ORTIZ, L. A. y SILVA-CORVALÁN, C. (2011). *La variable de expresión del sujeto pronominal. Guía de codificación*. Recuperado de <https://presea.linguas.net/Portals/0/Metodologia/guia_codificacion_sujetos_julio_2011.pdf>.
- BLAS ARROYO, J., CASANOVA, M. y VELANDO, M. (2006). *Discurso y sociedad: contribuciones al estudio de la lengua en contexto social*. Castellón: Universidad Jaume I.
- CAMERON, R. (1993). Ambiguous Agreement, Functional Compensation, and Non-Specific *tú* in the Spanish of San Juan, Puerto Rico and Madrid. *Language Variation and Change*, 5, 305-334.
- CARVALHO, A. M., OROZCO, R. y SHIN, N. L. (2015). *Subject Pronoun Expression in Spanish in Contact with Portuguese*. En A. M. Carvalho, R. Orozco y N. L. Shin (eds.), *Subject Pronoun Expression in Spanish* (pp. 143-165). Washington: Georgetown University Press.
- CERRÓN-PALOMINO, A. (2018). Variable Subject Pronoun Expression in Andean Spanish: a Drift from the Acrolect. *Onomázein*, 42, 53-73.
- CHOMSKY, N. (1981). *Lectures on Government and Binding*. Pisa: Foris.
- DEMONTÉ, V. y BOSQUE, I. (2000). *Gramática descriptiva de la lengua española*, tomo 1: *Sintaxis básica de las clases de palabras*. Madrid: Espasa-Calpe.
- ENRÍQUEZ, E. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Instituto Miguel de Cervantes.
- ERKER, D. y GUY, G. R. (2012). The Role of Lexical Frequency in Syntactic Variability: Variable Subject Personal Pronoun Expression in Spanish. *Language*, 88, 526-565.
- FERNÁNDEZ RAMÍREZ, S. (1951). *Gramática española: el pronombre*. Madrid: Arco/Libro.
- FERNÁNDEZ SORIANO, O. (1999). El pronombre personal. Formas y distribuciones. Pronombres átonos y tónicos. En I.

- Bosque y V. Demonte (dirs.), *Gramática descriptiva de la lengua española 1* (pp. 1209-1273). Madrid: Espasa.
- FLORES FERRÁN, N. (2007). Los Mexicanos in New Jersey: Pronominal Expression and Ethnolinguistic Aspects. En J. Holmquist, A. Lorenzino y L. Sayahi (eds.), *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics* (pp. 85-91). Somerville: Cascadilla Proceedings Project.
- GILI GAYA, S. (1989). *Curso superior de sintaxis española*. Barcelona: Vox.
- GIVÓN, T. (1970). Topic, Pronoun, Grammatical Agreement. En C. Li (ed.), *Subject and Topic* (pp. 151-188). Nueva York: New York Academic Press.
- GUERRERO, S. (2019). Expresión del sujeto pronominal en el corpus de PRESEEA de Santiago de Chile. Seminario Internacional Las Investigaciones Lingüísticas en el Mundo Hispánico. La Habana, Cuba.
- HOCHBERG, J. (1986). Functional Compensation for /s/ Deletion in Puerto Rican Spanish. *Language*, 62, 609-621.
- LASTRA, Y. y MARTÍN-BUTRAGUEÑO, P. (2015). Subject Pronoun Expression in Oral Mexican Spanish. En A. M. Carvalho, R. Orozco y N. L. Shin (eds.), *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective* (pp. 39-57). Washington: Georgetown University Press.
- LUJÁN, M. (2000). Expresión y omisión del pronombre personal. En I. Bosque y V. Demonte (dirs.), *Gramática descriptiva de la lengua española 1* (pp. 1277-1313). Madrid: Espasa.
- MANJÓN-CABEZA CRUZ, A., POSE FUREST, F. y SÁNCHEZ GARCÍA, F. (2016). Factores determinantes en la expresión del sujeto pronominal en el corpus PRESEEA Granada. *Boletín de Filología*, 51(2), 181-207.
- MARTÍN-BUTRAGUEÑO, P. (2020). An Approach to Subject Pronoun Expression Patterns in Data from the “Project for the Sociolinguistic Study of Spanish in Spain and America”, *Spanish in Context*, 17(2), 294-316.
- MARTÍNEZ-LARA, J. A., Guerrero, S. y González, C. (2020). (Yo) opino: la variación del sujeto pronominal en el habla de Santiago de Chile. En B. Rogers y M. Figueroa Candia (eds.), *Lingüística del castellano chileno: estudios sobre variación, innovación, contacto e identidad*. Málaga: Vernon.

