



UNIVERSIDAD
DE MÁLAGA



HEIDELBERG
CENTER FOR
IBERO-AMERICAN
STUDIES



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

ESTUDIO DE LA VARIACIÓN DEL SUJETO PRONOMINAL EN LA CIUDAD DE MÁLAGA: SOCIOLINGÜÍSTICA COGNITIVA DE SU PRODUCCIÓN Y PERCEPCIÓN

Alba Segura Lores

TESIS DOCTORAL

Facultad de Filosofía y Letras. Programa de Doctorado en Lingüística, Literatura y Traducción

Dissertation zur Erlangung des Doktorgrades an der Neuphilologischen Fakultät der Universität Heidelberg

DIRECTORES:

Francisco Moreno Fernández, Francisco Díaz Montesinos y Matilde Vida Castro

TUTOR:

Antonio Manuel Ávila Muñoz

HEIDELBERG, 2025





UNIVERSIDAD
DE MÁLAGA

AUTORA: Alba Segura Lores

 <http://orcid.org/0000-0002-0788-1950>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

<http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Cualquier parte de esta obra se puede reproducir sin autorización pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): riuma.uma.es



UNIVERSIDAD
DE MÁLAGA



Escuela de Doctorado

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña ALBA SEGURA LORES

Estudiante del programa de doctorado LINGÜÍSTICA, LITERATURA Y TRADUCCIÓN de la Universidad de Málaga, autor/a de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: ESTUDIO DE LA VARIACIÓN DEL SUJETO PRONOMINAL EN LA CIUDAD DE MÁLAGA: SOCIOLINGÜÍSTICA COGNITIVA DE SU PRODUCCIÓN Y PERCEPCIÓN

Realizada bajo la tutorización de ANTONIO MANUEL ÁVILA MUÑOZ y dirección de FRANCISCO DÍAZ MONTESINOS, MATILDE VIDA CASTRO Y FRANCISCOS MORENO FERNÁNDEZ
(si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 11 de JULIO de 2025

Fdo.: ALBA SEGURA LORES
Doctorando/a

Fdo.: ANTONIO MANUEL ÁVILA MUÑOZ
Tutor/a

Fdo.: FRANCISCO DÍAZ MONTESINOS



EFQM AENOR



Edificio Pabellón de Gobierno. Campus El Ejido.
29071
Tel.: 952 13 10 28 / 952 13 14 61 / 952 13 71 10
E-mail: doctorado@uma.es





UNIVERSIDAD
DE MÁLAGA



Escuela de Doctorado

MATILDE VIDA CASTRO

FRANCISCO MORENO FERNÁNDEZ
Director/es de tesis



EFQM AENOR



Edificio Pabellón de Gobierno. Campus El Ejido.
29071
Tel.: 952 13 10 28 / 952 13 14 61 / 952 13 71 10
E-mail: doctorado@uma.es





UNIVERSIDAD
DE MÁLAGA

Facultad de Filosofía y Letras

Málaga a 9 de julio de 2025

Matilde Ángeles Vida Castro, profesora de la Facultad de Filosofía y Letras (Universidad de Málaga),

HACE CONSTAR

Que ALBA SEGURA LORES, con DNI es estudiante de doctorado del Programa de Doctorado "Lingüística, Literatura y Traducción", con matrícula activa, y que ha realizado bajo mi dirección, la Tesis Doctoral titulada

"ESTUDIO DE LA VARIACIÓN DEL SUJETO PRONOMINAL EN LA CIUDAD DE MÁLAGA:

SOCIOLINGÜÍSTICA COGNITIVA DE SU PRODUCCIÓN Y PERCEPCIÓN"

Revisado el presente trabajo estimo que reúne los requisitos establecidos según la normativa vigente. Por lo tanto, **AUTORIZO** la admisión a trámite y defensa pública de esta Tesis Doctoral para optar al grado de Doctor en la Universidad de Málaga.

Y para que así conste, lo firmo en Málaga a 9 de julio de 2025,

Fdo. Matilde Ángeles Vida Castro



Campus de Teatinos s/n, 29071 Málaga
Tel.: 952 13 16 83/1684/1685/1687/3432/3435 - Fax: 952 13 16 23

UNIVERSIDAD
DE MÁLAGA





Heidelberg, 10 de julio de 2025

Francisco Moreno Fernández, profesor del Heidelberg Center for Ibero-American Studies (Universität Heidelberg),

HACE CONSTAR

Que ALBA SEGURA LORES, con DNI _____ es estudiante de doctorado del Programa de Doctorado "Lingüística, Literatura y Traducción", con matrícula activa, y que ha realizado bajo mi dirección, la Tesis Doctoral titulada

"ESTUDIO DE LA VARIACIÓN DEL SUJETO PRONOMINAL EN LA CIUDAD DE MÁLAGA:

SOCIOLINGÜÍSTICA COGNITIVA DE SU PRODUCCIÓN Y PERCEPCIÓN"

Revisado el presente trabajo estimo que reúne los requisitos establecidos según la normativa vigente. Por lo tanto, **AUTORIZO** la admisión a trámite y defensa pública de esta Tesis Doctoral para optar al grado de Doctor en la Universidad de Málaga.

Y para que así conste, lo firmo en Heidelberg a 10 de julio de 2025,

Fdo. Francisco Moreno Fernández



Málaga a 10 de julio de 2025

Francisco Díaz Montesinos, profesor de la Facultad de Filosofía y Letras (Universidad de Málaga),

HACE CONSTAR

Que ALBA SEGURA LORES, con DNI es estudiante de doctorado del Programa de Doctorado "Lingüística, Literatura y Traducción", con matrícula activa, y que ha realizado bajo mi dirección, la Tesis Doctoral titulada

"ESTUDIO DE LA VARIACIÓN DEL SUJETO PRONOMINAL EN LA CIUDAD DE MÁLAGA:

SOCIOLINGÜÍSTICA COGNITIVA DE SU PRODUCCIÓN Y PERCEPCIÓN"

Revisado el presente trabajo estimo que reúne los requisitos establecidos según la normativa vigente. Por lo tanto, **AUTORIZO** la admisión a trámite y defensa pública de esta Tesis Doctoral para optar al grado de Doctor en la Universidad de Málaga.

Y para que así conste, lo firmo en Málaga a 10 de julio de 2025,

Fdo. Francisco Díaz Montesinos





Málaga a 9 de julio de 2025

Antonio Manuel Ávila Muñoz, profesor de la Facultad de Filosofía y Letras (Universidad de Málaga),

HACE CONSTAR

Que ALBA SEGURA LORES, con DNI _____ es estudiante de doctorado del Programa de Doctorado "Lingüística, Literatura y Traducción", con matrícula activa, y que ha realizado bajo mi dirección, la Tesis Doctoral titulada

"ESTUDIO DE LA VARIACIÓN DEL SUJETO PRONOMINAL EN LA CIUDAD DE MÁLAGA:

SOCIOLINGÜÍSTICA COGNITIVA DE SU PRODUCCIÓN Y PERCEPCIÓN"

Revisado el presente trabajo estimo que reúne los requisitos establecidos según la normativa vigente. Por lo tanto, **AUTORIZO** la admisión a trámite y defensa pública de esta Tesis Doctoral para optar al grado de Doctor en la Universidad de Málaga.

Y para que así conste, lo firmo en Málaga a 9 de julio de 2025

Fdo. Antonio Manuel Ávila Muñoz



Campus de Teatinos s/n. 29071 Málaga
Tel.: 952 13 16 83/1684/1685/1687/3432/3435 - Fax: 952 13 16 23



INFORME SOBRE LAS PUBLICACIONES EN COAUTORÍA

Doctoranda: Alba Segura Lores

Tutor: Antonio Manuel Ávila Muñoz

Las publicaciones que se especifica a continuación y que avalan la tesis doctoral de Dña. **Alba Segura Lores** junta otras publicaciones en las que la doctoranda es la única firmante, no han sido utilizadas para avalar tesis anteriores:

- 1) Los estudiantes andaluces antes los nuevos retos de una sociedad en transformación. Tirant lo Blanch
- 2) Estudio de las variables predictoras de la expresión del sujeto pronominal en los corpus PRESEEA-Málaga. Nivel de instrucción bajo. Revista Anuario de Letras. Lingüística y Filología 10/2 (2022) de la Universidad Nacional Autónoma de México.

En Málaga, a 15 de julio de 2025





Deutscher Akademischer Austauschdienst
German Academic Exchange Service

This Doctoral Thesis has been financed by German Academic Exchange Service.
Scholarship Research Grants – Bi-nationally Supervised Doctoral Degrees/Cotutelle
2024/2025 (57693451).



OBSERVATORY
OF **SPANISH**
IN EUROPE

This Doctoral Thesis is part of the Project Observatory of Spanish in Europe

ÍNDICE

INTRODUCCIÓN.....	23
CAPÍTULO 1. VARIACIÓN DEL SUJETO PRONOMINAL: MARCO TEÓRICO Y ANTECEDENTES	27
1.1. OBJETO DE ESTUDIO: LA ALTERNANCIA AUSENCIA-PRESENCIA DEL SUJETO PRONOMINAL EN ESPAÑOL	27
1.2. MARCO TEÓRICO	31
1.2.1. La sociolingüística variacionista	31
1.2.1.1. La variación gramatical	33
1.2.2. La sociolingüística cognitiva	36
1.2.2.1. Sociogramática	39
1.3. ANTECEDENTES	41
1.3.1. Trabajos sobre el uso del sujeto pronominal	42
1.3.1.1. Perspectiva variacionista	42
1.3.1.2. Investigaciones en el contexto de PRESEEA.....	45
1.3.1.3. Factores explicativos de la alternancia	47
1.3.2. Trabajos sobre la percepción del sujeto pronominal	51
1.4. OBJETIVOS, PREGUNTAS E HIPÓTESIS DE INVESTIGACIÓN	53
1.4.1. Producción del sujeto pronominal	53
1.4.2. Percepción cognitiva del sujeto pronominal.	55
CAPÍTULO 2. LA PRODUCCIÓN DEL PRONOMBRE PERSONAL SUJETO EN LOS CORPUS PRESEEA-MÁLAGA.....	59
2.1. INTRODUCCIÓN AL ANÁLISIS MEDIANTE CORPUS	59
2.2. PROPUESTA METODOLÓGICA PARA EL ESTUDIO DEL PRONOMBRE PERSONAL EN LA CIUDAD DE MÁLAGA	61
2.2.1. Muestra	63

2.2.2. Selección de casos y variables lingüístico-sociales.....	64
2.2.3. Herramientas y análisis	69
2.3. ANÁLISIS ESTADÍSTICOS CORPUS PRESEEA-MÁLAGA	70
2.3.1. Análisis exploratorio	70
2.3.2. Análisis bivalente	73
2.3.2.1. Variables referentes al sujeto de la oración.....	73
2.3.2.2. Variables referentes a cuestiones pragmático-discursivas	82
2.3.2.3. Variables referentes al verbo de la oración	102
2.3.2.4. Variables sociales	112
2.3.3. Análisis multivalente.....	119
2.3.3.1. Análisis de modelos no paramétricos de regresión: árboles de inferencia y bosques aleatorios.....	119
2.3.3.2. Análisis de regresión logística con efectos mixtos.....	123
2.4. COMPARACIÓN DE LOS ANÁLISIS DE LOS CORPUS PRESEEA	132
2.4.1. Comparación diacrónica: El sujeto pronominal a lo largo del tiempo en Málaga	132
2.4.2. Comparación diatópica: El sujeto pronominal en la geografía hispana	138
2.4.2.1. El patrón de uso del pronombre personal sujeto	145
2.5. DISCUSIÓN Y CONCLUSIONES PARCIALES.....	147
CAPÍTULO 3. LA PERCEPCIÓN DE LA AMBIGÜEDAD MORFOLÓGICA Y LA CORREFERENCIALIDAD. ANÁLISIS EXPERIMENTAL MEDIANTE EYE- TRACKING	153
3.1. INTRODUCCIÓN AL ANÁLISIS EXPERIMENTAL MEDIANTE EYE-TRACKING.....	153
3.2. COMPRENSIÓN Y PROCESAMIENTO DEL DISCURSO	155
3.3. LECTURA CON EYE-TRACKING: DE LA FIJACIÓN OCULAR A LA COMPRENSIÓN.....	159
3.4. DISEÑO EXPERIMENTAL	163
3.4.1. Variables, materiales y áreas de interés.....	163

3.4.2. Participantes, procedimiento, equipamiento y tratamiento estadístico	168
3.5. ANÁLISIS	170
3.5.1. Ambigüedad referencial	170
3.5.2. Continuidad y cambio referencial	177
3.6. DISCUSIÓN Y CONCLUSIONES PARCIALES	183
CAPÍTULO 4. CONCLUSIONES	189
4.1. PRÓXIMOS OBJETIVOS	193
BIBLIOGRAFÍA	197
ANEXO I. GUÍA DE CODIFICACIÓN EXPRESIÓN DEL SUJETO PRONOMINAL EN CORPUS PRESEEA-MÁLAGA	214
ANEXO II. LISTADO DE LOS EFECTOS ALEATORIOS	219
ANEXO III. HISTORIAS DISEÑADAS PARA EL EXPERIMENTO CON EYE- TRACKING	228
ANEXO IV. MODELO ESTADÍSTICO COMPLETO EYE-TRACKING	234
ANEXO IV. ABSTRACT	238
ANEXO V. CONCLUSIONS	253

ÍNDICE DE TABLAS

Tabla 1. Muestra-tipo por cuotas propuesta para el proyecto PRESEEA (corpus 1995 y 2015)	63
Tabla 2. Presencia/ausencia de sujetos pronominales según el rasgo de animación del referente	75
Tabla 3. Presencia/ausencia de sujetos pronominales según la persona del sujeto	76
Tabla 4. Presencia/ausencia de sujetos pronominales según persona del sujeto	77
Tabla 5. Presencia/ausencia de sujetos pronominales según número	78
Tabla 6. Presencia/ausencia de sujetos pronominales según la especificidad	80
Tabla 7. Presencia/ausencia de sujetos pronominales según la persona del sujeto y la especificidad	81
Tabla 8. Presencia/ausencia de sujetos pronominales según la correferencialidad	84
Tabla 9. Presencia/ausencia de sujetos pronominales según la correferencialidad recodificada: continuidad y cambio referencial	85
Tabla 10. Presencia/ausencia de sujetos pronominales según el turno de habla	86
Tabla 11. Presencia/ausencia de sujetos pronominales según el tipo de discurso	88
Tabla 12. Presencia/ausencia de sujetos pronominales según el tipo de cláusula	91
Tabla 13. Presencia/ausencia de sujetos pronominales según el tipo de cláusula	91
Tabla 14. Presencia/ausencia de sujetos pronominales según la ambigüedad morfológica	93
Tabla 15. Presencia/ausencia de sujetos pronominales según la ambigüedad morfológica recodificada	95
Tabla 16. Presencia/ausencia de sujetos pronominales según el modo y la ambigüedad contextualizada	96
Tabla 17. Presencia/ausencia de sujetos pronominales según la persona del sujeto y la ambigüedad contextualizada	97
Tabla 18. Presencia/ausencia de sujetos pronominales según la ambigüedad fonética	98

Tabla 19. Presencia/ausencia de sujetos pronominales según la ambigüedad fonética contextualizada.....	100
Tabla 20. Presencia/ausencia de sujetos pronominales según la clasificación ADESSE	103
Tabla 21. Presencia/ausencia de sujetos pronominales según la clasificación de subcategorías ADESSE.....	104
Tabla 22. Presencia/ausencia de sujetos pronominales según el tiempo verbal desglosado	106
Tabla 23. Presencia/ausencia de sujetos pronominales según el tiempo verbal reagrupados	107
Tabla 24. Presencia/ausencia de sujetos pronominales según el modo.....	108
Tabla 25. Presencia/ausencia de sujetos pronominales según la perfectividad.....	109
Tabla 26. Presencia/ausencia de sujetos pronominales según la perífrasis verbal	110
Tabla 27. Presencia/ausencia de sujetos pronominales según la progresividad.....	111
Tabla 28. Presencia/ausencia de sujetos pronominales según la pronominalidad	111
Tabla 29. Presencia/ausencia de sujetos pronominales según el nivel de estudios.....	112
Tabla 30. Presencia/ausencia de sujetos pronominales según el sexo	113
Tabla 31. Presencia/ausencia de sujetos pronominales según la edad	117
Tabla 32. Explicación del modelo de regresión logística de efectos mixtos para el corpus de 1995	124
Tabla 33. Regresión logística efectos fijos corpus 1995	126
Tabla 34. Regresión logística efectos aleatorios (Infinitivo) corpus 1995.....	127
Tabla 35. Regresión logística efectos aleatorios (Informante) corpus 1995	128
Tabla 36. Explicación del modelo de regresión logística de efectos mixtos para el corpus de 2015	129
Tabla 37. Regresión logística efectos fijos corpus 2015	130
Tabla 38. Regresión logística efectos aleatorios (Infinitivo) corpus 2015.....	131
Tabla 39. Regresión logística efectos aleatorios (Informante) Corpus 2015	131

Tabla 40. Tasas de presencia del sujeto en distintas zonas de habla hispana.....	140
Tabla 41. Pesos probabilísticos y porcentaje de presencia de la 1ª y 2ª persona del singular	143
Tabla 42. Descripción de los ítems según la estructura y las variables lingüísticas	164
Tabla 43. Diseño experimental según cuadro latino	165
Tabla 44. Descripción y nomenclatura de las áreas de interés seleccionadas	167
Tabla 45. Descripción de la muestra analizada mediante <i>eye-tracking</i>	168
Tabla 46. Escala de magnitud de efectos	170
Tabla 47. Comparativa de C1, C2 y C3 según tiempo de lectura TRT en la OT	171
Tabla 48. Comparativa de C1, C2 y C3 según tiempo de lectura de TRT en la OC1 y OC2	173
Tabla 49. Comparativa de C4, C1, C2 y C3 según tiempo de lectura TRT en la OT	178
Tabla 50. Comparativa de C4, C1, C2 y C3 según el tiempo de lectura de TRT en la OC1 y OC2	179

INDICE DE GRÁFICOS

Gráfico 1. Recuento de casos en los corpus PRESEEA-Málaga	71
Gráfico 2. Distribución de casos por porcentaje de presencia	72
Gráfico 3. Presencia de sujetos pronominales según la correferencialidad y el turno de habla	87
Gráfico 4. Presencia de sujetos pronominales según la correferencialidad, el turno de habla y el tipo de discurso.....	90
Gráfico 5. Presencia/ausencia de sujetos pronominales según la ambigüedad fonética contextualizada y la especificidad	101
Gráfico 6. Presencia/ausencia de sujetos pronominales según el verbo y la persona del sujeto (porcentaje de presencia y frecuencia)	105
Gráfico 7. Presencia/ausencia de sujetos pronominales según el sexo y la ambigüedad morfológica contextualizada	114
Gráfico 8. Proporción de presencia/ausencia de sujetos pronominales según el nivel educativo y el sexo por individuo.....	116
Gráfico 9. Presencia/ausencia de sujetos pronominales según la edad, el sexo y el nivel de estudios.....	118
Gráfico 10. Importancia de variables lingüístico-sociales para el corpus 1995.....	120
Gráfico 11. Árbol de decisión para el corpus de 1995	121
Gráfico 12. Importancia de variables lingüístico-sociales para el corpus 2015.....	121
Gráfico 13. Árbol de decisión para el corpus de 2015	122
Gráfico 14. Mapa de distribución de trabajos dentro de PRESEEA	141
Gráfico 15. Variables significativas en trabajos previos ordenadas según importancia .	142
Gráfico 16. Patrón de uso del sujeto pronominal	146
Gráfico 17. Comparativa de C1, C2 y C3 según tiempos de lectura de FRT y RRT en la OT	172

Gráfico 18. Comparativa de C1, C2 y C3 según tiempos de lectura de FRT y RRT en la OC1 y OC2	174
Gráfico 19. Comparativa de C1 y C2 según tiempos de lectura de TRT, FRT y RRT en la OC1 y PV1	175
Gráfico 20. Comparativa de C2 y C3 según tiempos de lectura de TRT, FRT y RRT en la PV2.....	176
Gráfico 21. Comparativa de C2 y C3 según tiempos de lectura de TRT, FRT y RRT en la V2	177
Gráfico 22. Comparativa de C1, C2, C3 y C4 según tiempos de lectura de FRT y RRT en la OT.....	178
Gráfico 23. Comparativa de C4, C1, C2 y C3 tiempos de lectura de FRT y RRT en la OC1 y OC2	180
Gráfico 24. Comparativa de C3 y C4 según tiempos de lectura de TRT, FRT y RRT en la PV2.....	182
Gráfico 25. Comparativa de C2 y C4 según tiempos de lectura de TRT, FRT y RRT en la OC2 y PV2	183
Gráfico 26. Comparativa de las cuatro condiciones según el tiempo total de lectura del conjunto de los enunciados (OT).....	184
Gráfico 27. Ordenación de las cuatro condiciones según el tiempo total de lectura del conjunto de los enunciados (OT).....	187

INDICE DE FIGURAS

Figura 1. Tipos de movimientos oculares en lectura.....	161
Figura 2. Fijaciones para la primera lectura (FRT) y la relectura (RRT).....	162
Figura 3. Fijaciones para la lectura total (TRT)	163
Figura 4. Secuencia completa de la historia 1 con la condición 1.....	166
Figura 5. Áreas de interés mayores	167
Figura 6. Áreas de interés menores	167



UNIVERSIDAD
DE MÁLAGA

AGRADECIMIENTOS

El camino que me ha llevado a finalizar este trabajo ha estado lleno de personas con las que he tenido la gran suerte de coincidir y sin las cuales no podría haber llegado hasta aquí. En primer lugar, quisiera expresar mi agradecimiento más sincero a mis directores de tesis por compartir generosamente su experiencia y valiosas enseñanzas durante estos años. Al profesor Francisco Díaz Montesinos por su orientación desde las primeras fases de este trabajo, cuya perseverancia, rigor y sabios consejos me han guiado para encontrar siempre el mejor camino. A la profesora Matilde Vida Castro por transmitirme su positividad y calma frente a los retos surgidos, así como por su gran empatía ante mis nervios en cada presentación. Al Profesor Francisco Moreno Fernández por brindarme la oportunidad de realizar esta tesis en cotutela y por su cálida acogida en la Universidad de Heidelberg. Su carácter pragmático, su cercanía y su completa disponibilidad me han permitido comprobar que su excelencia humana está a la altura de su brillantez académica. A mi cuarto director, al Profesor Antonio Ávila Muñoz por creer en mí desde el principio. Ha sido mi brújula constante a lo largo de este camino. Le estaré siempre agradecida por acompañarme en los momentos buenos, y no tan buenos, en estos años de crecimiento académico, pero sobre todo personal.

Durante estos años he tenido el privilegio de poder formarme en dos universidades y conocer dos grupos de investigación extraordinarios. Por ello, quiero expresar mi profundo agradecimiento a los miembros del Grupo de Investigación Vernáculo Urbano Malagueño, donde descubrí la gran familia lingüística formada por el Profesor Juan Andrés Villena Ponsoda. Estos genes malagueños presentes en los miembros del grupo son un reflejo de su Maestro: una gran pasión por la sociolingüística y un firme compromiso académico sumado a una inconmensurable calidez humana. De manera especial, tengo que agradecerle al Profesor José María Sánchez Sáez por su contagioso entusiasmo por la estadística y su dedicación y amabilidad antes mis innumerables preguntas. De igual forma, agradezco enormemente al Heidelberg Center of Ibero-American Studies que me hayan hecho sentir como en casa y hayan convertido mi estancia en Alemania en un recuerdo imborrable: Katrin Berty, Héctor Álvarez Mella, Ana Gómez-Pavón, Magnus Mangold, David Chapman, María Manzanares y Mariel Elizondo. Debo agradecer a Martha Rudka por todo su apoyo en los procesos administrativos y por ofrecerme siempre su tiempo para orientarme en mi trabajo, y a Celia Hernández Pérez por sus infinitos audios que me fueron de gran ayuda a lo largo de todo el proceso experimental.

El experimento de *eye-tracking* ha sido posible gracias a la colaboración con el laboratorio StabLab del Institut für Statistik der Ludwig-Maximilians-Universität München, por su apoyo en el análisis estadístico de los datos, y con el laboratorio Leeduca de la Facultad de Psicología de la Universidad de Málaga, por abrirnos sus instalaciones para el desarrollo del experimento. Agradezco también a todas las personas que participaron voluntariamente en las diferentes pruebas.

En lo personal, quiero darle las gracias a mis amigos malagueños y almerienses, en especial a Celia Navío y a Rosario Mora, por su cariño y comprensión durante todos estos años. Quiero acordarme también de dos personas que fueron mis referentes: a María del Mar Ortiz Ibáñez por transmitirme su pasión por la lengua española y la sintaxis, y a Carmen Sánchez García por acompañarme e impulsarme a mejorar desde niña.

No hubiera llegado hasta aquí sin mis hermanas académicas: María Cruces Rodríguez y Jana Weckesser. A María, mi gemela de blusita blanca, debo agradecerle profundamente su completa lealtad y honestidad, así como su manera única para entenderme mejor que nadie. A Jana le estaré infinitamente agradecida por cuidarme tanto, por todas las mudanzas, por las tardes de Kaffee y Kuchen y por ser casa, incluso lejos de ella. Gracias a las dos por estar siempre conmigo, por compartir miedos y alegrías. No podría haber tenido mejores compañeras de viaje.

Mi agradecimiento más sincero a Álvaro Ramajo Ballester, mi gran apoyo, por su paciencia inagotable, por saberse de memoria mis presentaciones, por todo lo que ha hecho por mi tesis, pero, especialmente, por todo lo que hace por mí.

A Laura Segura Lores, una futura gran psicóloga, por darme siempre su opinión sincera, alegría y fortaleza en todo momento. Por último, a mis padres, por darme la libertad de elegir mi propio camino y apoyarme incondicionalmente hasta conseguirlo. A Consuelo Lores Ruiz por su corazón puro, por escuchar durante horas mis dudas e inseguridades y por ayudarme (y sufrir) con mis viajes, mudanzas y maletas. A Manuel Segura Gálvez por enseñarme la importancia del esfuerzo y la exigencia que supone el trabajo bien hecho, por ser un ejemplo de perseverancia y motivación por seguir aprendiendo y mejorando día a día.

Heidelberg, 12 de julio de 2025



UNIVERSIDAD
DE MÁLAGA

INTRODUCCIÓN

Observar la lengua como componente esencial de la comunicación humana implica considerar el estudio de dos procesos íntimamente relacionados: la producción, asociada al emisor que se encarga de construir un mensaje, y la percepción, vinculada al destinatario, quien debe descodificar el mensaje mediante la activación de sus capacidades analíticas y evaluativas. Si bien es fundamental el estudio de la producción lingüística —qué reglas de producción se activan en ciertos contextos y con qué frecuencia—, no debería descuidarse el análisis de la percepción entendida como uno de los mecanismos de cognición que permite la conexión entre el mundo interno y externo a través de los sentidos. Estudiar la lengua, más concretamente los procesos de variación, sin poner la atención en el carácter social implícito y en el proceso cognitivo que se desencadena en la mente humana, sería realizar un análisis parcial y sesgado de un fenómeno amplio y complejo.

En esta investigación, el estudio del carácter social del lenguaje no se limita a su asociación con variables que permitan estratificar la muestra analizada, ni se aborda como un elemento accesorio al hecho lingüístico, sino que se ajusta a una perspectiva integradora de base social que coincide con la óptica propuesta por Moreno Fernández (2012) y Caravedo (2014). Este enfoque parte de la premisa de una imbricación total entre la lengua

y la sociedad: no existe lengua natural sin dimensión social, ni estructura lingüística autónoma sin una implicación social. De esta manera, se supera la tradicional disyuntiva que considera lo estructuralmente lingüístico como interno y lo social como externo, de modo que ambos aspectos son complementarios e inseparables si se pretende adoptar un enfoque comprensivo y cohesivo del fenómeno lingüístico.

Esta Tesis Doctoral se vertebra sobre los fundamentos teóricos de la sociolingüística cognitiva interesada por la producción variable de las manifestaciones lingüísticas y, a su vez, por la percepción que los hablantes tienen sobre la variabilidad (Moreno Fernández, 2012, p. 89). De modo que nuestro estudio se centra en la variación gramatical en español, concretamente en la alternancia de la ausencia y presencia del pronombre personal sujeto desde la sociolingüística cognitiva. Este planteamiento estructura nuestro trabajo en dos bloques principales: los usos del sujeto pronominal y la percepción cognitiva de la variación ausencia-presencia del pronombre personal sujeto.

Este trabajo sigue el continuo interactivo de la filosofía cuantitativa-cualitativa de Newman y Benz (1998): partimos de una premisa que considera opcional la expresión del sujeto pronominal en español, la cual nos conduce a realizar una revisión bibliográfica de los estudios previos para contextualizar el fenómeno lingüístico. A partir de estas lecturas, surgen hipótesis sobre los posibles condicionantes de la expresión del sujeto y estas son verificadas mediante el análisis de los datos —previamente recogidos y codificados— y la interpretación de los resultados. Extraídas las primeras conclusiones sobre los patrones de uso, la investigación se expande hacia el ámbito perceptivo para examinar cómo determinadas estructuras son procesadas cognitivamente por los individuos.

De esta forma, la primera parte del trabajo, dedicada a la producción, aborda el uso del pronombre personal sujeto en el español hablado en la ciudad de Málaga recogido en dos corpus PRESEEA, correspondientes a dos cortes temporales (1995 y 2015). En ellos se analizan múltiples factores tanto lingüísticos —fonéticos, morfológicos, léxicos, pragmáticos, discursivos y semánticos— como sociales que permitan explicar la manifestación del sujeto pronominal (Capítulo II). Una vez analizada la producción, la segunda fase del trabajo, orientada hacia la exploración perceptiva, comprende un estudio mediante la técnica de *eye-tracking*. Con el fin de proporcionar evidencia experimental, se observa el patrón de procesamiento cognitivo en relación con la variación ausencia/presencia del pronombre personal sujeto (Capítulo III).

Estos dos objetivos generales —el análisis de la producción del uso del sujeto y el estudio de la percepción cognitiva— determinan la estructura de este trabajo organizada en cuatro capítulos, cuyo contenido se especifica a continuación.

El primer capítulo presenta el objeto de estudio y sitúa conceptualmente la teoría que sostiene el conjunto de la investigación: el variacionismo y la sociolingüística cognitiva. Presentamos los antecedentes tanto de los estudios de producción como de los trabajos sobre la percepción; y, por último, se generan las preguntas e hipótesis de investigación asociadas a cada uno de los objetivos generales planteados.

El segundo capítulo muestra el estudio de la producción lingüística, en particular, la alternancia de la expresión del sujeto pronominal en el español de la ciudad de Málaga. Para ello, nos basamos en los datos de los corpus PRESEEA-Málaga correspondientes a dos sincronías. En primer lugar, se describe la metodología de análisis, la selección de casos y los criterios empleados durante la fase de codificación. A continuación, se revelan los resultados obtenidos en cada uno de los cortes temporales analizados y, posteriormente, se establece una comparación diacrónica dentro de la ciudad de Málaga y diatópica con los resultados obtenidos en trabajos de otras ciudades. Este enfoque nos ayuda a identificar posibles cambios a lo largo del tiempo, así como a trazar un posible patrón de uso del pronombre personal sujeto en la geografía hispanohablante.

El tercer capítulo se centra en la percepción cognitiva de determinadas estructuras lingüísticas donde la ausencia-presencia del sujeto pronominal puede establecer relaciones de cambio o continuidad referencial e incluso puede ocasionar situaciones de ambigüedad potencial debido a la morfología flexiva del verbo. A través de la metodología experimental de *eye-tracking*, se analiza el procesamiento cognitivo que el hablante lleva a cabo al enfrentarse a dichos estímulos, con el fin de comprobar qué estructuras exigen mayores esfuerzos cognitivos para lograr su comprensión.

Por último, las conclusiones se presentan en torno a los dos bloques centrales de este trabajo: la producción y la percepción. Además, se exponen las limitaciones y líneas futuras de investigación y, finalmente, se ofrece una reflexión general que sintetiza la contribución de este estudio al entendimiento de la variación del sujeto pronominal en español.



UNIVERSIDAD
DE MÁLAGA

CAPÍTULO 1. VARIACIÓN DEL SUJETO PRONOMINAL: MARCO TEÓRICO Y ANTECEDENTES

1.1. Objeto de estudio: la alternancia ausencia-presencia del sujeto pronominal en español

A diferencia de lenguas como el inglés, el francés o el alemán, en algunas lenguas románicas, como el español o el italiano, existe la posibilidad de que determinados enunciados presenten el sujeto de manera expresa debido a la naturaleza flexiva de las formas verbales. De tal forma, los sujetos tácitos tienen propiedades pronominales y concuerdan en persona y número con el verbo, y en género y número con otros elementos del enunciado (Fernández Soriano, 1999; RAE & ASELE, 2009; etc.).

- (1) *Claro, ¿no viste tú/Ø la película? [MA_M31_025].*
- (2) *Nos asomábamos nosotras/Ø por el balcón y nosotras/Ø
veíamos el espectáculo [MA_M21_042].*

Desde las gramáticas tradicionales se ha considerado que la opción habitual sería la omisión del pronombre y que la presencia de este podría deberse principalmente a tres

valores: el uso redundante, el carácter enfático y la capacidad desambiguadora. Son muchas las afirmaciones que siguen esta línea. Salvá (1988) considera que el énfasis es necesario para que los pronombres se expresen en español y de esta forma se pueda llamar la atención hacia ellos. *El Esbozo* (1973) justifica la presencia del pronombre ante situaciones donde es posible aclarar una referencia ambigua o se pretende enfatizar un elemento respecto a otro; de lo contrario, “las desinencias personales de la conjugación española son tan claras y vivaces que casi siempre hacen innecesario y redundante el empleo del pronombre” (*El Esbozo*, 1973, p. 421).

Sin embargo, la naturaleza del énfasis tampoco parece ser un recurso claramente definido; incluso Gili Gaya (1961) señala que este puede estar motivado por cuestiones muy variadas y tiene unos límites borrosos y poco perceptibles. Por otra parte, el contexto discursivo pone de manifiesto que la presencia del pronombre personal también puede obedecer a otras condiciones, como la intención comunicativa y discursiva, que van más allá de los valores tradicionales de desambiguación y énfasis (Castellano Alemán, 2001). Por lo tanto, los tres valores señalados anteriormente no parecen explicar con completa claridad la alternancia de la presencia y ausencia del sujeto pronominal en español. De modo que la presencia de este tipo de constituyente en la lengua española no se justificaría categóricamente bajo las implicaciones de la redundancia o el énfasis en todos los casos, ni su ausencia conllevaría siempre a la ambigüedad (Devís Márquez, 2012).

En contraposición a estos argumentos, Fernández Soriano (1999, p. 1227) considera que la aparición del sujeto pronominal es una variable libre, cuya presencia responde más a razones estilísticas que gramaticales. No obstante, esta afirmación es negada con rotundidad por Bentivoglio (1987, p. 66) y Silva Corvalán (2003). En este sentido, el desarrollo posterior de los estudios variacionistas ha permitido esclarecer que —excluidos los contextos donde no es posible la alternancia— la expresión del sujeto se correlaciona probabilísticamente con una compleja red de factores pragmáticos, semánticos, sintácticos y morfológicos, como la correferencialidad, la topicalidad y la ambigüedad de la forma verbal, aunque estos factores no condicionen la variación de manera categórica.

Conviene insistir en que la metodología variacionista más actual considera con más cautela los contextos en los que la aparición del sujeto no es posible y también otros en los que se exige obligatoriamente la presencia de este, excluyéndolos de sus análisis. En esta

línea, la NGLE (33.4 y 33.5) recopila los casos de presencia y ausencia obligatoria que se enumeran a continuación:

- Contextos de ausencia obligatoria.

En español encontramos contextos en los que la ausencia del pronombre es obligatoria:

- Oraciones impersonales con *se* o impersonales reflejas.

Son oraciones que presentan marcas de 3ª persona del singular, cuya interpretación puede ser genérica al referirse a la gente en general o a todo el mundo, por ejemplo: *Aquí se habla inglés*. En otras ocasiones la interpretación no es genérica, sino indefinida, donde no se menciona a la persona sobre la que recae la acción, por ejemplo: *Se trabajó intensamente en el proyecto* (NGLE , 33.4.q).

- Verbos en presente de indicativo en titulares de prensa.

Aunque estos casos no van a aparecer en el estudio que aquí se desarrolla, existe en español un uso en el que el presente de indicativo con sujeto omitido no equivale al pronombre personal, sino a un sujeto nulo, por ejemplo: *Asesina a inocente y lo culpa* (Nuevo Herald, ejemplo extraído de NGLE , 33.4.r).

- Oraciones en 3ª persona del plural con sujeto omitido.

Cuando el sujeto designa individuos y permite una interpretación indefinida, se generan los sujetos nulos típicos de las oraciones impersonales en las que se atribuye una acción o propiedad a personas no identificadas: *Llaman a la puerta*. Pese a los rasgos de plural presentes en el sujeto y reflejados en la conjugación verbal, no necesariamente se hace referencia a un grupo específico de personas. En otras ocasiones, la interpretación de la omisión del sujeto en 3ª persona de plural puede ser más genérica, lo que sugiere que se trata de una acción realizada por cualquier persona, sin especificar quién la lleva a cabo, por ejemplo: *Hablan alemán*.

- Impersonales léxicos.

Algunos gramáticos sostienen que en este tipo de oraciones propias de los verbos que describen fenómenos meteorológicos existe un sujeto nulo o tácito. Este se

representa por un elemento pronominal vacío y carece de interpretación argumental, similar a los pronombres expletivos o pleonásticos en otras lenguas, por ejemplo: *Llueve mucho hoy, pero mañana hace calor*.

- Contextos de presencia obligatoria.

A pesar de la riqueza flexiva del verbo, existen ciertos casos en los que el sujeto debe estar presente en español. Devís Márquez (2012) atribuye la presencia del pronombre a cuestiones exclusivamente discursivas o gramaticales.

- Interpretación focal.

A diferencia de los sujetos omitidos, los sujetos expresos pueden poseer un acento contrastivo y desempeñar una función informativa con el propósito de introducir información nueva o contrastiva, como ejemplo: *¿Quién llamó? Llamé yo o Llamó él, y su hermana no*. Si en los ejemplos anteriores se omitiese el sujeto, las oraciones serían agramaticales; por lo tanto, la presencia del pronombre es obligatoria.

- Complementos apositivos cuantificadores, adjetivales u oracionales (Fernández Soriano, 1999, p. 1226).

El sujeto debe manifestarse de forma expresa cuando va acompañado de ciertos elementos como *solo*, *mismo* u oraciones. Por ejemplo: *Tú solo lo hiciste, tú mismo lo has resuelto* o *tú, que tienes dinero, podrás venir*. En estos contextos en los que el sujeto es obligatorio, puede tener como referente un objeto (rasgo no humano) : *Tus observaciones son todas ellas falsas o no es necesario escayolar este hueso. Se soldará él solo/mismo*.

- Estructuras que con valor de contraste o de individualización

Fernández Soriano (1999, p. 1227) indica que existen en español ciertas estructuras donde la interpretación del sujeto conlleva un valor de contraste o individualización. En este caso, oraciones como *donde él quiera* o *como tú digas* favorecerían la explicitación de un sujeto pleno.

1.2. Marco teórico

1.2.1. La sociolingüística variacionista

La lengua es, por su naturaleza, un sistema intrínsecamente variable que refleja su variabilidad en el uso. Los hablantes, al interactuar, seleccionan diversos recursos lingüísticos para expresar distintos significados, o también pueden emplear formas alternativas para referirse a un mismo concepto (Moreno Fernández, 2009, p. 21). Esta variación puede estar motivada, en mayor o menor medida, por factores lingüísticos y/o sociales, pero resulta indiscutible la conexión social inherente a cualquier estudio lingüístico, dado que sin esta dimensión social la lengua no podría cumplir su principal función.

En el ámbito de la sociolingüística, el elemento ideal de estudio es la lengua real en su uso, dentro de contextos sociales específicos y en situaciones concretas (Milroy, 1992; Moreno Fernández, 2009). Los hablantes pueden emplear una o varias formas para decir lo mismo, lo que se conoce como *variable lingüística* —más adelante profundizaremos en este concepto dentro del plano gramatical—, y la clave no es otra que encontrar el patrón de uso subyacente que explique el aparente caos de la variación. En los casos en los que la variable lingüística presenta variantes asociadas a un significado social, hablaríamos de *variable sociolingüística*, ya que implica una correlación entre un factor lingüístico y alguno de corte social relacionado con el hablante, la situación comunicativa, la audiencia, etc. (Labov, 1972, p. 237).

La atención sobre el significado social en la variación es una constante en la disciplina sociolingüística desde sus orígenes. Sin embargo, su tratamiento ha evolucionado a lo largo del tiempo y ha dado lugar a diferentes enfoques metodológicos y analíticos que pueden agruparse bajo la metáfora definida por Eckert (2012) como las Tres Olas de la sociolingüística. Antes de profundizar en estas Tres Olas, es conveniente subrayar que no deben entenderse como una mera sucesión cronológica, puesto que, en este sentido, el estudio pionero de Labov de Martha's Vineyard presenta ya elementos propios de la Primera y Segunda Ola.

De forma indiscutible, en la denominada Primera Ola, los estudios de Labov (1963; 1972) realizados en Martha's Vineyard y en Nueva York son un precedente en la

observación del significado social. En estos trabajos se pone de manifiesto cómo los hablantes son capaces de asignar un valor social a las variables lingüísticas con el propósito de aproximarse a una identidad o variedad —más vernácula o estándar—, y a raíz de la cual surgen preguntas relativas al arraigo lingüístico y al estigma social en la variación.

Sin embargo, los estudios desarrollados posteriormente en este periodo se centran en la relación entre el uso lingüístico y las macrocategorías capaces de clasificar el espacio social en grandes grupos según la edad, el sexo, la clase social o la etnia de los hablantes. A pesar de que estos estudios establecen un patrón regular aplicable a diferentes contextos urbanos, esta visión macroscópica de la sociedad no resulta suficiente para explicar la complejidad de la variación lingüística en su totalidad.

Con la llegada de la Segunda Ola denominada por Eckert, la atención se desplaza a las microvariables, a la reconstrucción etnográfica del individuo y al interés por la variación a nivel local. Por lo tanto, este segundo enfoque social se centra en definir grupos a partir de sus características sociales de pequeña y mediana escala, como la ocupación, el nivel económico familiar, el estilo de vida, su entorno social e, incluso, sus intereses y aspiraciones.

Otro gran hito de esta Segunda Ola son los estudios de redes, como el llevado a cabo por Milroy en la ciudad de Belfast, que demuestra que la densidad y la multiplicidad de las redes personales del sujeto influyen a la hora de adoptar una u otra variedad. Asimismo, Eckert (1989) realiza un estudio de dos grupos de adolescentes en Detroit —los Burnouts y los Jocks— que evidencia la importancia de las dinámicas locales en la variación lingüística. En el ámbito hispánico, destaca el estudio de Blanco Canales (1998) sobre una red de Alcalá de Henares y, en la ciudad de Málaga, los trabajos llevados a cabo por el grupo de investigación Vernáculo Urbano Malagueño (VUM). Estos últimos, además de estudiar la convergencia o divergencia de los malagueños hacia el estándar, observan la influencia de las redes sociales del individuo y la caracterización de grupos sociales diferenciados presentes en la ciudad (Ávila Muñoz, 1998; Ávila Muñoz, 2017; Villena Ponsoda, 2014; Villena Ponsoda et al., 2017; Villena Ponsoda & Vida Castro, 2017).

Hasta la fecha, la mayor parte de los estudios variacionistas se enmarcan entre la Primera y Segunda Ola. La Tercera Ola supone un avance hacia una aproximación más flexible de la variación, que no pone tanto el foco en la estructura social, sino en el

significado social de la identidad, contemplada como una actividad dinámica y constante donde los individuos configuran su rol. Esta Tercera Ola podría articularse a partir de dos conceptos clave: la indexicalidad, donde la variación se construye localmente, pero se sostiene a partir de cuestiones ideológicas que vinculan las interacciones cotidianas con la economía y la ideología (Eckert, 2012); y el *habitus* (Bourdieu, 1977) definido como el conjunto estable de disposiciones (patrones de comportamiento, afecto y pensamiento), que puede un individuo manifestar a lo largo de su vida en función de su posición social. Este último elemento actúa como un mediador que modela la acción individual en la inserción entre la estructura social y la conducta personal.