- MIRÓ VERA, R. y DE PINEDA PÉREZ, M. A. (1990). Determinación sociolingüística de la presencia/ausencia del pronombre personal sujeto. En M. T. Palet (ed.), *Sociolingüística andaluza 5. Habla de Sevilla y hablas americanas* (pp. 37-44). Sevilla: Universidad de Sevilla.
- OROZCO, R. (2015). Pronominal Variation in Colombian Costeño Spanish. En A. M. Carvalho, R. Orozco y N. L. Shin (eds.), *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective* (pp. 17-37). Washington: Georgetown University Press.
- OROZCO, R. y GUY, G. R. (2008). El uso variable de los pronombres sujetos: ¿qué pasa en la costa Caribe colombiana? En M. Westmoreland y J. A. Thomas (eds.), *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics* (pp. 70-80). Somerville: Cascadilla Proceedings Project.
- OROZCO, R. y HURTADO CUBILLOS, L. M. (2019, junio 26-28). A Sociolinguistic Study of Pronominal Expression in Mainland Colombian Spanish. *International Conference on Language Conference in Europe*. Leeuwarden, Países Bajos.
- OTHEGUY, R. y ZENTELLA, A. C. (2012). *Spanish in New York: Language Contact, Dialectal Leveling, and Structural Continuity*. Oxford: Oxford University Press.
- PESKOVA, A. (2015). *Sujetos pronominales en el español porteño. Implicaciones pragmáticas en la interfaz sintáctico-fonológica*. Berlín: De Gruyter Mouton.
- RANSON, D. (1991). Person Marking in the Wake Og /s/Deletion in Andalusian Spanish. *Language Variation and Change*, 3(2), 133-152.
- REAL ACADEMIA ESPAÑOLA. (1973). *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- REAL ACADEMIA ESPAÑOLA y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.
- REAL ACADEMIA ESPAÑOLA y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2010). *Nueva gramática de la lengua española: Manual*. Madrid: Espasa.
- REPEDE, D. (2019). Condicionantes sociolingüísticos de la expresión del sujeto pronominal en el sociolecto alto del cor-

- pus PRESEEA-Sevilla. *Revista de Investigación Lingüística*, 22, 397-423.
- ROSELLÓ VERDEGUER, J. (2021). Factores determinantes en la expresión del sujeto pronominal en español hablado de Valencia. *LINRED*, XVII, 1-24.
- SANKOFF, D., TAGLIAMONTE, S. y SMITH, E. (2005). *GoldVarb X. A Variable Rule Application for Macintosh and Windows*. Toronto: University of Toronto.
- SHIN, N. L. y OTHEGUY, R. (2013). *Social Class and Gender Impacting Change*. Cambridge: CUP.
- SILVA-CORVALÁN, C. (1982). Subject Variation in Spoken Mexican-American Spanish. En J. Amastae y L. Elías-Olivares (eds.), *Spanish in the United States* (pp. 93-120). Cambridge: CUP.
- SILVA-CORVALÁN, C. (1989). *Sociolingüística: Teoría y análisis*. Madrid: Alhambra.
- SILVA-CORVALÁN, C. (1994). *Language Contact and Change. Spanish in Los Angeles*. Óxford: Clarendon Press.
- SILVA-CORVALÁN, C. (2003). Otra mirada a la expresión del sujeto como variable sintáctica. En F. Moreno Fernández, F. Gimeno Menéndez, J. A. Samper, M. L. Gutiérrez Araus, M. Vaquero y C. Hernández (eds.), *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales* (pp. 849-860). Madrid: Arco/Libros.
- SILVA-CORVALÁN, C. (2015). Infantes bilingües y hablantes de herencia adultos: ¿Qué los vincula? *Boletín de Filología*, 50(1), 165-191.
- VIDA, M. (2007). *El español hablado en Málaga. Nivel de estudios bajo*. Málaga: Sarria.