Las distintas Olas de variación sociolingüística descritas por Eckert (2012) no deben entenderse como corrientes reemplazables de manera sucesiva, sino como enfoques complementarios que enriquecen la comprensión de la relación compleja entre lenguaje y sociedad. Además, la autora aboga por una revisión del concepto de variable *lingüística*, el cual hasta el momento se había usado de manera análoga a la variación en el plano fonético-fonológico. Sin embargo, dado que los factores sociales tienen influencia en los cambios de significado, la variable lingüística podría asemejarse a una variación más compleja, similar a la morfosintáctica ya propuesta por Lavandera (1978). Por ello, resulta necesario llevar a cabo una revisión conceptual de la variable lingüística desde la perspectiva morfosintáctica.

1.2.1.1. La variación gramatical

Conviene subrayar que los primeros estudios de corte variacionista se centran en el plano fonético-fonológico, cuyas características favorecen una clara identificación de la unidad categorial y de las variantes estudiadas. En cambio, al aplicar esta metodología al nivel morfosintáctico, aparecen múltiples problemas que han sido ampliamente discutidos (García, 1985; Lavandera, 1978; Romaine, 1981; Sankoff, 1988).

Las discrepancias entre el plano fonético-fonológico y el gramatical son evidentes. Silva-Corvalán (2001, p. 129-132) afirma que no se puede equiparar la variación fonológica con la sintáctica, puesto que en la lengua existe menos variación sintáctica que fonológica. Además, la variación sintáctica es más difícil de medir y cuantificar, debido a que los contextos en los que ocurre son menos frecuentes y más difíciles de delimitar.

Aunque sea fácil determinar si existen o no diferencias de significado entre las supuestas variantes léxicas, quizás el desafío semántico en el nivel sintáctico sea mayor. El obstáculo semántico ha sido el más criticado en la variación sintáctica, ya que, de acuerdo con la idea original de Labov (1972, p. 271), las variantes son idénticas en cuanto a su valor de verdad, pero difieren únicamente en su valor social y/o estilístico. A pesar de que en el nivel fonético-fonológico pueda estudiarse sin inconvenientes aparentes, en el nivel léxico-semántico intervienen una complejidad de factores que hacen que la equivalencia de variantes, en términos estrictos o la igualdad en cuanto a condiciones de verdad, resulte dudosa (Serrano, 2007, p. 103).

Para acercar posturas, Silva-Corvalán (2001) propone una opción conciliadora con el fin de definir la variable sintáctica y abordar el problema de la sinonimia. Considera que cuando no hay variación en el significado, la definición coincidiría con la propuesta por Labov: “dos o más realizaciones de un elemento común; o dos o más formas de decir la misma cosa”. A esta definición responderían los casos donde el significado fuera el mismo y solo variara la estructura sintáctica en cuestión. En consonancia con lo planteado por Silva-Corvalán (2001) y López Morales (2009), Moreno Fernández (2009, p. 121) establece una distinción entre dos tipos de sinonimia o equivalencia: la lógica y la pragmática.

- La equivalencia lógica, también conocida como de verdad o referencial, se produce cuando dos variables tienen la misma referencia, aunque existan disparidades entre ellas. Por ejemplo: *Si el profesor falta a clase, los alumnos se alegran/alegrarán.*
- La equivalencia pragmática se refiere a los casos en los que a pesar de que las diferencias léxicas y sintácticas sean muy importantes, quedan neutralizadas por el discurso y la intención comunicativa es la misma en todas las variantes. Por ejemplo: *Está oscuro, ¿puedes leer con tan poca luz?, enciende la luz, ¿podrías encender la luz?*

Aunque es innegable que observar, analizar y establecer el grado de identidad entre variantes gramaticales es una tarea sumamente compleja, es precisamente esta complejidad la que se propone como objeto de estudio en el campo de la variación sintáctica. Como señala Silva-Corvalán (2001, p. 129-131), las dificultades derivadas del análisis de los elementos portadores de significado han supuesto un desafío al conectar de manera más estrecha la sociolingüística con la semántica y la pragmática del discurso. Además, los dos

tipos de equivalencias mencionadas podrían vincularse a su vez con variables sociales, de modo que sería posible indagar en los diferentes usos sociales y estilísticos que emanan de la lengua. De esta forma, la posibilidad de interpretar una covariación con factores sociales y estilísticos se convierte en una tarea delicada, cuando se comprueba que las variantes conllevan diferencias de significado.

Todo este proceso gesta una concepción más amplia y flexible de la variación lingüística, en la cual se considera la intersección de factores sintácticos, discursivos, pragmáticos, e incluso fonéticos y fonológicos, en lugar de circunscribir el análisis a una visión simplificada basada en la identidad exacta de las variantes para determinar sus correlaciones sociales (Serrano, 2007, p. 109). Todo ello, lejos de frenar el avance de las investigaciones, las críticas hacia este tipo de estudios han contribuido a consolidar el marco variacionista en el análisis de la dimensión sintáctica.

Otro de los aspectos más discutidos sobre la variación sintáctica es su carácter social. Numerosos trabajos demuestran la mínima influencia de factores sociales externos sobre la elección de una u otra variable; en cambio, los factores lingüísticos internos parecen ser los grandes condicionantes de la variación y los que juegan un papel crucial en los estudios semánticos y pragmáticos. Romaine (1981, p. 17) señala que, para comprender los factores internos y externos que afectan a la gramática, es necesario adoptar un enfoque pragmático y también semántico, que vincule el significado con su función. En esta línea misma, Martín Butragueño (1994) sostiene que el primer paso sería analizar la función de las variantes dentro del contexto discursivo y, de manera más ambiciosa, subraya la importancia de incorporar un componente cognitivo dentro del enfoque variacionista, ya que podría ser de gran utilidad para, mediante un análisis más profundo, observar las diferencias de significado.

Gracias a enfoques más actuales, como el propuesto por la sociolingüística cognitiva (Moreno Fernández, 2012), los límites de estudio se amplían hacia una visión más integradora, cuyos pilares fundamentales son los recursos cognoscitivos implicados en el discurso, el uso lingüístico contextualizado y la percepción social. Este planteamiento ofrece un marco provechoso para abordar la complejidad de la variación, especialmente la sintáctica, al evidenciar que los factores semánticos, pragmáticos y cognitivos convergen con aquellos de orden social y estilístico en la observación de la compleja red de correlaciones presentes en el uso lingüístico.

1.2.2. La sociolingüística cognitiva

Antes de profundizar en el concepto de *sociolingüística cognitiva* y los fundamentos que sustentan este trabajo, resulta pertinente contextualizar tres elementos esenciales sobre la naturaleza de las lenguas: el lenguaje, la lengua y el uso lingüístico. El lenguaje se concibe como la facultad humana que se manifiesta en diversas variedades lingüísticas y que, ejercida de forma colectiva, tiene su origen y configuración en la vida social con fines comunicativos. La lengua puede considerarse un fenómeno histórico resultante de la interacción recíproca entre procesos de pensamiento, habilidades perceptivo-motoras, capacidades cognoscitivas y factores sociopragmáticos. En este marco, los usos lingüísticos se conciben como realidades esencialmente variables, producidas y percibidas por los individuos. Estos usos lingüísticos no solo reflejan la forma de las lenguas, sino que también contribuyen a su configuración, y tienen lugar en contextos específicos de la realidad social, también entendidos como modelos cognitivos de interacción verbal.

Por ello, las lenguas, consideradas como sistemas adaptativos complejos (Beckner et al., 2009), se caracterizan por la intervención de múltiples agentes, el uso acumulativo de las interacciones, la concurrencia de factores en la conducta del hablante y el surgimiento de patrones de interacción, de experiencia y de cognición. La cooperación comunicativa entre los individuos convierte a la lengua en un fenómeno emergente con una doble dimensión: una parte social crecientemente compleja compartida por los interlocutores y, a su vez, otra más individualizada basada en factores de naturaleza cognitiva, psicomotriz, perceptiva y experiencial, donde las variantes preferidas por el uso lingüístico acaban integrándose gradualmente en la mente individual. En consecuencia, si la intención del investigador es observar la lengua atendiendo a la complejidad y el dinamismo que esta entraña, es necesario recurrir a constructos teóricos que permitan una concepción multidimensional e integradora de los componentes lingüísticos, sociales y cognitivos.

A la luz de todo lo anterior, la sociolingüística cognitiva se concibe como un modelo metateórico que pretende analizar relacionamente la lengua en su entorno natural, cultural, social y situacional, cuyo fin es abordar lo lingüístico y lo social como realidades analizadas desde la cognición humana. Ciertamente es que, al incorporar el componente cognitivo, podría pensarse que esta teoría continúa el enfoque presentado principalmente

por Geeraerts y Kristiansen (Geeraerts, 2005; Kristiansen, 2001, 2003, 2008; Kristiansen & Dirven, 2008). Sin embargo, aunque en parte puede ser así, la perspectiva adoptada por Moreno Fernández (2012), y en la que se basa este trabajo, pretende ser más integradora. En lugar de detenerse en ámbitos sociológicos muy específicos, busca proporcionar una visión panorámica y multidimensional de lo cognitivo en la lengua y de su uso social en todos los planos. Por lo tanto, la sociolingüística cognitiva se centra en el estudio de los recursos cognoscitivos implicados en el procesamiento y en el uso contextualizado del lenguaje. Esta perspectiva pone especial énfasis en el conocimiento que los hablantes tienen de la lengua en su uso social, donde se consideran aspectos como los entornos comunicativos, los procesos de interacción, la variación y el cambio lingüístico, y cómo estos son percibidos por los hablantes.

Desde esta óptica, el estudio de la variación de la lengua no se concibe como un fenómeno objetivo y autónomo, sino como un proceso que opera en una doble dimensión: la variación se exterioriza en la producción y adquiere su significado intrínseco en la percepción. En palabras de Caravedo (2018), limitar el análisis de la variación exclusivamente a la producción, sin tener en consideración la percepción, sería incurrir en un error. Para suplir esta limitación, la sociolingüística cognitiva aborda tal multidimensionalidad interesándose por la producción manifestada en el uso lingüístico, pero también por la percepción que los propios hablantes tienen sobre la variación (Moreno Fernández, 2012, p. 89). En este marco, resulta fundamental precisar qué entendemos por producción y percepción.

La producción consiste en el *output*, es decir, las manifestaciones lingüísticas concretas que emite el hablante en la interacción conversacional y que dan lugar a los actos de habla donde pueden adscribirse todos los demás usos sociolingüísticos (Moreno Fernández, 2012, p. 77). La variación sociolingüística aborda principalmente el estudio organizado, objetivo y sistemático de la producción de los hablantes (Caravedo, 2014, p. 36), cuyo análisis puede basarse, por ejemplo, en la exploración de corpus orales. El estudio de la producción resulta fundamental, ya que la sociolingüística cognitiva exige la observación sistemática al analizar la lengua basándose en su uso y, por lo tanto, los análisis cuantitativos y cualitativos son esenciales para conocer cómo se produce y cómo se establecen reglas generales de predicción en el uso (Moreno Fernández, 2012, p. 166-171).

La percepción, de acuerdo con la definición de Caravedo (2018, p. 70-71), constituye uno de los componentes principales que intervienen en el desarrollo de la cognición: actúa como un mecanismo básico para acercarse al conocimiento de una lengua y resulta esencial en la actividad cognitiva. La percepción permite poner en juego los sentidos con el fin de (re)conocer el mundo externo; es decir, es un puente que conecta el interior con el exterior. La percepción está ligada a la sensorialidad —visual, auditiva, táctil y olfativa—, y a través de las diferentes fuentes, especialmente mediante la auditiva y la visual, se recibe información externa necesaria. Por lo tanto, se entiende por percepción la recepción del material lingüístico producido por el hablante y su consiguiente decodificación y procesamiento por parte del oyente (*input*).

En cuanto a los aspectos metodológicos, Moreno Fernández (2012, p. 166) señala que la sociolingüística cognitiva requiere la integración de análisis cuantitativos y cualitativos, así como la aplicación de procedimientos experimentales. No obstante, se advierte de que este enfoque no está ceñido a propuestas metodológicas específicas, sino que evalúa las alternativas metodológicas existentes y considera la aplicación de aquellas que resulten más pertinentes en función del fenómeno lingüístico tratado.

La flexibilidad metodológica y el amplio paradigma de estudio de la sociolingüística cognitiva produce una convergencia natural con los principios teóricos propuestos por la Tercera Ola de la sociolingüística. En este sentido, aunque Pizarro Pedraza (2016) esboza un posible punto de encuentro entre ambas corrientes, su análisis no contempla la óptica adoptada por Moreno Fernández. Ambos enfoques coinciden en dos aspectos cruciales: la importancia del contexto y la influencia del conocimiento. Por un lado, el contexto es un elemento crucial para sustentar el análisis del significado social y lingüístico; por otro, el conocimiento no solo lingüístico, sino también cultural, social e histórico forma parte del proceso de reconstrucción de significados por parte de los individuos.

En relación con la metodología, Pizarro Pedraza (2016) sostiene que la sociolingüística cognitiva basa la recogida de datos mayoritariamente en corpus lo que, en el caso de corpus recogidos *ad hoc*, implicaría un elevado coste en términos de tiempo. Para mitigar esta limitación, propone la planificación de hipótesis sociolingüísticas *a priori*. En cambio, la concepción de la sociolingüística cognitiva abordada aquí sigue el principio de la filosofía de Newman y Benz (1998), cuyo ciclo de investigación se basa en

la formulación de hipótesis nuevas a partir de los resultados obtenidos. Además, al no adscribirse a un paradigma metodológico específico, promueve la adopción de los procedimientos analíticos necesarios para la consecución de los objetivos de investigación establecidos. De este modo, la propuesta de una sociolingüística cognitiva de la Tercera Ola ya se encontraría recogida en el marco más amplio, integrador y multidimensional de la sociolingüística cognitiva propuesta por Moreno Fernández.

1.2.2.1. Sociogramática

Desde la perspectiva de la sociolingüística cognitiva, la gramática sociocognitiva se fundamenta en los principios de la lingüística cognitiva y la gramática de usos. En primer lugar, comparte con la lingüística cognitiva (Langacker, 1987) la concepción de la lengua como un sistema en el que todos los niveles —fonético, gramatical, léxico y discursivo— se interrelacionan entre sí dentro del marco comunicativo. En este contexto, el uso lingüístico contextualizado y dinámico influye en la configuración interna de la lengua.

En segundo lugar, la gramática sociocognitiva incorpora la perspectiva teórica de la gramática basada en el uso (Bybee, 2001, 2010), la cual postula que la organización cognitiva del sistema lingüístico se construye a partir de la experiencia acumulada. De esta forma, las representaciones cognitivas que subyacen al uso se elaboran a través de la categorización de expresiones y la construcción de modelos a partir de sus formas lingüísticas, su significado y el contexto en el que han aparecido. En consecuencia, las categorizaciones y construcciones que se producen en el uso de la lengua son el resultado de procesos emergentes que dependen de interacciones contextualizadas e implican elementos morfosintácticos, pero también, fónicos y léxico-semánticos.

Una gramática sociocognitiva se constituye a partir de la interacción y se configura desde la convención social. Por lo tanto, la gramática entendida de esta forma no sería un conjunto de reglas estancas ni una estructura formal que se refleja en los usos lingüísticos, más bien una red de representaciones tejida con muestras categorizadas de usos fijados convencionalmente en construcciones de diversa complejidad. En definitiva, un sistema adaptativo complejo. Además, la gramática sociocognitiva debe incluir dos principios fundamentales: el principio experiencial, que hace necesaria la experiencia del hablante (asociaciones y percepciones) para la existencia de las construcciones, y el principio de

preeminencia, que explica cómo las diferencias de percepción están ligadas a diferencias gramaticales.

Otro aspecto que cabe destacar para el estudio de la gramática sociocognitiva es la frecuencia de uso. En el marco de la configuración gramatical y de la gramática basada en el uso, el registro de las probabilidades de ocurrencias y de las coocurrencias de las formas lingüísticas resulta esencial para comprender los procesos de estructuración gramatical. El dominio de una lengua no radica exclusivamente en el conocimiento de su sistema gramatical, sino también implica la capacidad de registrar las probabilidades de ocurrencia de las formas lingüísticas.

En este sentido, los hablantes poseen un conocimiento implícito de las construcciones convencionales y del impacto que la frecuencia de uso ejerce sobre las estructuras lingüísticas, cuya influencia puede incidir en el procesamiento y la organización del lenguaje, como la velocidad de acceso a las unidades lingüísticas, la evidencialidad (el efecto del *priming*) de las propiedades morfofonológicas en palabras de alta y baja frecuencia, así como en los procesos de gramaticalización (Bybee, 2001; Ellis, 2002).

Moreno Fernández (2012, p. 135) insiste en la relevancia del componente cuantitativo en el funcionamiento del sistema lingüístico y detalla una serie de fenómenos que lo evidencian:

- Los patrones lingüísticos con una alta frecuencia de uso tienden a ser más productivos que aquellos con una frecuencia menor.
- Los miembros específicos de las categorías gramaticales presentan una frecuencia de ocurrencia inferior en comparación con los elementos no específicos.
- Los elementos gramaticalizantes más frecuentes y con un significado más general dentro de una construcción pueden llegar a perder su capacidad de funcionar como verbos principales.
- Las construcciones que incluyen unidades de alta frecuencia suelen establecer conexiones más débiles con otras formas lingüísticas, lo que las hace más independientes y menos propensas a contribuir a la formación de clases productivas.
- Los ítems lingüísticos con una alta frecuencia de uso presentan una mayor fuerza léxica, lo que los hace más resistentes al cambio morfológico y analógico. Además,

funcionan como base para el cambio lingüístico y poseen una mayor autonomía dentro del sistema gramatical.

La sociolingüística cognitiva ofrece una aproximación al estudio de la gramática que, si bien se inscribe dentro de los postulados del cognitivismo, enfatiza la dimensión social, situacional y contextual de las formas gramaticales. Desde esta perspectiva, los hablantes poseen un conocimiento estructural de la lengua; aprenden y emplean secuencias y construcciones de manera contextualizada; participan activamente en la conformación de categorías lingüísticas; identifican analogías y regularidades; y son conscientes de la frecuencia de uso de las distintas unidades. Asimismo, los individuos tienen la capacidad de establecer equivalencias gramaticales en la interacción comunicativa y de modificar los significantes en función de los elementos lingüísticos y extralingüísticos presentes.

De este modo, la sociolingüística cognitiva reivindica un espacio propiamente social dentro de la gramática más allá de la variación. En esta concepción, la interacción social, la dinámica comunicativa y la percepción del hablante resultan factores esenciales para comprender el origen, la organización y los cambios de la gramática, en estrecha interdependencia con el nivel fónico y léxico-semántico.

1.3. Antecedentes

Durante casi cincuenta años, algunos investigadores han mostrado un gran interés por el fenómeno de la expresión del sujeto pronominal, lo que ha originado una extensa bibliografía sobre el tema. Esta alternancia sintáctica ha sido objeto de estudio en el ámbito de la sociolingüística desde diversos enfoques, como el contacto con otras lenguas o la adquisición del lenguaje, pero también desde la variación dialectal dentro de comunidades monolingües y bilingües. A pesar de que se ha estudiado principalmente el uso del pronombre personal sujeto, observar cómo los hablantes perciben estructuras sintácticas concretas también puede contribuir al mayor conocimiento de este fenómeno lingüístico.

En el marco de la producción, los estudios sobre el empleo del sujeto pronominal han sido tratados desde diversas perspectivas y, en ocasiones, se han basado en corpus variados, lo que ha supuesto dificultades a la hora de comparar los resultados de diferentes trabajos ante la falta de unos principios metodológicos homogéneos. En este sentido, los

estudios podrían dividirse en dos etapas, ya que es posible identificar un hito diferenciador con la aparición del Proyecto para el Estudio Sociolingüístico del español en España y América (PRESEEA). La creación de PRESEEA ha permitido superar muchas de las problemáticas anteriores, como la falta de corpus comparables, la ausencia de una metodología unificada y la falta de representatividad en las muestras.

1.3.1. Trabajos sobre el uso del sujeto pronominal

A continuación, se revisan los estudios más relevantes sobre el pronombre personal sujeto organizados de acuerdo con la perspectiva adoptada en nuestro trabajo. En primer lugar, se incluyen aquellas investigaciones que, desde un enfoque variacionista, han contribuido notablemente a la consolidación de las bases teóricas y metodológicas para el análisis de la producción de la expresión del sujeto pronominal, tanto dentro como fuera del marco de PRESEEA. En segundo lugar, aunque en menor número, se agrupan los estudios previos que han explorado la alternancia del sujeto pronominal desde el prisma de la percepción.

1.3.1.1. Perspectiva variacionista

Los primeros estudios sobre el sujeto pronominal en español han realizado contribuciones cruciales en cuanto a los fundamentos metodológicos y lingüísticos, al detectar e implementar algunos aspectos que han marcado el transcurso posterior de la investigación sobre este fenómeno.

Desde el punto de vista metodológico, se puede considerar el trabajo de Rosengren (1974) como el inicio de una nueva forma de investigar sobre el pronombre personal sujeto, ya que, aunque esté basado en datos procedentes de obras teatrales, emplea una base empírica sólida gracias a un corpus extenso y al empleo de procedimientos estadísticos. Asimismo, Barrenechea y Alonso (1973, p. 333) confirman que, para tratar con rigor los motivos subjetivos que rigen la elección del hablante, es necesaria la cuantificación de los factores objetivos para posteriormente hacer las conjeturas estilísticas pertinentes. Además, consideran que es más adecuado contemplar las cifras de la ausencia y presencia (*índices de saturación*), que reducir la comparativa a las tasas explicitación (*índices de magnitud*).

Los corpus empleados no siempre son homogéneos en cuanto a sus características. Los primeros trabajos sobre el sujeto pronominal se basan en datos de distinta naturaleza: mientras que Rosengren (1974) desarrolla sus investigaciones a partir de textos literarios, en los que el pronombre se inscribe en un contexto más formal, trabajos posteriores, como los de Barrenechea y Alonso (1973). En cambio, Bentivoglio (1980, 1987), Enríquez (1984) y Silva-Corvalán (1982) optan por basarse en interacciones más espontáneas como las obtenidas en entrevistas semidirigidas.

El trabajo de Barrenechea y Alonso (1973) en Buenos Aires y el de Enríquez (1984) en Madrid tratan el habla culta recogida en el *Proyecto de estudio coordinado de la norma lingüística culta de las principales ciudades de Iberoamérica y de la Península Ibérica* (PILEI). El estudio de Silva-Corvalán (1982) que analiza el habla de mexicanos en Los Ángeles y los trabajos de Bentivoglio (1980, 1987) sobre el habla de Caracas son pioneros en la investigación del pronombre personal sujeto en el discurso oral. En las entrevistas realizadas para el desarrollo de estos estudios, los investigadores seleccionan previamente ciertos temas y los van introduciendo a lo largo del transcurso de la propia entrevista. Este método permite captar una gran variedad de estilos comunicativos, que oscilan entre, lo que en términos labovianos se conoce como *careful speech* (Labov, 1972, p. 79), un discurso más controlado y reflexivo, hasta momentos en los que los participantes se acercan más a sus formas de habla más espontáneas e informales (Bentivoglio, 1987, p. 25).

Por otro lado, desde la perspectiva lingüística, se han introducido algunos matices relevantes en la conceptualización teórica del sujeto pronominal. Enríquez (1984) ofrece una actualización de la categoría persona, al considerar a esta como un elemento lingüístico, abstracto y universal, que se encuentra íntimamente ligado al contexto comunicativo y establece una relación directa entre los participantes en el discurso y el propio enunciado. Además, realiza una necesaria distinción entre la persona gramatical y la persona conceptual (Enríquez, 1984, p. 29-66): la persona gramatical se actualiza a través de los pronombres y las formas verbales; la persona conceptual no siempre coincide estrictamente con los valores gramaticales de persona que señala, lo que permite explicar asociaciones semánticas homogéneas y heterogéneas como, por ejemplo, en las personas del plural.

Estos estudios presentados constituyen aportaciones significativas en la identificación de factores condicionantes en el uso del pronombre sujeto. Uno de los aspectos más relevantes analizados es la naturaleza de la persona del sujeto, donde los rasgos de la animación y de la especificidad del referente parecen restringir considerablemente la presencia del pronombre. De modo que las referencias no humanas e indeterminadas limitan la explicitación del sujeto pronominal, ya que este suele aparecer con una referencia de persona humana claramente definida y determinada contextualmente (Enríquez, 1984, p. 335-336).

En esta misma línea, Barrenechea y Alonso (1973, p. 335) destacan que la 1ª persona del singular parece ser la que mayor uso acumula, aunque sugiere que esta tendencia puede deberse al tipo de entrevista empleado en la recogida de datos en la que el informante habla sobre sus experiencias personales. No obstante, al considerar la presencia del sujeto en términos generales, se detecta que los pronombres de 2ª persona (*usted*, *ustedes* y *vos*) son más frecuentes cuando funcionan como sujetos indeterminados que cuando hacen referencia directamente al interlocutor. Por su parte, Rosengren (1974, p. 221-234) señala una tendencia generalizada al uso de los pronombres más en singular que en plural, así como mayor empleo de la 1ª y 2ª persona del singular.

Otro aspecto determinante en la aparición del sujeto pronominal es la clase semántica del verbo. Los verbos relacionados con procesos psíquicos son los que mayor tasa pronominal obtienen (Enríquez, 1984) y, además, la presencia es particularmente elevada cuando los sujetos de 1ª y 2ª persona del singular están acompañados de verbos como *creer*, *hacer*, *decir*, *oír*, *saber* o *ver* (Rosengren, 1974). Por su parte, Bentivoglio (1987, p. 50), a partir de la ordenación de la tipología verbal propuesta por Givón (1984), establece una clasificación semántica de los verbos dividida en cinco categorías: *verbos de actividad cognoscitiva y mental*, *verbos de percepción y sensación*, *verbos de dicción y pronunciación oral*, *verbos de volición* y *el resto*.

En cuanto a las variables pragmático-discursivas, el trabajo de Silva-Corvalán (1982) introduce la variable de la correferencialidad —el cambio o el mantenimiento de referencia—, que se consolida como uno de los factores más explicativos de la alternancia entre la aparición o ausencia del sujeto pronominal. Además, subraya la importancia del contexto para el tratamiento de la ambigüedad morfológica, la cual no debe ser estudiada de manera aislada, ya que en el contexto pueden estar presentes algunos factores

desambiguadores como el rasgo [+humano], los pronombres clíticos o la continuidad referencial.

Barrenechea y Alonso (1973, p. 333) también coinciden al considerar el contexto como solucionador de la ambigüedad y advierten de que estudiar esta variable sin atender al contexto puede arrojar resultados similares entre formas ambiguas y no ambiguas. En cambio, cuando se incorpora el contexto al análisis, se observa que este facilita la identificación de la persona a la que se hace referencia y que solo en circunstancias muy reducidas queda sin resolver la ambigüedad, lo cual compromete la presencia del pronombre para asegurar la eficacia en la comunicación.

Otros estudios han centrado sus análisis en la ambigüedad fonética, como Morales (1997), Cameron (1993) y Ranson (1991), los cuales abordan la relación entre la presencia del sujeto y la pérdida de la /-s/ en la coda silábica del verbo. Estos estudios exploran cómo la elisión de este segmento fónico puede generar una mayor ambigüedad en la marcación morfológica del sujeto, lo que, a su vez, podría favorecer la aparición del pronombre como estrategia de clarificación referencial en determinados contextos discursivos. Este fenómeno se conoce como hipótesis funcional (Hochberg, 1986).

En definitiva, los trabajos previos han asentado las bases metodológicas y teóricas que sustentan los análisis contemporáneos sobre el uso del pronombre personal sujeto. Estas investigaciones han puesto de manifiesto que determinados factores —la persona gramatical, el rasgo humano, la especificidad del sujeto, los verbos de procesos mentales, la ambigüedad morfológica y el cambio de referente— parecen condicionar el uso del sujeto pronominal y, en algunos casos, estas variables también incrementan la frecuencia con la que este se explicita.

1.3.1.2. Investigaciones en el contexto de PRESEEA

Entre los fenómenos lingüísticos prioritarios abordados en el contexto del proyecto PRESEEA, se encuentra el estudio de la expresión del sujeto pronominal, el cual ha dado lugar a la creación de diferentes guías de codificación de esta variable lingüística. A partir de esta propuesta metodológica, comienzan a proliferar trabajos dentro del ámbito de PRESEEA y, gracias a ellos, se permite el examen sistemático y comparativo de este fenómeno en distintas comunidades hispanohablantes.

Sin embargo, es necesario advertir de que no todos los trabajos presentan los mismos enfoques. Algunos se orientan hacia una perspectiva más contrastiva con el objetivo de comparar varias zonas geográficas para identificar similitudes y diferencias entre ellas. Entre estos estudios destacan los primeros trabajos realizados en los corpus PRESEEA por Prieto Muñoz (2011) sobre el habla de San Juan de Puerto Rico, Bogotá y Valencia; Claes (2011), por su parte, busca determinar si la zona de las Antillas y el Caribe continental constituyen una única zona dialectal respecto al sujeto pronominal; y, finalmente, García Salido (2013) compara cuatro corpus diferentes entre los que se encuentran los corpus de PRESEEA-Alcalá y de PRESEEA-Valencia.

Más recientemente, otros estudios han ampliado la perspectiva comparativa a diferentes áreas geográficas. Por ejemplo, Martínez Lara (2016) compara el habla de Caracas, Medellín y Montevideo; Pérez Córdoba y Gomes Camacho, (2019) examinan las diferencias de uso del pronombre personal sujeto en Colombia, más concretamente las ciudades de Valledupar, Cartagena y Barranquilla; Manjón-Cabeza Cruz y Granados Romero, 2022 estudian las dos ciudades andaluzas de Málaga y Granada; y Orozco y Orozco (2022) abordan el análisis del habla de Medellín y Cali.

En cambio, algunos estudios han optado por centrar sus análisis exclusivamente en un área geográfica y emplean muestras más amplias para profundizar en el estudio de la variación pronominal dentro de un mismo entorno sociolingüístico. Ejemplo de ello son los trabajos de Martínez Lara, (2013) y Pérez Brabandere (2010) en el habla de Caracas, Repede (2019) en la ciudad de Sevilla, Martínez Lara et al. (2021) en Santiago de Chile, Martínez Lara (2022) en Montevideo, Manjón-Cabeza Cruz et al. (2016) en Granada o Roselló Verdeguer, (2021) en Valencia.

A pesar de que la metodología común del proyecto facilita las comparaciones, es conveniente señalar que algunos trabajos presentan ciertas particularidades que pueden condicionar las comparaciones entre ellos. Entre estas diferencias se encuentra el tratamiento exclusivo de la 1ª persona (Pérez Brabandere, 2010) o la 3ª persona (Martínez Lara, 2013, 2015, 2022), así como la consideración de un único nivel educativo (Repede, 2019). Por estos motivos, las comparaciones entre estos trabajos deben ser tratadas con cautela. A continuación, revisamos el tratamiento otorgado a los factores condicionantes del sujeto.

1.3.1.3. Factores explicativos de la alternancia

Los estudios previos han identificado y analizado diversas variables que pueden considerarse explicativas de la variación de la expresión del sujeto pronominal. PRESEEA proporciona dos directrices fundamentales para el desarrollo de los trabajos dentro de este proyecto: *la Guía de codificación de la variable expresión del sujeto pronominal* de Bentivoglio et al. (2011) y la recientemente actualizada *Guía PRESEEA de estudio de la expresión del sujeto pronominal* de Guerrero González y Manjón-Cabeza Cruz (2022). Por ello, antes de proceder al análisis de las variables seleccionadas en este estudio, resulta fundamental examinar cómo han sido tratadas en investigaciones anteriores con el propósito de establecer un punto de partida sólido y garantizar la coherencia metodológica.

Los trabajos identifican la expresión del sujeto pronominal —con sus variantes de ausencia y presencia del pronombre personal sujeto— como variable dependiente del objeto de estudio. No obstante, esta uniformidad en la selección de la variable dependiente no se extiende a la delimitación de los contextos de variación. Las investigaciones previas a la implementación de PRESEEA constatan una evidente carencia de criterios homogéneos que permitan diferenciar de manera sistemática los casos en los que se produce la alternancia del sujeto de aquellos en los que este permanece invariable. En este sentido, la unificación metodológica constituye un avance sustancial, dado que la primera guía (Bentivoglio et al., 2011) ya considera los contextos de presencia y ausencia obligatoria que deben ser excluidos del análisis de variación.

Uno de los objetivos del proyecto PRESEEA es la identificación de un patrón lingüístico que permita describir los factores condicionantes de la expresión del sujeto pronominal en las distintas áreas geográficas estudiadas. Al margen del proyecto PRESEEA, el trabajo de Flores Ferrán (2007) recopila, tras tres décadas de estudios sobre el sujeto pronominal, los principales aspectos metodológicos y las variables que han demostrado ser determinantes en la presencia y ausencia del pronombre. Sin embargo, la contribución de Martín Butragueño (2020) supone un avance significativo en el esbozo de un patrón lingüístico subyacente en los estudios desarrollados en el marco de PRESEEA. Su aproximación no solo permite sintetizar los hallazgos previos, sino también ofrecer una visión global sobre el funcionamiento de esta variable sintáctica en la geografía hispanohablante.

En cuanto a las variables independientes, revisamos de los factores más explicativos del objeto de estudio en función de su ámbito de incidencia. Para ello, agrupamos a continuación aquellas que conciernen al sujeto de la oración, las relacionadas con el verbo, las que pertenecen al plano pragmático-discursivo y las de naturaleza social.

Entre las variables referentes al sujeto de la oración destaca la persona del sujeto, la cual ejerce una gran influencia en la ausencia y presencia del sujeto pronominal. Sin embargo, su tratamiento en la literatura no ha sido idéntico en todos los trabajos. Algunos optan por desglosarla en función del número gramatical (singular y plural) y de la persona gramatical (1ª, 2ª o 3ª persona), y otros se acogen a un análisis individualizado en el que asignan a cada pronombre las características de persona y número correspondientes. Aunque la primera guía de PRESEEA (Bentivoglio et al., 2011) incluye el pronombre *uno/a*, este se elimina en la segunda guía (Guerrero González & Manjón-Cabeza Cruz, 2022) debido al bajo número de casos detectados.

Los resultados obtenidos en los diferentes trabajos parecen converger en una misma dirección: un uso predominante y generalizado de la 1ª persona del singular y, en determinadas áreas geográficas como el Caribe, se observaba también un empleo elevado de la 2ª persona del singular (Claes, 2011). Del mismo modo, los estudios previos coinciden al obtener resultados en los que el plural parece desfavorecer la presencia del sujeto pronominal. Todo parece indicar que, como señala Claes (2011, p. 200), el comportamiento asociado a la persona y número gramatical “no constituye una característica de la gramática a nivel de los dialectos bajo análisis, sino del diasistema hispánico”.

Otra variable referente al sujeto de la oración es la especificidad. Entre los estudios que analizan esta variable, debemos destacar el de Cameron (1993), quien advierte que la especificidad no se analiza en todos trabajos previos del ámbito sociolingüístico, y, además, señala que la 2ª persona del singular puede tener diferentes usos según el grado de especificidad del sujeto. Un fenómeno similar se documenta en la 3ª persona del plural. Lapidus Shin y Otheguy (2005) afirman que la especificidad puede manifestarse en distintos grados clasificados como no referenciales, semi referencial y corporativo. No obstante, las guías de PRESEEA (Bentivoglio et al., 2011; Guerrero González & Manjón-Cabeza, 2022) solo contemplan las opciones de sujeto específico o no específico.

En segundo lugar, también se han tratado variables referentes a las propiedades verbales que, si bien son diversas, se pueden agrupar en dos categorías: las que hacen referencia al tiempo, modo y aspecto del verbo, y aquellas que remiten a la clase semántica. Algunos trabajos, como el de Pérez Brabandere (2010) o Claes (2011), abordan los predictores de tiempo, aspecto y modo de forma conjunta, en cambio las guías de metodológicas de PRESEEA (Bentivoglio et al., 2011; Guerrero González & Manjón-Cabeza, 2022) recomiendan el análisis individualizado de cada variable. A pesar de no establecerse un patrón de uso determinante para el conjunto de trabajos revisados, los resultados sugieren que el imperfecto tiende a ser el tiempo verbal que mayor presencia del sujeto promueve en comparación con el resto de los tiempos verbales.

En cuanto a la clase semántica del verbo, esta variable ha demostrado ser de una de las más explicativas en trabajos previos, aunque su análisis se ha llevado a cabo desde clasificaciones muy diversas (Orozco y Orozco, 2022). Trabajos pioneros como el de Bentivoglio (1980) identifican una tendencia de la 1ª persona a aparecer acompañada de determinados verbos con mayor frecuencia y, a partir de esta observación, la autora considera necesario incorporar la clasificación semántica de los verbos de Givón (1984). Más tarde, Enríquez (1984) profundiza en el contenido semántico del verbo y espera que aquellos relacionados con una actividad psíquica sean los que mayor tasa pronominal requieran, ya que en estos se encuentra implícita la opinión del hablante.

Otros estudios más recientes han realizado sus análisis desde tipologías diversas: Travis y Torres Cacoullos (2012) emplean una clasificación binaria (verbo psicológico frente a otros verbos); Erker y Guy (2012) estudian la relación entre la frecuencia léxica y la aparición del sujeto; y Orozco y Hurtado (2020) aplican la clasificación kinésica. Las investigaciones realizadas en el marco del PRESEEA siguen la codificación mínima propuesta en la primera guía (Bentivoglio et al., 2011) que distingue entre verbos de procesos mentales, verbos de estado, verbos dicendi, verbos de actividad y otra clase, pero tras la actualización de la guía (Guerrero González & Manjón-Cabeza, 2022) se recomienda el uso de la clasificación ADESSE (García-Miguel et al., 2003).

En cuanto a los factores que conciernen al plano pragmático-discursivo, numerosos estudios incluyen la continuidad o el cambio referencial. Esta variable ha recibido diferentes denominaciones. Algunos autores (Cameron, 1993; Flores Ferrán, 2004) la abordan como cambio de referencia o cambio del referente del sujeto, mientras que Silva

Corvalán (1982, 2003) opta por el término correferencialidad. Desde esta perspectiva más amplia, la correferencialidad permite examinar la relación entre los sujetos y otros factores, como el cambio o el mantenimiento del referente previo en forma de objeto o sujeto. Las guías de PRESEEA (Bentivoglio et al., 2011; Guerrero González & Manjón Cabeza, 2022) sugieren una codificación similar a la propuesta Silva-Corvalán, de modo que distinguen entre correferencialidad con sujeto o con objeto y la no correferencialidad, a la que también se incorporan una cuarta opción no aplicable propuesta por Orozco y Guy (2008, p. 74). En líneas generales, los trabajos parecen coincidir al considerar que el cambio de referente indica una mayor presencia del sujeto pronominal.

Otra variable considerada en múltiples trabajos es la ambigüedad morfológica. Sin embargo, la mayoría de los estudios la abordan sin considerar el contexto comunicativo, donde pueden intervenir otros factores que permitan desambiguar la forma verbal (Cameron, 1993; García Salido, 2013; Hurtado, 2005; Martínez Lara et al., 2021; Paredes Silva, 1993). Un planteamiento similar se observa en los trabajos dentro PRESEEA, ya que las guías de codificación (Bentivoglio et al., 2011; Guerrero González & Manjón Cabeza, 2022) se limitan a identificar si la forma verbal es ambigua o no, independientemente del contexto comunicativo en el que se producen. Como consecuencia, los trabajos basados en los corpus PRESEEA replican esta metodología (Manjón-Cabeza Cruz et al., 2016; Manjón-Cabeza Cruz y Granados Romero, 2022; Martínez Lara, 2013, 2022; Pérez Córdoba & Gomes Camacho, 2019; Repede, 2019; Roselló Verdeguer, 2021). Los resultados obtenidos hasta el momento indican que las formas ambiguas requieren un mayor uso del pronombre personal sujeto, pero esta necesidad de explicitación es aún mayor cuando confluyen los factores de ambigüedad morfológica de la forma verbal y ambigüedad contextual (Pérez Brabandere, 2010; Silva-Corvalán, 1982).

La variable de ambigüedad fonética también se incorpora al análisis de la expresión del sujeto pronominal, especialmente en estudios sobre áreas geográficas cuyas características fonéticas propias —como el habla caribeña o andaluza, entre otras— pueden resultar ambiguas (Cameron, 1993; Hochberg, 1986; Ranson, 1991). Han sido varias las teorías planteadas respecto a la ambigüedad fonética, entre las cuales se puede destacar la Hipótesis Funcional de Hochberg (1986), que se basa en que la pérdida de información semánticamente relevante —la pérdida de la /-s/ o /-n/ en la morfología

flexiva— produce la activación de otros mecanismos compensatorios para reconducir la ambigüedad causada.

En cambio, Ranson (1991) detecta el efecto contrario al analizar el habla de Puente Genil (Córdoba), ya que los pronombres suelen tener menos aparición con formas ambiguas que con formas no ambiguas. Más tarde, Cameron (1993) analiza la ambigüedad fonética con una muestra más amplia (10 hablantes de Madrid y 10 de San Juan de Puerto Rico) y, además, considera factores como el rasgo de animación del sujeto [+ humano], la correferencialidad y la especificidad. A pesar de que las guías de PRESEEA (Bentivoglio et al., 2011; Guerrero González & Manjón-Cabeza Cruz, 2022) no han recogido la ambigüedad fonética como variable, varios estudios realizado en el marco de PRESEEA han decidido incorporarla en sus análisis, como las ciudades de Granada (Manjón-Cabeza Cruz et. al, 2016; Manjón-Cabeza Cruz & Granados Romero, 2022) y Sevilla (Repede, 2019).

Por último, entre las variables sociales que se han tenido en cuenta en el estudio del pronombre personal sujeto se encuentran aquellas utilizadas para la estratificación de la muestra de los corpus PRESEEA: edad, sexo y nivel de estudios. A pesar de que estas no han resultado ser significativamente influyentes en la mayoría de estudios, Martín Butragueño (2020) sostiene que la variación social, entendida en términos de macrovariación, presenta una magnitud relativamente reducida dentro de cada comunidad de habla. Además, el autor no identifica un patrón estable que se mantenga de manera homogénea en las distintas áreas hispanohablantes analizadas, lo que sugiere que los factores sociales tienen un impacto menor que los condicionantes lingüísticos en la expresión del sujeto pronominal

1.3.2. Trabajos sobre la percepción del sujeto pronominal

A pesar de que la investigación sobre la expresión del sujeto pronominal en español tradicionalmente se ha centrado en su uso, también encontramos estudios que, desde la óptica de la percepción, analizan cómo los oyentes o lectores procesan y resuelven posibles ambigüedades ocasionadas en el discurso. Estos trabajos se centran en la influencia que tienen la presencia o ausencia de distintos elementos —sintagmas nominales, pronombres átonos o, en particular, pronombres tónicos, como el pronombre personal— en la interpretación del referente. Dichos contextos ambiguos suelen darse cuando no hay un

referente único claro o cuando coexisten entre varios candidatos a ser sujeto u objeto de la oración, de modo que el receptor debe activar mecanismos de desambiguación para identificar correctamente a quién se está haciendo alusión.

Desde el ámbito experimental, aunque la mayoría de los estudios en español traten sobre los procesamiento cognitivos implicados en la resolución de la ambigüedad, las técnicas experimentales empleadas en estas investigaciones no han sido siempre las mismas. Entre las diversas metodologías se encuentran los cuestionarios *offline* (Alonso-Ovalle et al., 2002; Chamorro, 2018; Gelormini Lezama & Almor, 2013; Gelormini Lezama & Almor, 2011) en los que se presentan secuencias de enunciados escritos o en forma de audio, donde se pide al participante que juzgue la coherencia de la oración o indique a qué referente se refiere en cada caso. Otra estrategia es la lectura donde cada individuo lee a su propio ritmo, conocida como *self-paced reading*. En estos casos, los participantes leen un texto u oraciones y deben presionar un botón para medir el tiempo de lectura y, por lo tanto, registrar los tiempos de procesamiento (Filiaci et al., 2014; Keating et al., 2016). Por último, encontramos los trabajos donde se recurre a la herramienta de seguimiento ocular (Chamorro et al., 2016; Cristerna Román, 2020) con el objetivo de recopilar los tiempos de procesamiento.

La herramienta de *eye-tracking* permite recoger, además del tiempo dedicado a la lectura, las fijaciones realizadas por los participantes durante la lectura de los estímulos, ya sea en oraciones aisladas o textos de mayor extensión. Al mismo tiempo, se pueden registrar los movimientos oculares tanto al avanzar en el texto como al retroceder para revisar la información leída previamente. En una situación prototípica, el texto presenta dos o más personajes que concuerdan en género, número o función sintáctica, y el pronombre pronominal (tónico o átono) puede contribuir al esclarecimiento del referente. Frente a situaciones de ambigüedad, los lectores utilizan diferentes estrategias de resolución; por ejemplo, cuando la información proporcionada resulta insuficiente para esclarecer el referente, dedican un mayor tiempo en la primera lectura o vuelven a leer el enunciado con el fin de reinterpretar la referencia.

Es conveniente destacar que los enfoques metodológicos de estos estudios pueden ser muy diversos, dado que las técnicas experimentales utilizadas —como las pruebas de percepción y procesamiento— resultan altamente aplicables a distintos ámbitos, como es el caso del aprendizaje de lenguas o el estudio de trastornos del lenguaje, entre otros.

1.4. Objetivos, preguntas e hipótesis de investigación

Una vez revisados los conceptos teóricos y los estudios previos en relación con nuestro objeto de estudio, con el propósito de llevar a cabo el estudio de la variación del sujeto pronominal en la ciudad de Málaga desde el enfoque de la sociolingüística cognitiva, la estructura de este trabajo se divide en dos bloques principales, la producción y la percepción. A partir de los interrogantes a los que se pretende responder en cada bloque, formulamos las hipótesis generales y definimos los objetivos correspondientes. Las hipótesis específicas relativas a aspectos más precisos se detallarán en las introducciones correspondientes a cada bloque (2.1 y 3.1).

1.4.1. Producción del sujeto pronominal

Al abordar la producción del sujeto pronominal en la ciudad de Málaga, surgen preguntas transversales a cualquier comunidad de habla hispana. En primer lugar, resulta necesario indagar en qué lleva a los hablantes a explicitar el sujeto en una lengua como el español, cuya morfología verbal presenta una gran riqueza flexiva. Una vez excluidos los contextos obligatorios de presencia y ausencia del sujeto pronominal, es relevante cuestionarse qué condiciones lingüísticas motivan al hablante a omitir o a explicitar el sujeto.

En base a los antecedentes revisados, suponemos que existirán factores lingüísticos relacionados con el sujeto de la oración, con los verbos, con las situaciones pragmático-discursivas y con las características sociales del individuo que condicionen la variación. De este modo, surge el interrogante sobre qué factores lingüístico-sociales fomentan más la presencia del sujeto y, asimismo, qué variables restringen la aparición del sujeto pronominal.

Más concretamente, este trabajo procura dar respuesta a qué factores lingüísticos y sociales inciden en mayor medida en la expresión del sujeto pronominal en la ciudad de Málaga. Asimismo, dado que nuestro análisis se basa en dos sincronías dentro de la misma ciudad, resulta relevante comprobar cómo ha evolucionado el comportamiento del sujeto pronominal a lo largo del tiempo en una misma ciudad, y qué diferencias o similitudes se detectan entre los dos cortes temporales. Por último, gracias a la metodología homogénea

establecida por el proyecto PRESEEA que permite la comparación los resultados de Málaga con los de otras ciudades, nos planteamos investigar qué factores condicionantes detectados en la urbe malagueña son extrapolables a otras áreas hispanohablantes y, en un sentido más amplio, cuál podría ser el patrón de uso general del sujeto pronominal en español.

Así pues, se formulan las siguientes hipótesis iniciales orientadas a esclarecer las causas que propician la aparición u omisión del pronombre personal sujeto en el español hablado en Málaga.

1. *Hipótesis de la gramaticalización subyacente.* Los condicionantes lingüístico-sociales identificados como significativos en investigaciones anteriores basadas en comunidades de habla similares (Manjón-Cabeza Cruz et al., 2016; Repede, 2019; Roselló Verdeguer, 2021) afectarán también al habla de la ciudad de Málaga. Esto corroboraría la existencia de una gramática subyacente común a todas las variedades del español, aunque superficialmente existan diferencias en las tasas pronominales (Orozco et al., 2024).
2. *Hipótesis del impacto de las variables lingüístico-sociales.* Las variables lingüísticas tendrán mayor peso que las sociales en la variación del sujeto pronominal (Martín Butragueño, 2020a).
3. *Hipótesis de restricción interna de la lengua.* La aparición u omisión del sujeto estará condicionada por características inherentes a la naturaleza del sujeto, como la especificidad o el rasgo de animación.

A partir de estos interrogantes e hipótesis que servirán de faro a nuestra investigación, establecemos los siguientes objetivos:

- a. Identificar los factores condicionantes en el uso. Este primer objetivo trata de analizar principalmente el uso del sujeto pronominal en el español hablado en la ciudad de Málaga. Para ello, se emplean los datos procedentes de los dos corpus orales de PRESEEA-Málaga. Concretamente, se busca determinar con qué frecuencia se produce la ausencia y presencia del sujeto pronominal y qué factores —tanto lingüísticos como sociales— condicionan su uso.

Además del objetivo principal, definimos dos objetivos secundarios con el fin de establecer comparativas diacrónicas y sincrónicas respecto a la ciudad de Málaga.

- b. Evaluar la evolución interna en Málaga (1995 y 2015). Este objetivo pretende detectar los posibles cambios de uso en la expresión del sujeto a lo largo de dos décadas en la ciudad de Málaga. Además de comparar la frecuencia de aparición del sujeto pronominal en el habla malagueña en cada corpus, se identifican variaciones significativas según factores lingüísticos y sociales condicionantes.
- c. Contrastar los resultados de Málaga con otras ciudades del proyecto PRESEEA. A pesar de las disparidades, la homogenización metodológica permite comparar los resultados obtenidos en las distintas zonas hispanohablantes estudiadas con el objetivo de trazar un posible patrón general que justifique el uso del sujeto pronominal.

Con el propósito de abordar el estudio del sujeto pronominal lingüístico desde el prisma de la sociolingüística cognitiva, el siguiente bloque surge a partir los datos sobre el uso obtenidos en el análisis de corpus (la producción) para profundizar en la percepción cognitiva de la variación la ausencia y presencia del sujeto pronominal.

1.4.2. Percepción cognitiva del sujeto pronominal.

Entre las variables lingüísticas estudiadas en el entorno controlado del corpus, se encuentran la ambigüedad morfológica y la correferencialidad, entendida esta última como continuidad y cambio referencial. Ante estos factores, el sujeto pronominal parece ser explicitado con mayor frecuencia y, por ello, se plantea la necesidad de explorar cómo se perciben cognitivamente estas estructuras cuando el sujeto está presente u omitido.

De esta manera, surgen interrogantes sobre qué función cumple el sujeto en oraciones cuyos verbos son morfológicamente ambiguos. Si se parte de la premisa de que el sujeto contribuye a una lectura más fluida de las oraciones, nos preguntamos cómo afectará la redundancia pronominal en los enunciados y, por el contrario, cuál será el efecto de la ausencia total del sujeto en la comprensión del enunciado. Asimismo, en los casos en los que el sujeto esté presente, pero introduzca un nuevo referente respecto al presentado anteriormente, resulta relevante considerar cómo repercutirá el cambio referencial en el procesamiento de la oración.

Por lo tanto, en lo que respecta a la percepción cognitiva, también nos surgen preguntas generales sobre qué estructuras sintácticas serán las que menor coste de

procesamiento requieran y, por el contrario, cuáles exigirán un mayor sobreesfuerzo cognitivo para ser comprendidas.

Estas preguntas han sido formuladas para guiar el análisis y desarrollar de manera más específica las hipótesis propuestas a continuación:

1. Hipótesis de la resolución pronominal. La presencia del pronombre personal facilitará la resolución de la ambigüedad, al constituir un anclaje cognitivo que reduce la carga interpretativa (Ariel, 2001).
2. Hipótesis del grado de aparición. Aunque el sujeto puede ser un elemento útil para entender los enunciados, se anticipan diferentes tendencias de procesamiento cognitivo según el grado de aparición o de omisión. Estas diferencias pueden favorecer la interpretación o, por el contrario, dificultar la lectura, ya sea por cuestiones de recursividad o por la introducción de nuevos referentes (Badecker & Straub, 2002; Gernsbacher, 1989; Grant et al., 2020).

Por lo tanto, tras establecer las preguntas e hipótesis esperadas en esta investigación, el objetivo principal de este bloque es el siguiente:

- a. Determinar la función que cumple el sujeto pronominal en el procesamiento cognitivo de los enunciados mediante la herramienta de eye-tracking. Este objetivo pretende identificar cómo afecta la variación entre la explicitación u omisión del pronombre personal sujeto en el entendimiento de la oración. Además, se busca establecer qué estructuras suponen un mayor o menor esfuerzo cognitivo.

El objetivo general deriva en dos objetivos específicos que están vinculados con las variables lingüísticas analizadas:

- b. Observar las estrategias cognitivas implicadas en la resolución de la ambigüedad. A partir de los diferentes parámetros de tiempo de lectura, se pretende identificar disparidades en los patrones de procesamiento al comparar oraciones que pueden resultar ambiguas con aquellas que no presentan ambigüedad.
- c. Comparar los efectos de la continuidad referencial frente al cambio de referente. Este objetivo se orienta a analizar como repercute la introducción de un nuevo referente en la carga cognitiva durante la primera lectura y su relectura.

En definitiva, la visión desde la percepción cognitiva permite complementar la dualidad esperada desde la sociolingüística cognitiva, al integrar la producción lingüística con los procesos de comprensión y recreación mental de la variación.



UNIVERSIDAD
DE MÁLAGA

CAPÍTULO 2. LA PRODUCCIÓN DEL PRONOMBRE PERSONAL SUJETO EN LOS CORPUS PRESEEA-MÁLAGA

2.1. Introducción al análisis mediante corpus

El presente capítulo está dedicado al estudio del uso del sujeto pronominal desde un enfoque metodológico propio de la investigación sociolingüística. A partir de los materiales recopilados en los corpus orales de comunidades de habla específicas, se indaga sobre qué factores condicionantes de carácter lingüístico (fonéticos, morfológicos, léxicos, pragmáticos, discursivos y semánticos) o de naturaleza social motivan la omisión y presencia del sujeto pronominal. Con el propósito de establecer las relaciones entre estas variables, se hace necesaria la utilización de métodos cuantitativos para discernir de qué modo la frecuencia de concurrencia de distintos factores lingüístico-sociales explica la variación.

Aunque el mapa lingüístico de la ciudad de Málaga lleva siendo dibujado desde hace tres décadas, la variación ausencia-presencia del sujeto pronominal no había recibido hasta ahora una atención particular en la ciudad. Gracias a que la urbe malagueña dispone de materiales sonoros que recogen las voces locales desde los años 50 (Ávila Muñoz et al.,

2024) hasta la actualidad (Vida Castro, 2007; Ávila Muñoz et al., 2008; Lasarte Cervantes et al., 2009; Cruces Rodríguez, en prensa; Vida Castro, en prensa), hoy podemos comprobar si se han experimentado cambios significativos en el uso del pronombre personal sujeto en las dos sincronías que hemos seleccionado para nuestro estudio (1995-2015).

Esta contribución, además de enriquecer el mapa lingüístico de la ciudad de Málaga, amplía su alcance al integrarse a su vez en el Proyecto para el Estudio Sociolingüístico del Español en España y América (PRESEEA). Al trabajar con los corpus PRESEEA-Málaga, los resultados obtenidos en esta ciudad son comparables con los de otras zonas de habla hispana y, de esta forma, los acuerdos metodológicos propuestos por el proyecto nos permiten dibujar un patrón sobre el uso del pronombre personal sujeto a lo largo de la geografía hispánica.

En este contexto, el análisis cuantitativo de los datos extraídos de los corpus PRESEEA-Málaga permite no solo describir el comportamiento lingüístico observado en dos momentos sincrónicos distintos, sino también formular principios explicativos que permitan comprender los mecanismos subyacentes a su variación. A partir de la observación sistemática de los factores que intervienen en la elección entre la presencia y la omisión del sujeto, se plantean las siguientes hipótesis específicas que aspiran a esclarecer las motivaciones funcionales, cognitivas, semánticas y pragmático-discursivas que condicionan esta variación.

1. *Hipótesis del egocentrismo directo e indirecto.* La 1ª persona del singular será la forma pronominal con mayor frecuencia de aparición, debido al egocentrismo intrínseco a la naturaleza de las lenguas (Givón, 1976). Esto implicará no solo un mayor uso de este pronombre respecto al resto de personas gramaticales, sino también una mayor presencia del sujeto explícito.
2. *Hipótesis de la especificidad referencial.* La inespecificidad referencial restringirá el uso del pronombre, ya que su presencia podría comprometer el carácter inespecífico del referente (Cameron, 1993)
3. *Hipótesis de la pérdida referencial contextual.* Cuando se ha mencionado previamente un elemento en el discurso o cuando se produce una ambigüedad debida a la coincidencia de rasgos entre varios elementos, la recuperación del referente se llevará a cabo a través de la presencia del sujeto. Esta pérdida

referencial está influenciada por diversos factores pragmático-discursivos, tales como las intervenciones extensas, la existencia de múltiples referentes y la interrupción por parte del entrevistador, entre otros. Además, es de esperar que las secuencias próximas al cambio de turno conversacional sean más proclives al mantenimiento del sujeto pronominal para marcar la participación activa del emisor (Blas Arroyo, 2005).

4. *Hipótesis pragmático-discursiva*. El cambio referencial tendrá un impacto más significativo en la presencia del pronombre que los cambios de turno conversacional y el tipo de discurso. No obstante, ciertas tipologías como la argumentación requerirán un mayor empleo del pronombre debido a su naturaleza discursiva (Lastra & Martín Butragueño, 2015).

5. *Hipótesis de la semántica verbal*. Los verbos de procesos mentales, percepción sensorial, opinión y entendimiento patrocinarán una mayor aparición del sujeto explícito debido al rasgo de subjetividad que los caracteriza (Enríquez, 1984). Dado que las acciones expresadas por estos verbos implican procesos cognitivos del individuo, será más probable la aparición explícita del sujeto.

6. *Hipótesis de la ambigüedad descontextualizada*. A pesar de que se pueda inducir cierto grado de ambigüedad por la elisión de la /-s/ en la coda de la morfología flexiva del verbo o por el sincretismo de la 1ª y 3ª persona en algunos tiempos verbales, como el imperfecto, se anticipa que habrá marcas textuales que mitiguen dicha ambigüedad. En consecuencia, cuando la ambigüedad no pueda ser resuelta por otros elementos, habrá una mayor presencia del pronombre (Silva-Corvalán, 1982).

A partir de las hipótesis expuestas, se plantea el diseño metodológico para llevar a cabo este trabajo, así como los criterios de selección de casos, las variables consideradas y las herramientas empleadas para el análisis estadístico.

2.2. Propuesta metodológica para el estudio del pronombre personal en la ciudad de Málaga

A pesar de los numerosos trabajos en el ámbito hispano sobre el pronombre personal sujeto, estos no se han basado siempre en una metodología homogénea ni en lo

que se refiere a la recogida de datos ni al procedimiento de selección de casos y variables. El proyecto PRESEEA —orientado a la exploración de diferentes áreas urbanas hispanohablantes con el objetivo de obtener una muestra representativa del universo lingüístico analizado y asegurar la comparabilidad entre las diferentes comunidades de habla— ofrece unas directrices metodológicas generales que aseguran la coherencia y la comparabilidad de los datos desde una perspectiva sociolingüística.

A partir del marco metodológico común propuesto por PRESEEA, se establece la muestra necesaria para cada población atendiendo a las macrocategorías sociales de edad, género y nivel de instrucción, y, además, los fenómenos lingüísticos considerados prioritarios disponen de una guía de codificación en la que se aconseja la exclusión de ciertos contextos de análisis y la selección de variables comunes. De esta forma, este esquema metodológico inicial posibilita que cada comunidad de habla se adhiera a los principios generales establecidos por PRESEEA, pero que también pueda incorporar variables que reflejen sus particularidades sociolingüísticas específicas.

En el caso concreto de la ciudad de Málaga, consideramos necesario incorporar el estudio de la ambigüedad fonética que puede producirse en la morfología flexiva del verbo por la frecuente elisión de la /-s/ en posición postnuclear (Vida Castro, 2004). Esta elisión puede hacer confluir las formas verbales, si la /-s/ se elide completamente, en casos como la segunda y 3ª persona de singular del presente de indicativo (cantas/canta). Del mismo modo, prescindimos del pronombre vos por no ser característico del habla malagueña. Una aportación novedosa respecto a estudios anteriores radica en la contextualización de la ambigüedad, ya que prestamos especial atención a otros factores discursivos que intervienen en dicha situación comunicativa y que pueden ser esenciales para su resolución.

A continuación, se detallan las características esenciales de la muestra estudiada en la ciudad de Málaga correspondiente a los dos cortes temporales citados y los aspectos metodológicos para el desarrollo del análisis, como los criterios de exclusión de casos y la selección de variables analizadas.

2.2.1. Muestra

La selección de informantes se preestratifica a partir de unas cuotas de afijación uniforme que permiten clasificar a la comunidad de habla en subpoblaciones según las siguientes variables macrosociales: la edad, dividida en tres grupos generacionales —primer grupo entre 20 y 34 años, segundo grupo entre 35-54 años y tercer grupo más de 55 años—, el sexo —mujeres y hombres—, y el nivel de instrucción —estudios básicos: sin estudios o con educación obligatoria; estudios intermedios: bachillerato, formación profesional; estudios superiores: universitarios—. Además, se emplea el mismo número de informantes para cada subgrupo poblacional.

Esta Tesis Doctoral analiza el habla oral de la ciudad de Málaga basándose en dos corpus que recogen las voces de la ciudad: *El español hablado en Málaga*, al que haremos referencia como PRESEEA-Málaga 1995, ya que de esa fecha data la recogida de los materiales (Vida Castro, 2007; Ávila Muñoz et al., 2008; Lasarte Cervantes et al., 2009), y *Corpus oral complementario del español hablado en la ciudad de Málaga*, que nombraremos como PRESEEA-Málaga 2015 (Vida Castro, en prensa).

El volumen muestral del corpus de 1995 se distribuye de manera uniforme con cuatro informantes por cada casilla (N=72) y en cuanto al corpus de 2015, se asignan tres informantes por casilla (N=54). En definitiva, la muestra que abordamos en este trabajo está formada por un total de 126 informantes, cuya distribución se presenta en la Tabla 1.

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

Tabla 1. Muestra-tipo por cuotas propuesta para el proyecto PRESEEA (corpus 1995 y 2015)

Ambos corpus están formados por entrevistas semidirigidas en las que se tratan los temas propuestos por PRESEEA: saludos, el tiempo, lugar donde vive, familia y amistad, costumbres, peligro de muerte, anécdotas, deseos de mejora económica y final de entrevista. Esta diversidad temática favorece la aparición de distintos fenómenos lingüísticos en los distintos niveles de la lengua (Moreno Fernández, 2005).

Cada entrevista posee una duración de entre 45 y 60 minutos aproximadamente y, para registrar las entrevistas de 1995 se recurrió a una grabadora analógica y para el corpus de 2015 se utilizó la grabadora Roland R-05. Las entrevistas solían desarrollarse en ambientes con escasa interferencia acústica, ya que la mayoría de ellas tuvieron lugar en un despacho de la facultad o en el domicilio del informante. Pese a que esta situación podría limitar la realización de ciertos análisis fonéticos, se prescindió del uso de espacios insonorizados para propiciar una interacción más natural, en la que el informante pudiera encontrarse más relajado y espontáneo y, por lo tanto, pudiera expresarse en su variedad más vernácula. Una vez grabadas las entrevistas, se transcribieron y etiquetaron los materiales sobre los que, posteriormente, se anotan los posibles casos de estudio.

Además de la entrevista semidirigida, cada informante cumplimenta un cuestionario sociológico que recoge información sobre su lugar de nacimiento y residencia actual, formación académica, situación económica, entorno familiar y social, relación con los medios de comunicación, reparto de tareas en el hogar, actitud hacia la norma, entre otras muchas. A pesar de que los cuestionarios sociológicos que complementan las entrevistas de cada corpus varían ligeramente, ambos proporcionan datos suficientes para reconstruir los diferentes perfiles sociales integrados en la muestra más allá de las tres variables sociales de preestratificación.

2.2.2. Selección de casos y variables lingüístico-sociales

Para llevar a cabo el análisis de los datos, seguimos las directrices establecidas por las guías de PRESEEA (Bentivoglio et al., 2011; Guerrero González & Manjón-Cabeza Cruz, 2022), donde se recomienda evitar los primeros minutos de las entrevistas y aconsejan comenzar a analizar el discurso a partir del minuto diez. En consecuencia, seleccionamos de cada entrevista un fragmento representativo de cinco minutos, correspondiente al intervalo comprendido entre los minutos diez y quince. Aunque la guía de 2022 (Guerrero González & Manjón-Cabeza Cruz, 2022) establece una selección

temporal de la entrevista, no especifica el número de casos por informante, aspecto que debe ser observado antes de realizar los análisis cuantitativos de las variables sociales para garantizar que las diferencias en la producción lingüística no interfieren en el análisis y, por lo tanto, que realmente se esté estudiando el comportamiento de los individuos según sus características sociales.

Otra consideración relevante de este trabajo es la selección de la cláusula, definida como la estructura formada por el sujeto, el verbo y los complementos (sean o no argumentales), como la unidad para basar el análisis. Además, cuando las variables lo requieren —por ejemplo, en contextos de correferencialidad o en la delimitación del turno de habla—, codificamos la relación de la cláusula de análisis con respecto a la inmediatamente anterior.

Antes de proceder a la observación de las variables expuestas en nuestra guía de codificación (Anexo I), se hace constar que se han excluido del análisis los casos de presencia obligatoria del pronombre: frases idiomáticas (*¡yo qué sé!*), usos enfáticos con *solo/a* y *mismo/a* (*se rodean de las conchas que se ponen ellos mismos* [MA_H21_734]) y usos focales contrastivos o informativos (*él tiene 50 años y yo tengo 59* [MA_M21_042]). Asimismo, los contextos de ausencia obligatoria tampoco han sido considerados en el análisis como oraciones impersonales (*había un profesor muy bueno de <término> matemáticas* [MA_M22_703]) o en frases hechas (*¡ya está!* [MA_H22_014], *¡vamos!* [MA_H22_012]).

Los materiales de este estudio se codifican según la *Guía de codificación de la variable expresión del sujeto pronominal* por Bentivoglio et al. (2011), la cual ha sido recientemente actualizada en la *Guía PRESEEA de estudio de la expresión del sujeto pronominal* de Guerrero González y Manjón-Cabeza Cruz (2022). En nuestro estudio, analizamos variables que provienen de ambas propuestas metodológicas y, además, incorporamos factores que no están contemplados previamente en ninguna de las dos guías. De este modo, consideramos como variable dependiente la alternancia de ausencia y presencia del pronombre personal sujeto y, como variables independientes, se contemplan 19 de naturaleza lingüística y tres sociales. A continuación, se detallan las variables estudiadas y en el apartado dedicado a los resultados bivariantes (2.3.2), presentamos las variables acompañadas de ejemplos con el objetivo de ilustrar los resultados obtenidos.

En cuanto a las variables compartidas por ambas guías (Bentivoglio et al., 2011; Guerrero González & Manjón-Cabeza Cruz, 2022), presentamos a continuación cómo estas han sido codificadas o, en su defecto, recodificadas para esta investigación:

- **Persona del sujeto.** Las dos guías de codificación aconsejan el estudio del paradigma pronominal según la persona (1ª, 2ª y 3ª) y número (singular y plural). Además, incluyen el pronombre vos, pero al tratarse del habla de Málaga donde no es usual el uso de este pronombre, pero optamos por no contemplarlo en el presente estudio.
- **Especificidad.** Ambas guías discriminan entre referentes específicos y no específicos, así se tratan también en este trabajo.
- **Correferencialidad y turno de habla.** En un análisis inicial, estas variables se analizan según las indicaciones de las guías de codificación. Por un lado, correferencial con sujeto, correferencial con objeto, no correferencial y no aplicable; y, por otro lado, cambio de turno y mismo turno. Sin embargo, para relacionarla con variables como la ambigüedad y para establecer comparaciones posteriormente con el estudio de *eye-tracking*, la correferencialidad se recodifica como cambio de referente y mantenimiento del referente.
- **Tiempo verbal y modo.** A pesar de que estas dos variables son consideradas en las guías de codificación, presentan matices diferentes. Por un lado, la guía de 2022 introduce, como novedad, el modo imperativo, ya que este no se considera en la versión de 2011. En relación con los tiempos verbales, nuestra codificación sigue el desglose detallado de las formas simples y compuestas, tal como sugiere la guía de 2022. No obstante, debido a la baja frecuencia de aparición de ciertos tiempos verbales, se opta posteriormente por una simplificación siguiendo el criterio de la guía de 2011, que organiza los tiempos en categorías más amplias como presente, pretérito, imperfecto, futuro morfológico, futuro perifrástico y condicional.
- **Ambigüedad morfológica.** Las dos guías PRESEEA incluyen esta variable para estudiar la ambigüedad producida entre la 1ª y la 3ª persona del singular al presentar formas verbales idénticas en algunos tiempos verbales, como el imperfecto. De esta forma, sin tener en cuenta otros elementos presentes en el entorno discursivo, discriminan entre forma morfológicamente no ambigua y forma morfológicamente ambigua.

En segundo lugar, la guía de codificación más actual propone dos variables nuevas, no consideradas previamente:

- Clasificación semántica Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español (ADESSE). La guía de codificación de 2022 sugiere el uso de la clasificación de ADESSE, que organiza los verbos en torno a seis macrocategorías: procesos mentales, relacionales, materiales, verbales, existenciales y de modulación.
- Tipo de discurso. Pese a que la guía de 2022 recomienda codificar los ejemplos en función a las cinco posibilidades de estilo discursivo —narrativo, argumentativo, explicativo, descriptivo y diálogo rápido—, en este trabajo agrupamos las secuencias explicativas dentro de la tipología discursiva de la argumentación con el fin de poder establecer comparaciones con trabajos previos.

Para este análisis, recuperamos tres variables de la guía de codificación anterior que consideramos que pueden contribuir a la explicación del fenómeno lingüístico analizado. Estas son la progresividad (formas progresivas y no progresivas), perfectividad (formas simples y compuestas) y tipo de cláusula (afirmativa, negativa e interrogativa con y sin partícula), que proporcionan una visión más detallada de los verbos y de la oración analizada.

Además de las variables mencionadas previamente, proponemos el (re)análisis de las siguientes variables:

- Referente del sujeto. Esta variable se refiere al rasgo de animación del sujeto y, en este trabajo, la codificamos según si el referente tiene un rasgo [+humano], [+animado], [+inanimado] o [+neutro]. A pesar de que algunos trabajos dentro de PRESEEA (Claes, 2011; Martínez Lara, 2013; Orozco & Hurtado, 2020, 2021) tratan el rasgo de animación del sujeto, exclusivamente el [+humano], ninguna de las dos guías de codificación previas la consideran entre sus variables a analizar.
- Ambigüedad fonética. La ambigüedad fonética se incorpora para explorar la interacción entre la presencia del sujeto y la elisión de la /-s/ final en las formas verbales, característica propia del habla malagueña. En este trabajo nos basamos en el paradigma de Hochberg (1986) para clasificar la ambigüedad en los siguientes tipos: consideramos formas no ambiguas (Tipo A) a los verbos que no presentan

ambigüedad en su paradigma, como en *comí*, *comiste*, *comió*, *comieron*; en cambio, como formas ambiguas entendemos (Tipo B) tanto donde la ambigüedad potencial surge en las 2ª y 3ª personas si la /-s/ final se elide, por ejemplo, *comes* y *come*, como en (Tipo C) los casos en los que la 1ª y 3ª persona del singular tienen formas verbales idénticas, pero además, pueden converger también con la 2ª persona del singular en caso de elisión de /-s/, como en *comería*, *comerías* y *comería*.

- Ambigüedad morfológica y fonética contextualizada. Aunque partimos de la codificación inicial de estos dos factores con el propósito de establecer comparaciones con trabajos previos, estas variables posteriormente se recodifican bajo el criterio de Silva-Corvalán (1982). Esta recodificación tiene en cuenta elementos como la presencia de pronombres átonos y referencias [+humanas], que desambiguan contextualmente la forma verbal analizada. Ampliar el análisis de estas variables —ambigüedad morfológica y ambigüedad fonética— mediante la consideración del contexto en el que tienen lugar, se aproxima a una visión más natural sobre cómo se integran estas cláusulas en el discurso. De este modo, distinguimos entre formas no ambiguas, aquellas que son contextualmente no ambiguas, y las que se consideran inherentemente ambiguas.
- Subtipos ADESSE. Con el objetivo de profundizar en el análisis de la clase semántica verbal, incorporamos las subcategorías de la clasificación ADESSE, lo que nos permite explorar de manera más precisa del comportamiento de los verbos en relación con la presencia del pronombre personal sujeto. Así, los verbos de procesos mentales incluyen las subclases de percepción, sensación, cognición y elección; los verbos relacionales se dividen en atribución y posesión; los verbos materiales abarcan los subtipos de espacio, cambio, otros hechos y comportamiento; los verbales se subdividen en comunicación, valoración y emisión de sonido; los existenciales comprenden las microcategorías de existencia, fase-tiempo y vida; y finalmente, los de modulación incluyen causación, disposición, aceptación y verbos de apoyo.
- Pronominalidad (formas pronominales y no pronominales) y perífrasis verbal (formas no perífrásticas, perífrasis modales y perífrasis aspectuales) se extraen del trabajo previo realizado en la ciudad de Granada por Manjón-Cabeza Cruz et al. (2016). Consideramos que incluir estas variables puede resultar provechoso para el

estudio del sujeto en la ciudad de Málaga, entre otras razones por la proximidad geográfica de ambas localidades.

Por último, en cuanto a las variables sociales, contemplamos las tres variables de estratificación de la muestra citadas previamente: edad (grupo 1, 2 y 3), sexo (mujeres y hombres) y nivel de estudios (bajo, medio y alto).

2.2.3. Herramientas y análisis

Para la recopilación y organización de la base de datos, hemos empleado diferentes herramientas técnicas que se adaptan a las necesidades específicas de nuestro análisis. La selección de los casos —alternancia de sujeto, presencia y ausencia obligatorias— se realiza manualmente a partir de la lectura de la entrevista y los ejemplos seleccionados para el análisis se recopilan en una hoja de Excel.

El proceso de codificación de las variables lingüísticas está optimizado mediante la utilización de *Autolingua* (Ramajo Ballester, 2022), lo que asegura una mayor sistematicidad en la anotación de las variables. Esta herramienta desarrollada en el lenguaje de *Python* está diseñada específicamente para facilitar la codificación automatizada de determinadas variables lingüísticas. Para su implementación, el programa usa la biblioteca *Verbecc*, que recoge las conjugaciones completas de verbos en español, y permite codificar algunas variables relacionadas con las formas verbales: persona del sujeto, tiempo del verbo, modo verbal, progresividad, perfectividad y ambigüedad morfológica. La herramienta reconoce el verbo conjugado y, según sus características de la forma verbal, cumplimenta de manera automática las variables. Tras ello, se procede a la revisión y codificación manual del resto de variables.

En cuanto a las herramientas empleadas para el desarrollo de los análisis estadísticos, en la primera fase, correspondiente al análisis exploratorio de las variables, se ha utilizado el entorno de R para la elaboración de tablas cruzadas y la aplicación de pruebas estadísticas básicas, como chi-cuadrado, V de Cramer y Phi. Para la segunda fase del análisis, los gráficos de los árboles de decisión se han generado mediante el paquete *rpart* y *rpart.plot*, la visualización de la importancia de las variables se ha llevado a cabo con la herramienta en línea *Language Variation Suite* (Scrivner & Díaz-Campos, 2016) y, con el fin de realizar regresiones logísticas binarias, se ha recurrido al paquete *Rbrul* (Johnson,

2009). Cabe señalar que, en los análisis desarrollados en este trabajo, se emplea el uso de punto como signo separador de los decimales, con el objetivo de seguir la recomendación de la *Ortografía de la lengua española* (RAE, 2024).

2.3. Análisis estadísticos corpus PRESEEA-Málaga

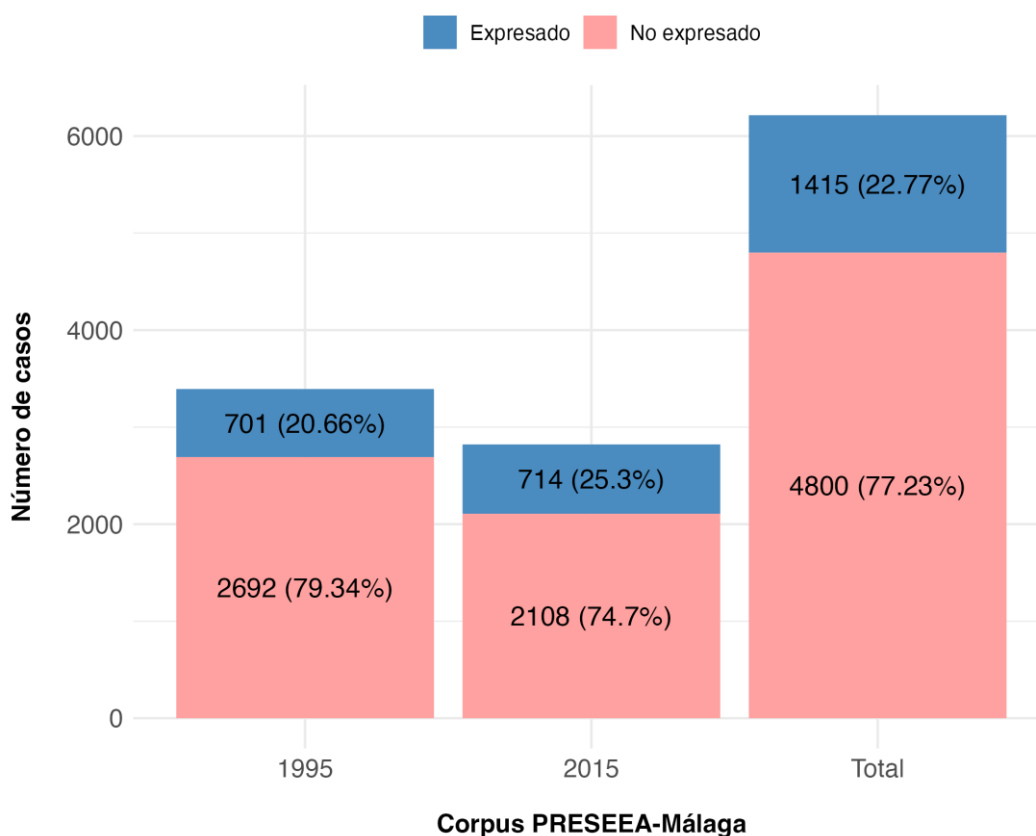
El análisis de nuestros datos se lleva a cabo en tres fases. En primer lugar, un análisis exploratorio de las tasas pronominales en los corpus PRESEEA-Málaga, con el propósito de relacionar los resultados obtenidos en esta ciudad con otras investigaciones previas dentro del ámbito de PRESEEA. En la segunda etapa, se procede a realizar un análisis bivalente orientado a identificar qué variables son significativas, evaluar la fuerza de su asociación y explorar posibles interacciones. Por último, se efectúa un análisis de regresión logística para modelar las relaciones entre la variable dependiente y los factores relevantes identificados en fases previas.

2.3.1. Análisis exploratorio

Antes de proceder al recuento de los casos de presencia y ausencia en lo que puede tener lugar la variación, excluimos aquellos ejemplos de presencia y de ausencia obligatoria del pronombre personal detallados previamente, como frases hechas, usos enfáticos y oraciones impersonales. Además, quedan fuera del análisis las cláusulas con ausencia del sujeto pronominal cuyo referente presenta el rasgo no humano, así como los escasos ejemplos de usted. Estas particularidades se detallarán en el análisis bivalente al tratar la variable de referente del sujeto y persona del sujeto. En definitiva, una vez eliminados estos casos por no presentar variación, los 4147 casos registrados en un primer momento en el corpus de 1995 quedan reducidos a 3393, y los 3419 del corpus de 2015, a 2822.

El Gráfico 1 muestra la proporción de presencia y ausencia del sujeto pronominal en cada corte temporal estudiado en la ciudad de Málaga, y también la suma de casos analizados en esta investigación. A primera vista, se observa una mayor predominancia de la omisión pronominal, que representa entre el 75-80% del total de ocurrencias. Además, resulta significativo que el corpus correspondiente a 2015, a pesar de presentar un menor

número de casos totales, registra una tasa de explicitación pronominal ligeramente mayor, un 25% de presencia, en comparación con el corpus de 1995, donde dicha tasa se sitúa en torno al 20%.



Chi-cuadrado:18.609, g.d.l: 1, p-valor: 1.605e-05, Phi: 0.05471893

Gráfico 1. Recuento de casos en los corpus PRESEEA-Málaga

Antes de realizar los análisis bivalentes, se considera necesario determinar si los corpus se pueden tratar de forma conjunta en la ciudad o si, por el contrario, ambos cortes temporales presentan diferencias significativas que justificasen un análisis independiente. Para ello, se realiza la prueba de chi-cuadrado con el objetivo de evaluar si la distribución de la presencia/ausencia del pronombre en cada corpus difiere estadísticamente en las sincronías estudiadas.

Los resultados estadísticos obtenidos revelan un valor de $p < 0.001$, lo que indica que la probabilidad de que las diferencias observadas dependa del azar es prácticamente nula. En consecuencia, nos lleva a rechazar la hipótesis nula de que no existan diferencias

entre los dos corpus. No obstante, el valor de Phi (0.054) indica que la magnitud de la asociación es muy débil. A partir de estos resultados, se justifica estadísticamente la decisión de llevar a cabo un análisis diferenciado para cada una de las sincronías estudiadas en la ciudad de Málaga, lo que permite una interpretación más precisa y contextualizada de los patrones de variación observados.

Otro aspecto que debe ser considerado antes de proceder con los análisis es la distribución del número de casos y el porcentaje de presencia obtenido por cada individuo. A pesar de que las guías de codificación no establecen un número mínimo de ocurrencias para cada hablante, resulta metodológicamente pertinente verificar que los datos se distribuyen de manera homogénea y que no presentan sesgos atribuibles al volumen de casos por informante.

El Gráfico 2 muestra cómo se distribuye el número de casos en el eje horizontal, respecto al porcentaje de presencia situado en el eje vertical. Además, los puntos representados en el gráfico hacen referencia a las características macrosociales de cada individuo, según la forma determina el grupo de edad y el color el sexo.

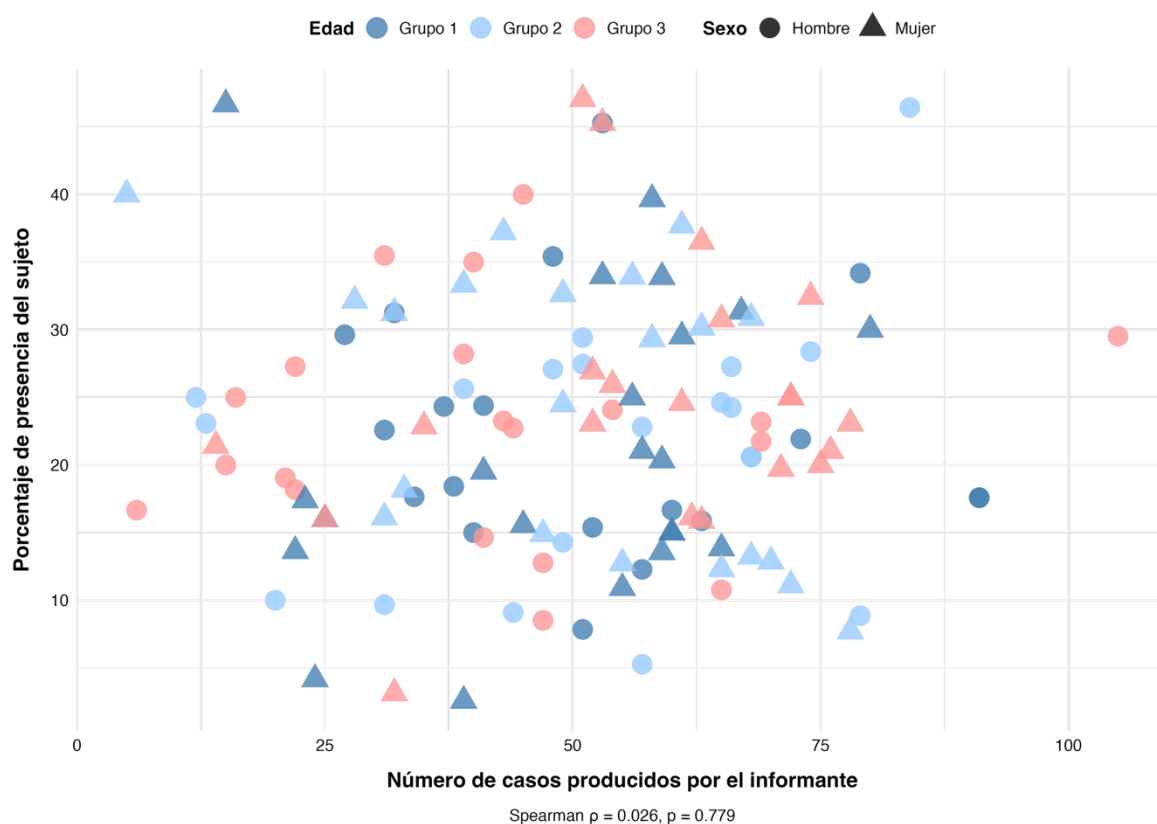


Gráfico 2. Distribución de casos por porcentaje de presencia

A partir de la visualización de esta representación gráfica, la distribución de la nube de datos no parece indicar ningún patrón y, de esta forma, podríamos afirmar que no existe condicionamiento alguno entre el número de casos aportados por cada individuo y su porcentaje de expresión. Esta afirmación se ve respaldada por los resultados del coeficiente de correlación de Spearman (ρ 0.026), que indica la ausencia absoluta de una correlación entre las variables consideradas (valor de p 0.779). En consecuencia, estos datos pueden ser analizados con garantías de representación individual sin riesgo de distorsión en los resultados por sesgos.

2.3.2. Análisis bivariante

A continuación, se describen los resultados obtenidos en ambos corpus, organizados según el tipo de variable independiente considerada: aquellas que remiten al sujeto de la oración o al verbo que lo acompaña, las que aluden a factores de índole pragmático-discursivo, o bien las variables de naturaleza social. Además del análisis bivariante entre la variable dependiente —ausencia o presencia del sujeto pronominal— y cada una de las variables lingüístico-sociales, presentamos algunas interacciones relevantes entre factores, con el propósito de conocer en mayor detalle cómo aumenta o disminuye el uso del sujeto pronominal en determinados contextos.

2.3.2.1. Variables referentes al sujeto de la oración

En cuanto a los factores relacionados directamente con el sujeto de la oración, se procede a continuación al análisis de las tres variables fundamentales: animación del referente, persona del sujeto y especificidad.

Animación del referente

El referente al que remite la desinencia verbal de persona o el pronombre personal sujeto ocupa una posición en la jerarquía de animación. En 1-5 se muestran ejemplos con ausencia y presencia del pronombre personal tónico según el rasgo de las entidades a las que aluden. En (3) la desinencia verbal de *tengo* y en (4) *tú* y la forma verbal de *sabes* aluden al hablante y al oyente respectivamente y, por tanto, a entidades marcadas por el rasgo [+ humano]. En (5) el sujeto tácito de *se queda* remite anafóricamente a *el pulpo*,

caracterizado por el rasgo [+animado]. El sujeto omitido del ejemplo (6) que acompaña al verbo *era* está vinculado anafóricamente a *un disco*, cuyo referente es inanimado. Por último, el sujeto tácito de *me cuesta* (7) hace referencia a *ir al pueblo*.

- (3) [...] *ya ø tengo más ganas de estar en mi casa que de estar allí* [MA_M12_729].
- (4) [...] *¿tú sabes dónde estaba la <nombre propio> Lauri </nombre propio>, allí en Pedregalejo?* [MA_M32_722].
- (5) [...] *y metí yo las manos entre< (:)> las las rocas y< (:)> lo lo ø saqué porque el pulpo estaba suelto, si ø se queda enganchado en las piedras no <(:)> no hay forma de <(:)> sacarlo.* [MA_H21_734].
- (6) [...] *me ø escuché un disco una vez de él y ø era< (:)> versiones <simultáneo> de otras canciones* [MA_H11_916].
- (7) [...] */ antes era rara la semana que no iba yo por allí / pero ahora cuan <palabra cortada> desde que faltaron ellos ya <vacilación> me ø cuesta más trabajo ya parece ser como si no me< (:)> interesara mucho mi pueblo me da pena y todo ¿no?* [MA_M31_077].

Por lo general, mientras que el sujeto tácito se usa sin limitación alguna, sea cual sea el rasgo de la entidad designada, las formas pronominales tónicas con función de sujeto parecen tener restringidas las posibilidades de emplearse cuando el referente designado no es humano (Enríquez, 1984; Fernández Soriano, 1999; Luján, 1999; RAE & ASELE, 2009). Son factibles, sin embargo, expresiones como *él nos enseña*, si se aplican a un libro (RAE & ASELE, 2009, p. 2549) por personificación. De la misma manera, el cuantificador *todo* (8) y el adverbio focalizador *solo* (9) parecen anular esta restricción (Fernández Ramírez, 1987):

- (8) *Las aspiraciones del hombre son infinitas [...] ¿Pero son todas ellas legítimas?*

- (9) [...] una intervención es muchas veces una consecuencia
tan fácil de los hechos anteriores, que casi se puede decir
que se desprende ella sola sin esfuerzo

En los trabajos previos sobre la variación del pronombre personal sujeto, no suele informarse sobre los casos en los que los pronombres remiten a referentes no humanos. Por ejemplo, en Claes (2011, p. 195) se lee: “excluimos de la investigación todos los contextos en que el pronombre nulo está bajo el control de un sujeto léxico inanimado. En estos casos, se ha demostrado que la aparición de un pronombre pleno hace la frase agramatical”. Esta información sí la recoge Enríquez (1984, p. 176-179) en su estudio sobre el español hablado en Madrid, pero solo encuentra cinco casos (un porcentaje global de tan solo un 0.15%) de presencia pronominal con referentes no humanos. En la Tabla 2 exponemos los datos de presencia y ausencia de sujetos pronominales según el rasgo de animación de sus referentes.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Humano	2692	79.34	701	20.66	3393	85.64	2108	74.70	714	25.30	2822	84.29
Animado	14	100	0		14	0.35	3	100	0		3	0.09
Neutro	391	100	0		391	9.87	403	100	0		403	12.04
Inanimado	164	100	0		164	4.14	120	100	0		120	3.58
Total	3261	82.31	701	17.69	3962	100	2634	78.67	714	21.33	3448	100

Tabla 2. Presencia/ausencia de sujetos pronominales según el rasgo de animación del referente

En ambos corpus los sujetos que poseen referentes con rasgo [+humano] son los más numerosos. Los casos en los que la desinencia verbal de persona alude a referentes no humanos —animado, neutro e inanimado suman un total de 569 casos en el corpus de 1995, así como 526 casos en el corpus del 2015— no registran ni un solo ejemplo de explicitación pronominal. Con el fin de observar estudiar los casos donde se produce la variación, solo consideramos en el análisis los referentes que presentan el rasgo [+humano] (3393/3962 en el corpus de 1995 y 2822/3448 en el corpus de 2015), ya que en estos casos se produce la alternancia entre ausencia y presencia del pronombre como de persona del

sujeto. Todos los ejemplos encontrados en los corpus con referentes animados, inanimados y neutros corresponden a casos de 3ª persona singular y plural.

Persona del sujeto

La persona gramatical constituye un factor omnipresente en la literatura previa sobre el sujeto pronominal y, además, es uno de los más influyentes en la expresión o no del sujeto pronominal. Excluidos del cómputo los cinco ejemplos del pronombre de cortesía *usted(es)* en el corpus de 1995 por no presentar alternancia, la cuantificación de la presencia/ausencia del pronombre sujeto de acuerdo con la persona gramatical es la reflejada en la Tabla 3. Además, la prueba de chi-cuadrado indica una asociación muy significativa entre las variables en ambos corpus (300.468 para el corpus de 1995 y 309.822 para el corpus de 2015) con un valor de p extremadamente bajo. La V de Cramer también muestra que la asociación es moderada en el corpus de 1995 (0.297), pero se incrementa en el corpus de 2015 (0.331).

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
1ª singular	894	66.97	441	33.03	1335	39.35	702	58.89	490	41.11	1192	42.24
2ª singular	331	72.11	128	27.89	459	13.53	311	75.85	99	24.15	410	14.53
3ª singular	430	87.04	64	12.96	494	14.56	267	86.13	43	13.87	310	10.99
1ª plural	519	92.51	42	7.49	561	16.53	338	88.48	44	11.52	382	13.54
2ª plural	9	90	1	10	10	0.29	20	100	0	0	20	0.71
3ª plural	509	95.32	25	4.68	535	15.74	470	92.52	38	7.48	509	18.00
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 300.4685, g.d.l: 5, p-valor: 7.942052e-63, V de Cramer: 0.2975826						Chi-cuadrado: 309.8227, g.d.l: 5, p-valor: 7.736406e-65, V de Cramer: 0.3313432					

Tabla 3. Presencia/ausencia de sujetos pronominales según la persona del sujeto

En este primer análisis podemos constatar los siguientes tres hechos:

- La ausencia del sujeto pronominal predomina en todas las personas, con porcentajes que oscilan entre el 59% aproximadamente para la persona 1ª persona hasta 100% de los casos para la 2ª persona del plural en el corpus de 2015, o entre

casi el 67% para la primera del singular y el 96% tercera del plural en el corpus de 1995.

- La 1ª persona del singular es la que presenta más casos totales (1335 en el corpus del 1995 y 1192 en el corpus de 2015) y, además, es la que más favorece la presencia del sujeto pronominal (33% y 41%), seguida de la 2ª persona del singular (28% y 24%).
- La 2ª persona del plural no alcanza ni el 1% de casos del total del corpus. Esto puede deberse a la dinámica de la entrevista, donde hay un entrevistador y un informante, y no varios entrevistadores y varios entrevistados.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
1ª persona	1413	74.53	483	25.47	1896	55.88	1040	66.07	534	33.93	1574	55.78
2ª persona	340	72.49	129	27.51	469	13.87	331	76.98	99	23.02	430	15.24
3ª persona	939	91.34	89	8.66	1028	30.30	737	90.10	81	9.90	818	28.99
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 130.5655, g.d.l: 2, p-valor: 4.446975e-29, V de Cramer: 0.1961653						Chi-cuadrado: 165.7681, g.d.l: 2, p-valor: 1.009074e-36, V de Cramer: 0.2423661					

Tabla 4. Presencia/ausencia de sujetos pronominales según persona del sujeto

En cuanto a la persona gramatical, se constata que el uso de la 1ª persona predomina en ambos corpus (55%), seguida de la 3ª persona (30%) y, en menor medida, la 2ª persona (15%). Estos resultados (Tabla 4) van en la línea de lo esperado por Givón (1984, p. 153), donde el uso de la 1ª persona es la más frecuente, ya que el hablante tiende a tomarse a sí mismo como punto de referencia universal y como argumento más presupuesto, lo que contribuye a manifestar el carácter egocéntrico del individuo en el discurso.

Por otro lado, la Tabla 5 considera únicamente el número de la persona gramatical. De esta comparativa, se revela que una mayor predominancia de los casos en singular (67%) que en plural (32%), así como una mayor tasa de explicitación pronominal en las formas del singular (entre un 27% y un 33%).

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Singular	1655	72.33	633	27.67	2288	67.43	1280	66.95	632	33.05	1912	67.75
Plural	1037	93.85	68	6.15	1105	32.57	828	90.99	82	9.01	910	32.25
Total	2692	79.34	701	20.66	3382	100	2108	74.80	714	25.20	2833	100
	Chi-cuadrado: 209.0579, g.d.l: 1, p-valor: 2.204941e-47, Phi: 0.2482227						Chi-cuadrado: 187.3160, g.d.l: 1, p-valor: 1.225237e-42, Phi: 0.2576374					

Tabla 5. Presencia/ausencia de sujetos pronominales según número

Este patrón de uso se detecta también de manera recurrente en estudios previos como el de Lastra y Martín Butragueño (2015) y el de Orozco y Hurtado (2021). Para explicar esta tendencia, Cameron (1993, p. 328) lo justifica de la siguiente manera:

“If we conceive of plural subjects as sets, we find that discourse is typically structured so that the great majority of plural subjects occur in contexts where their set members are either explicitly or inferably present within the immediately preceding discourse. Such contexts favor null subject expression. Therefore, plural subjects are frequently null overall.”

Especificidad

La persona es una categoría déictica caracterizada por señalar tanto a uno de los participantes en el discurso como a alguien (o a algo) distinto de ellos presente en el contexto lingüístico o extralingüístico. Por lo tanto, los pronombres en función sujeto o las desinencias personales del verbo pueden tener una referencia que permita una interpretación específica o inespecífica en función del referente en cuestión. De manera que consideramos que un sujeto “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” (RAE & ASELE 2009, p. 1134). En este sentido, el sujeto nulo en expresiones como *sabes* (10) y *llega* (12), así como los sujetos pronominales *yo* y *ella* de los verbos *voy* (11) y *dice* (13) respectivamente, constituyen ejemplos de interpretación específica:

(10) *no sé si Ø sabes lo que es una orza [MA_M21_041].*

(11) *cuando haya teatro que a mí me guste y eso yo voy [MA_M21_008].*

- (12) *esa mujer sale de su casa dice ella a las dos y media/ y ø
llega a su casa a las diez y media o las once de la noche
[MA_M21_008].*

No obstante, ni las desinencias personales del verbo ni los pronombres personales tónicos refieren siempre a una entidad específica (Casielles Suárez, 1996). La 2ª persona del singular (13) puede emplearse para referirse no al interlocutor directo, sino a un conjunto de individuos no determinados, y también puede aludir al propio hablante escudándose bajo un *tú* inespecífico (14). Además, por desconocimiento del referente del sujeto o por no considerarse oportuno hacerlo expreso, se usa en español el verbo en 3ª persona del plural (15) sin sujeto explícito con el valor de persona indeterminada o con valor genérico (Jaeggli, 1986, p. 46-48; RAE & ASELE, 2009, p. 2553-2554). En determinados casos, la 3ª persona del singular remite a un antecedente cuya interpretación es inespecífica. Así, el sintagma nominal definido *la gente* en el ejemplo (17) tiene valor genérico y es el antecedente del sujeto tácito de *pide*. En algunos usos se hace difícil determinar si el pronombre o la desinencia de persona con interpretación definida tiene lectura específica o inespecífica. En el ejemplo (16), el sujeto tácito alude a ciertos individuos particulares de los que el hablante desconoce su identidad, pero de la situación comunicativa creada se infiere que es su hipotético interlocutor.

- (13) *yo eso de que la gente no se pueda quitar la camiseta en
agosto no lo veo normal, si ø estás en una feria, claro, si tú
vas por la calle a comprando, yo entiendo perfectamente
que cada uno lleve su ropa y pa[r]a entrar a las tiendas se
ø tenga que poner una ropa [MA_M11_038].*

- (14) *yo estaba en un colegio religioso y hombre, después te ø
alegras porque ø tienes unos principios, ¿no?
[MA_M31_025].*

- (15) *[...] entonces había que llen<palabra cortada>,
rellenarlos, to[d]os los casilleros que había/ garbanzos,
habichuelas, to[d]o, lentejas, y entonces cuando ø querían
“medio kilo de eso”, o “un cuarto de kilo [MA_M31_016].*

- (16) *¡anda! pues entonces mira cómo no ha veni<[d]>o la profesora y ø nos van a poner de una clase y otra </sic> pos ø vámonos </estilo directo> [MA_M21_105].*
- (17) *los días diarios, está eso muerto, vamos [...] más que los fines de semana la gente pues descansa y ø pide <simultáneo>más comida a domicilio [MA_M11_053].*
- (18) *una persona sufre un accidente y ø tiene una deformación en la cara [MA_H13_035].*
- (19) *[...] ø vas al departamento de convalidación, ø entregas una instancia, esta asignatura de <nombre propio>París<nombre propio> se parece a esta de la <nombre propio>UMA<nombre propio> de <nombre propio>Málaga<nombre propio>, pues me la miráis y me decís si firmáis la convalidación. [MA_M13_036].*

Los resultados generales de la variable especificidad muestran que es significativa en ambos corpus, aunque especialmente en el corpus de 2015, tiene una mayor fuerza de asociación (V de Cramer: 0.198). En cuanto a la distribución de los datos, se observa una mayor tendencia a la lectura específica (83% y 76%) que inespecífica (16% y 23%). En este contexto, la aparición del pronombre es menos frecuente en casos inespecíficos, ya que solo alcanzan un 13% de presencia frente a un 22% en casos con referente específico en el corpus de 1995, y un 9% en comparación con un 30% en corpus de 2015.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
No específico	495	86.54	77	13.46	572	16.86	602	90.25	65	9.75	667	23.64
Específico	2197	77.88	624	22.12	2821	83.14	1506	69.88	649	30.12	2155	76.36
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 21.22468, g.d.l: 1, p-valor: 4.084713e-06, Phi: 0.07909128						Chi-cuadrado: 110.7606, g.d.l: 1, p-valor: 6.676611e-26, Phi: 0.1981135					

Tabla 6. Presencia/ausencia de sujetos pronominales según la especificidad

La Tabla 7 recoge los resultados de la interacción de la categoría de persona con la interpretación específica o inespecífica de la misma. Se observa que únicamente en la 2ª y

en la 3ª persona se registran interpretaciones tanto específicas como inespecíficas, mientras que, en la 1ª persona del singular y del plural, únicamente se encuentran casos con un referente específico.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
1ª sing. específico	894	66.96	441	33.03	1335	39.34	702	58.89	490	41.10	1192	42.23
2ª sing. específico	121	70.34	51	29.65	172	5.06	63	64.94	34	35.05	97	3.43
2ª sing. inespecífico	210	73.17	77	26.82	287	8.45	248	79.23	65	20.76	313	11.09
3ª sing. específico	418	86.72	64	13.27	482	14.20	223	83.83	43	16.16	266	9.42
3ª sing. inespecífico	12	100	0	0	12	0.35	44	100	0	0	44	1.55
1ª pl. específico	519	92.51	42	7.48	561	16.53	338	88.48	44	11.43	382	13.53
2ª pl. específico	9	90	1	10	10	0.29	17	100	0	0	17	0.60
2ª pl. inespecífico	0	0	0	0	0	0	3	100	0	0	3	0.10
3ª pl. específico	236	90.42	25	9.57	261	7.69	163	81.09	38	18.90	201	7.12
3ª pl. inespecífico	273	100	0	0	272	8.01	307	100	0	0	307	10.87
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100

Tabla 7. Presencia/ausencia de sujetos pronominales según la persona del sujeto y la especificidad

Los casos de 2ª persona del singular y de 3ª persona del plural en contextos no específicos siguen la tendencia afirmada por Cameron (1993, p. 170), quien ya detecta que los efectos de referencia inespecífica son diferentes para el singular y el plural. Por ende, la 2ª persona del singular es la única que presenta casos de ausencia y presencia del pronombre personal sujeto en contextos donde el sujeto es no específico, pero en el resto de las personas gramaticales solo se observan casos inespecíficos con ausencia del sujeto pronominal.

En cuanto a la 2ª persona del singular, evidenciamos diferencias tanto en el uso como en la tasa de aparición del sujeto pronominal según la especificidad del referente. En ambos corpus la presencia del pronombre es superior con referentes específicos (30% en el corpus de 1995 y 35% en el corpus de 2015) que con los inespecíficos (20-26%, respectivamente), y, además, la distribución total de casos refleja una proporción considerablemente mayor en interpretaciones inespecíficas de la segunda del singular

(62% en el corpus de 1995, 76% en el corpus de 2015) que de lecturas específicas. Cabe destacar que las diferencias de uso de la 2ª persona del singular solo logran ser significativas en el corpus de 2015 (valor de p : 0.006 y V de Cramer: 0.135), en contraste con el corpus de 1995 que no resultan significativas (valor de p : 0.585 y V de Cramer: 0.025), lo que puede indicar que se haya producido un cambio en el uso de la 2ª persona según su especificidad en las dos sincronías estudiadas.

Por otro lado, los 56 ejemplos encontrados de 3ª persona del singular en nuestros corpus se manifiestan con frecuencia en estructuras coordinadas copulativas y, además, son correferenciales con una persona genérica mencionada previamente como *alguien*, *gente* o *persona*, como se muestra en los ejemplos anteriores del (15) al (19).

Por último, los casos de 3ª persona del plural en los que se omite el pronombre pueden ser tanto genéricos como indeterminados. En cuanto a los sujetos indeterminados, el verbo señala el número de individuos a los que hace referencia, por lo que podría tratarse de un único individuo y el uso del [+plural] del verbo sería puramente una formalidad gramatical que no indica información sobre el número de referentes del sujeto. En este caso, una oración con plural indeterminado tendría la misma estructura que la oración en singular; por ejemplo: *alguien llama a la puerta* (Jaeggli, 1986, p. 46-48). En cambio, los sujetos genéricos no están claramente definidos, pero los referentes forman parte del conocimiento compartido por los interlocutores. Además, suelen hacer referencia a un elemento mencionado previamente en el discurso, con el que establece una relación de correferencialidad como el ejemplo (16).

2.3.2.2. Variables referentes a cuestiones pragmático-discursivas

En este apartado abordamos el análisis de las variables relacionadas con aspectos pragmático-discursivos: correferencialidad, turno conversacional, tipo de discurso, tipo de cláusula, ambigüedad morfológica y ambigüedad fonética potencial.

Correferencialidad

La correferencialidad se posiciona como una de las variables más significativas estadísticamente en los estudios previos. En los primeros trabajos sobre la variación del sujeto pronominal, Silva-Corvalán (1982) discrimina entre el mantenimiento y el cambio

de referente con sujeto o con objeto y, de esta forma, la misma referencia constituye información antigua que se supone que sigue en la mente del receptor durante el siguiente enunciado, como consecuencia de haber sido mencionada previamente o porque está implícita en el discurso. Cabe recordar que en este trabajo se considera como unidad de análisis la cláusula compuesta por el sujeto, el verbo y los complementos, sean o no argumentales; por lo tanto, el análisis de la correferencialidad se codifica respecto a la cláusula inmediatamente anterior.

Las guías de codificación de PRESEEA distinguen entre los distintos de cambio o de mantenimiento del referente previo. Por ello, consideramos como correferencialidad con sujeto aquella cláusula analizada que remite al sujeto precedente, como se ilustra en el ejemplo (20), donde el sujeto omitido el verbo *llega* hace referencia a *ella*. El mantenimiento del referente anterior también puede darse con el objeto, como es el caso del ejemplo (21), en el que el sujeto de *tenía que andar* presenta continuidad referencial con el objeto de la cláusula anterior (*me*). Por el contrario, puede producirse un cambio de referencia respecto a la presentada en la cláusula precedente, como en el ejemplo (22).

(20) *esa mujer sale de su casa dice ella / a las dos y media / y ø
llega a su casa a las diez y media o las once <[d]><[e]>
la noche [MA_M21_008].*

(21) *me dijeron los médicos que ø tenía que andar siquiera una
hora al día [MA_M31_077].*

(22) *cuando ya nació el niño, ø volvieron otra vez a su casa.
[MA_M32_043].*

La Tabla 8 evidencia que la variable es significativa en ambos corpus, por lo que se puede confirmar que la relación observada no es aleatoria y la fuerza de asociación es más elevada en el corpus de 2015 (V de Cramer: 0.159). La distribución de los casos se concentra principalmente en dos niveles: un 40-45% del total de ejemplos corresponde a la correferencia con el sujeto previo, mientras que entre un 46% y un 52% se clasifican como no correferenciales. Las subvariables restantes como no aplicable (2%) o correferencial con objeto (6%) presentan una incidencia marginal.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Corref. sujeto	1291	83.83	249	16.17	1540	45.39	926	83.20	187	16.80	1113	39.44
Corref. objecto	192	89.30	23	10.70	215	6.34	140	72.92	52	27.08	192	6.80
No Corref.	1167	73.67	417	26.33	1584	46.68	1009	68.59	462	31.41	1471	52.13
No aplicable	42	77.78	12	22.22	54	1.59	33	71.74	13	28.26	46	1.63
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 63.06861, g.d.l: 3, p-valor: 1.298379e-13. V de Cramer: 0.1363373						Chi-cuadrado: 72.0997, g.d.l: 3, p-valor: 1.515509e-15. V de Cramer: 0.159841					

Tabla 8. Presencia/ausencia de sujetos pronominales según la correferencialidad

Con el objetivo de simplificar la interacción de estas variables con otros factores en las siguientes fases del análisis, reagrupamos la correferencialidad en dos categorías principales: continuidad y cambio referencial¹, tal y como se presentan en la Tabla 9. Además, la aplicación de la prueba de chi-cuadrado revela que no existen diferencias significativas entre no correferencialidad y no aplicable en ambos corpus (valor de p 0.605 para el corpus de 1995 y valor de p 0.770 para el corpus de 2015).

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Continuidad	1483	84.50	272	15.50	1755	51.72	1066	81.69	239	18.31	1305	46.24
Cambio	1209	73.81	429	26.19	1638	48.28	1042	68.69	475	31.31	1517	53.76
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
Chi-cuadrado: 58.43649, g.d.l: 1, p-valor: 2.099609e-14. Phi: 0.1312351							Chi-cuadrado: 62.02061, g.d.l: 1, p-valor: 3.398806e-15. Phi: 0.1482482					

¹ En los análisis posteriores, utilizamos la simplificación de la variable a continuidad y cambio referencial, con el fin de obtener una visión más general cuando se relacione la correferencialidad con otras variables, como la ambigüedad fonética y morfológica del verbo, el turno de habla o el tipo de discurso. Además, el estudio experimental de *eye-tracking* (Capítulo III) considera solo la comparación entre la continuidad y el cambio de referencia, por lo que simplificar esta variable facilita la posterior unificación de conceptos.

Tabla 9. Presencia/ausencia de sujetos pronominales según la correferencialidad recodificada: continuidad y cambio referencial

Los resultados presentados en la Tabla 9 se alinean con investigaciones previas ante la obtención de un alto porcentaje de casos correferenciales (51% y 46%), lo cual contrasta con la baja presencia del pronombre (15% y 18%). Lastra y Martín Butragueño (2015, p. 14) consideran que esta diferencia puede explicarse en función de que, a mayor cohesión textual, menor es la necesidad de la expresión explícita del pronombre, ya que múltiples mecanismos como el conocimiento general y contextual compartido permiten que el interlocutor pueda identificar al referente en cuestión. Además, el discurso suele estar acompañado de recursos que aseguran la coherencia narrativa, de modo que, aunque la oración pueda ser ambigua desde el punto de vista sintáctico, no presenta ningún problema desde la perspectiva discursiva o lógica (Lastra & Martín Butragueño, 2015, p. 15).

Turno de habla

En el contexto de una entrevista es habitual que el discurso del informante se vea interrumpido por el entrevistador o por acontecimientos ajenos a la conversación, como la aparición de un tercer participante o ruidos ambientales. Tras estas interrupciones, el hablante puede reiterar el sujeto pronominal para asegurarse de que sigue presente en la mente del oyente (Bentivoglio, 1987, p. 38).

Esta variable codifica si la unidad de análisis se encuentra inmersa en el discurso continuo del hablante o si, por el contrario, el entrevistado retoma su turno de palabra tras haber sido interrumpido. Cabe destacar que las interjecciones del entrevistador no las consideramos como un cambio de turno conversacional, puesto que se interpretan como estímulos comunicativos y no como intervenciones propiamente dichas (Bentivoglio, 1987, p. 40). De tal forma, el ejemplo (23) codifica como se produce el cambio de turno conversacional al comienzo de la cláusula *yo creo*, puesto que tiene lugar después de la intervención del entrevistador. En cambio, el ejemplo (24) muestra como varias cláusulas (*yo tengo, yo fui, estuvimos, pregunté y conocían*) se encuentran dentro del mismo turno de habla.

(23) E: *¿Pero esos niños en concreto o de esa edad?*

I: yo creo que es por esa edad, habrá de to[d]o, que en ese barrio hay de to[d]o. [MA_M11_057].

- (24) *I: bueno, mira, yo tengo mi hija ahora en Tarragona, yo fui a, estuve en febrero en Tarragona y en un hotel donde ø estuvimos le ø pregunté si ø conocía Málaga y la ø conocían. [MA_M31_015].*

En cuanto a los datos de nuestros corpus, la Tabla 10 presenta los resultados generales de la variable turno de habla. Aunque la variable es significativa en los dos cortes temporales, la fuerza de asociación es débil (Phi: 0.066 para el corpus de 1995 y Phi: 0.094 para el corpus de 2015). De los resultados obtenidos cabe recalcar, en primer lugar, que la mayoría de los casos se encuentran en secuencias de mismo turno conversacional (72% y 82%), lo cual se puede atribuir a que, en las entrevistas semidirigidas, el entrevistador procura interrumpir lo menos posible y genera esencialmente monólogos del informante (Guerrero González, 2017, p. 83). Es plausible suponer que, en contextos de habla cotidiana, la espontaneidad de la situación comunicativa y la no jerarquización de la toma de los turnos conversacionales provocarían que los porcentajes estuviesen más equilibrados.

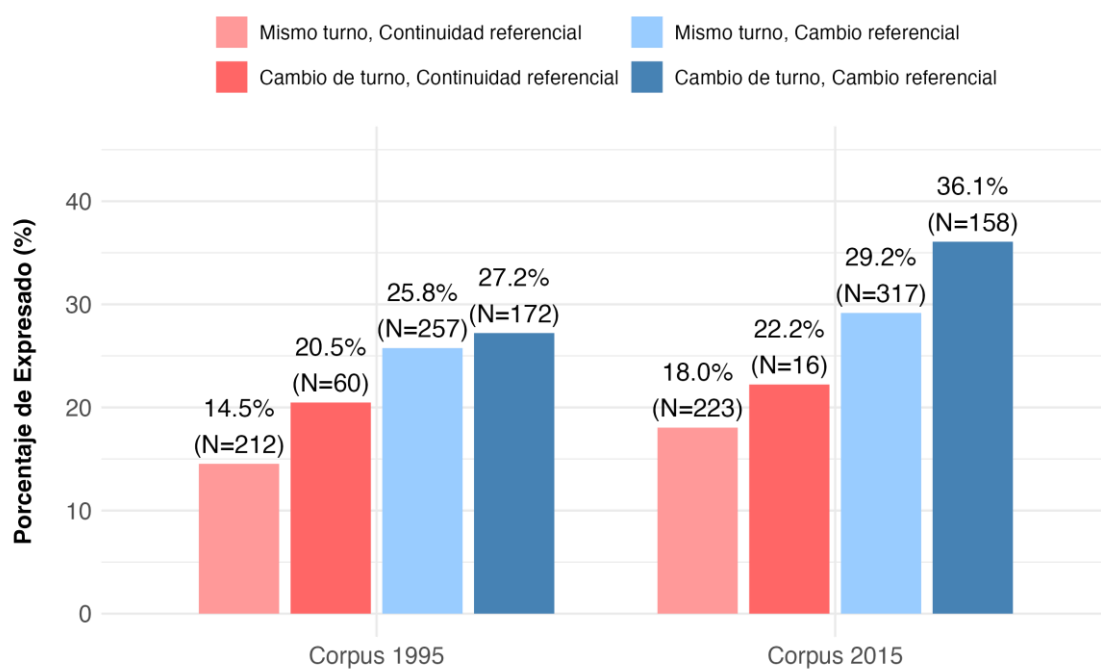
	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Mismo turno	1998	80.99	469	19.01	2467	72.71	1773	76.65	540	23.35	2323	81.96
Cambio turno	694	74.95	232	25.05	926	27.29	335	65.82	174	34.18	509	18.04
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 14.63333, g.d.l: 1, p-valor: 0.0001305844, Phi: 0.06567192						Chi-cuadrado: 25.36013, g.d.l: 1, p-valor: 4.756471e-07, Phi: 0.09479758					

Tabla 10. Presencia/ausencia de sujetos pronominales según el turno de habla

En segundo lugar, la Tabla 10 evidencia un aumento de la presencia del pronombre personal cuando se produce un cambio de turno conversacional (25% en el corpus de 1995 y un 34% en el corpus de 2015), en comparación con aquellas situaciones en las que el hablante continúa su intervención sin ser interrumpido. En estas situaciones, la presencia del pronombre ronda entre el 19-23%. Estos resultados coinciden con trabajos previos, como el de (Li & Thompson, 1981, p. 333), que demuestran que el cambio de turno

conversacional promueve la presencia del pronombre cuando las cláusulas analizadas se encuentran en diferentes turnos de habla.

Mediante la interacción entre las variables pragmático-discursivas se puede analizar el papel que juegan en conjunto la correferencialidad y el turno conversacional en la variación del pronombre personal sujeto. El Gráfico 3 presenta los porcentajes de presencia según los diferentes contextos analizados considerando las variables de correferencialidad y turno de habla. A primera vista, se observa una mayor explicitación del pronombre en contextos no correferenciales —tanto en situaciones de continuidad como de cambio de turno—, lo que sugiere que la influencia de la correferencialidad consigue ser más determinante que la del propio turno conversacional.



Corpus PRESEEA-Málaga

Chi-cuadrado 1995: 65.013, g.d.l: 3, p-valor: 4.983e-14, V de Cramer: 0.1384234

Chi-cuadrado 2015: 70.116, g.d.l: 3, p-valor: 4.03e-15, V de Cramer: 0.1576272

Gráfico 3. Presencia de sujetos pronominales según la correferencialidad y el turno de habla

En ambos corpus se observa una tendencia similar, aunque las diferencias son más pronunciadas en el corpus de 2015, donde se constata que la presencia del pronombre es menor cuando se produce continuidad referencial en la misma intervención (14-18%). Esta explicitación pronominal aumenta ligeramente (entre un 4% y un 6%) cuando se registra un cambio de turno, y alcanza niveles significativamente superiores en contextos de

cambio referencial, tanto en el mismo turno (25-29%) como, especialmente, cuando se combinan el cambio de turno y el cambio de referente (27-36%).

Además, esta combinación de variables presenta una influencia distinta en los dos cortes temporales estudiados, ya que en el corpus de 1995 observamos que cuando se mantiene la continuidad referencial y se produce un cambio en el turno de habla, las diferencias en la presencia del pronombre son significativas (valor de p : 0.012). En cambio, los contextos de cambio de referente se ven significativamente afectados por el cambio o mantenimiento del turno de palabra en el corpus de 2015 (valor de p : 0.011). Por ende, estos resultados indican que la correferencialidad ejerce un mayor efecto sobre la presencia del pronombre, y que, además, la interacción cambio de turno de habla y cambio de referente promueve una mayor explicitación del sujeto pronominal.

Tipo de discurso

Durante las entrevistas semidirigidas que conforman el corpus se introducen diferentes temas con el fin de incentivar al informante a usar diferentes estilos discursivos en sus intervenciones, de manera que se pueden distinguir secuencias argumentativas, narrativas, descriptivas y de diálogo rápido.

La Tabla 11 revela que, en el corpus de 1995 predominan las secuencias narrativas y argumentativas, mientras que en el corpus de 2015 se observan mayoritariamente narraciones y descripciones.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Diálogo	78	77.23	23	22.77	101	2.98	273	65.94	141	34.06	414	14.67
Descripción	136	81.93	30	18.07	166	4.89	666	76.55	204	23.45	870	30.83
Narración	1643	81.22	380	18.78	2023	59.62	814	77.08	242	22.92	1056	37.42
Argumentación	835	75.70	268	24.30	1103	32.51	355	73.65	127	26.35	482	17.08
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
Chi-cuadrado: 14.19936, g.d.l: 3, p-valor: 0.002645978. V de Cramer: 0.06469078							Chi-cuadrado: 21.83426, g.d.l: 3, p-valor: 7.062126e-05. V de Cramer: 0.08796111					

Tabla 11. Presencia/ausencia de sujetos pronominales según el tipo de discurso

Sin embargo, las situaciones de diálogo rápido son las que alcanzan una presencia más elevada del pronombre (22% en el corpus de 1995 y un 34% en el corpus de 2015), en contraste con los pasajes narrativos que obtienen los porcentajes más bajos de presencia (18% en el corpus de 1995 y un 22% en el corpus de 2015). Nuestros resultados van en consonancia con lo expuesto por Travis (2007, p. 131), quien afirma que las situaciones conversacionales, al estar sometidas a cambios rápidos de tema, pueden requerir un mayor empleo de los pronombres personales que las secuencias narrativas, donde se espera una mayor continuidad referencial.

Además, otro tipo de discurso que favorece la aparición del pronombre es la argumentación con unos porcentajes de presencia que rondan el 24-26%. Estos datos respaldan las conclusiones de Lastra y Martín Butragueño (2015), quienes afirman que este estilo discursivo enfatiza la posición de los diferentes puntos de vista mediante la presencia del pronombre. Al observar de manera conjunta las variables de tipo de discurso, turno de habla y correferencialidad (Gráfico 4), los valores estadísticos obtenidos reflejan que esta combinación de variables es significativa para ambos corpus con un valor de p muy bajo y una V de Cramer (0.157 en 1995 y 0.177 en 2015) que supera la fuerza de asociación de las variables por separado.

En el Gráfico 4 identificamos algunas similitudes y disparidades entre los dos corpus analizados, como que las secuencias argumentativas ante situaciones de cambio de turno conversacional y cambio de referente son las que requieren una mayor tasa de explicitación pronominal en ambas sincronías (un 32% en el corpus de 1995 y un 40% en el corpus de 2015). Esto también sucede en los pasajes narrativos (33%) y de diálogo rápido (37%) en el corpus de 2015. En cuanto a la narración y el diálogo rápido, en el corpus de 1995 presentan un patrón diferente, ya que, tanto cuando se produce el cambio de turno (22-27%) en contextos de continuidad referencial, como cuando se cambia de referente en el mismo turno de habla (25-27%), las tasas de presencia del pronombre son más elevadas que en situaciones de no correferencialidad y de cambio en el turno conversacional.

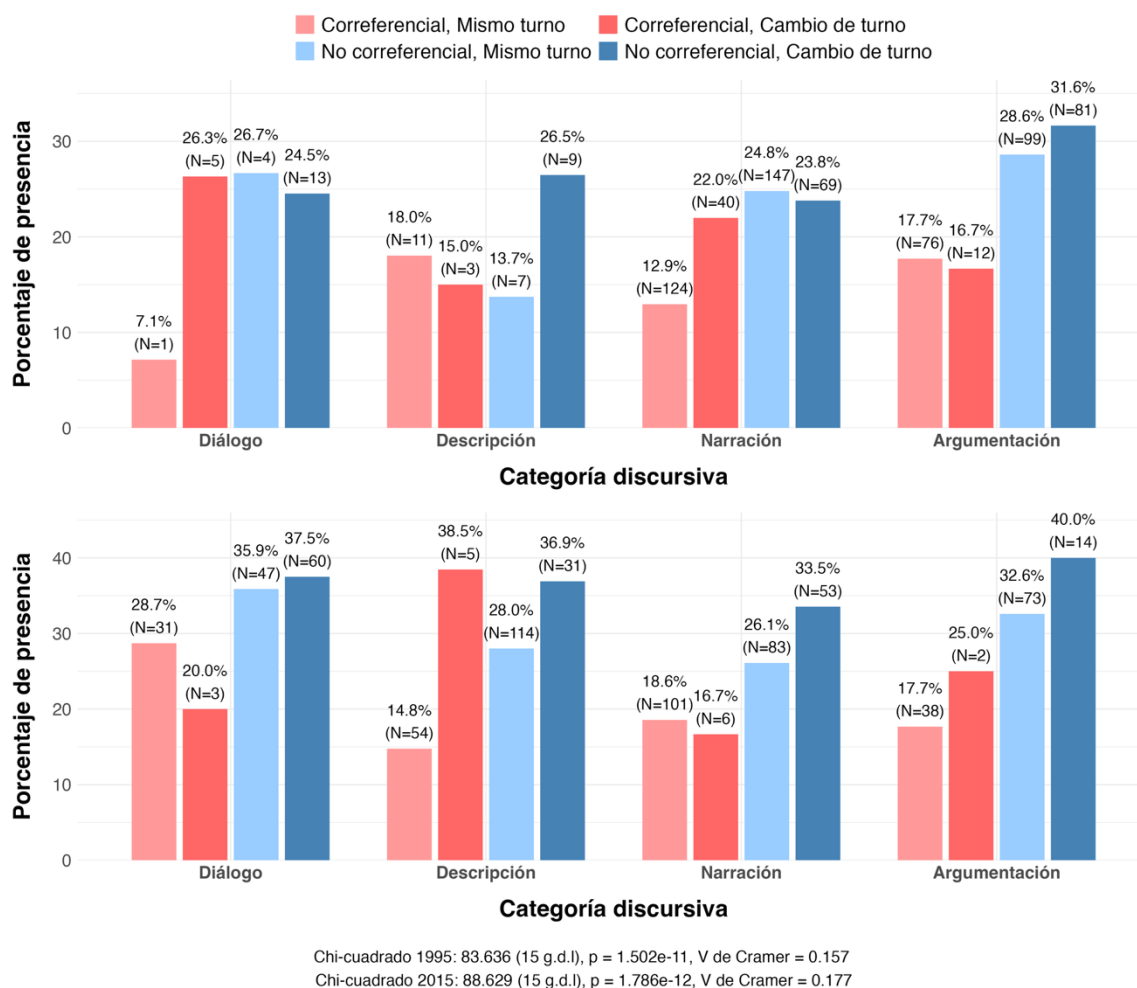


Gráfico 4. Presencia de sujetos pronominales según la correferencialidad, el turno de habla y el tipo de discurso

En cuanto a las secuencias descriptivas, en el corpus de 1995 los porcentajes de presencia son más bajos que el resto de las tipologías discursivas y, además, alcanza su mayor tasa de presencia en contextos no correferenciales y con cambio de turno. No obstante, estos resultados deben ser tomados con cautela debido al bajo número de casos ($N=9$ en el corpus de 1995 y $N=31$ en el corpus de 2015). En cambio, para el corpus de 2015 la interrupción de turno conversacional y el cambio de referente parecen favorecer la presencia del pronombre (37-38%).

Tipo de cláusula

Las cláusulas se clasifican según sean afirmativas, negativas o interrogativas con partícula o interrogativas sin partícula. La Tabla 12 recoge los casos detectados en ambos

corpus y, además, demuestra que esta variable no resulta ser significativa en ninguno de los dos cortes temporales analizados (valor de p 0.075 para el corpus de 1995 y valor de p 0.123 para el corpus de 2015).

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Afirmativa	2320	79.75	589	20.25	2909	85.74	1798	74.21	625	25.79	2423	85.86
Negativa	289	74.87	97	25.13	386	11.38	264	79.52	68	20.48	335	11.76
Interrogativa con partícula	23	82.14	5	17.86	28	0.83	16	66.67	8	33.33	24	0.85
Interrogativa sin partícula	60	85.71	10	14.29	70	2.06	30	69.77	13	30.23	43	1.52
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
Chi-cuadrado: 6.875506, g.d.l: 3, p-valor: 0.07597347. V de Cramer: 0.04501533							Chi-cuadrado: 5.764342, g.d.l: 3, p-valor: 0.1236557. V de Cramer: 0.04519562					

Tabla 12. Presencia/ausencia de sujetos pronominales según el tipo de cláusula

Debido al escaso número de oraciones interrogativas con y sin partícula, la Tabla 13 muestra los resultados de la reagrupación según los tres tipos de cláusula: afirmativa, negativa e interrogativa. Esta reordenación de los factores logra que la variable resulte significativa en el corpus de 1995 (valor de p : 0.034), aunque su fuerza de asociación entre las variables sigue siendo muy débil (V de Cramer: 0.044).

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Afirmativa	2320	79.75	589	20.25	2909	85.74	1798	74.21	626	25.84	2423	85.86
Negativa	289	74.87	97	25.13	386	11.38	264	79.52	68	20.48	335	11.76
Interrogativa	83	84.69	15	15.31	98	2.89	46	68.66	21	31.34	67	2.37
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 6.719878, g.d.l: 2, p-valor: 0.03473738. V de Cramer: 0.04450295						Chi-cuadrado: 5.733183, g.d.l: 2, p-valor: 0.0568925. V de Cramer: 0.04507331					

Tabla 13. Presencia/ausencia de sujetos pronominales según el tipo de cláusula

En términos de distribución de los casos, la gran mayoría de ejemplos estudiados corresponden a cláusulas afirmativas (85% del total del corpus) y, además, las tasas de explicitación pronominal obtenidas en los dos cortes temporales —a excepción de las

oraciones interrogativas en el corpus de 2015 que alcanzan un 30% de presencia, pero con una representatividad muy baja (2%)— son muy próximas entre las cláusulas afirmativas y negativas con solo una diferencia del 5%, cuya similitud ya había sido expuesta por Enríquez (1984).

Ambigüedad morfológica

La variable ambigüedad morfológica indica si la morfología flexiva del verbo resulta ambigua o no cuando la desinencia verbal es idéntica para dos personas del paradigma (Guerrero González & Manjón-Cabeza Cruz, 2022) tal y como sucede en formas del imperfecto del indicativo, en las formas condicionales y en los tiempos del subjuntivo. Con el propósito de establecer comparaciones con otras áreas de habla hispana analizadas previamente dentro del proyecto, se consideran formas no ambiguas aquellas cuya desinencia verbal determina de manera inequívoca la persona a la que hace referencia, y como formas ambiguas las que el verbo pueda tener una doble interpretación para la 1ª y 3ª persona del singular. En el ejemplo (25) la morfología verbal *hemos seguido pagando* señala a la 1ª persona del plural y en el ejemplo (26) para el verbo *me quedé trabajando* el sujeto sería 1ª persona del singular, no solo por la marca verbal, sino también por el pronombre átono que lo acompaña. En cambio, en el ejemplo (27) el sujeto de *podía repararlo* no resulta tan evidente, ya que podría ser tanto la 1ª como la 3ª persona del singular.

(25) *no ø hemos seguido pagando el piso porque yo me hice un seguro de vida. [MA_H22_014].*

(26) *sí, es que ø hice las prácticas y en ese mismo sitio ø me quedé trabajando. [MA_H11_022].*

(27) *si ø podía repararlo en la casa del cliente, lo ø reparaba allí. [MA_H31_050].*

La Tabla 14 muestra que la variable analizada resulta significativa en ambos corpus y, además, los dos cortes temporales coinciden tanto en la distribución de casos como en los porcentajes de presencia en formas ambiguas, ya que estas constituyen entre un 10-12% del total y obtienen un 33% de explicitación pronominal.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
No ambiguo	2412	81.10	562	18.90	2974	87.65	1907	75.61	615	24.39	2522	89.37
Ambiguo	280	66.83	139	33.17	419	12.35	201	67.00	99	33.00	300	10.63
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 44.80271, g.d.l: 1, p-valor: 2.179208e-11, Phi: 0.1149106						Chi-cuadrado: 10.0766, g.d.l: 1, p-valor: 0.001501645, Phi: 0.05975557					

Tabla 14. Presencia/ausencia de sujetos pronominales según la ambigüedad morfológica

Sin embargo, limitar el análisis exclusivamente a la ambigüedad derivada de la desinencia verbal sin considerar otros elementos contextuales que actúan como mecanismos desambiguadores, podría arrojar resultados que no reflejen de manera precisa la ambigüedad potencial o la función real que cumple el pronombre personal sujeto. Con el propósito de realizar un análisis más detallado de la ambigüedad morfológica, adoptamos el criterio propuesto por Silva-Corvalán (1982, p. 110), que subraya la necesidad de discernir entre los verbos que están precedidos de pronombres reflexivos clíticos y los que no lo están, dado que estos pronombres átonos resolverían la ambigüedad entre la 1ª y 3ª persona producida por la forma verbal (*yo me comía, él se comía*). Así, en el ejemplo (28) la forma verbal, que de otro modo sería idéntica para ambas personas en el imperfecto, se desambigua gracias al pronombre átono *se*, el cual indica que era *ella* la que se quedaba afónica.

(28) *con mucha frecuencia ø se quedaba afónica.*
[MA_M33_723].

Si aplicamos este criterio a nuestros datos, los ejemplos codificados como formas ambiguas —41 en el corpus de 1995 y 29 en el corpus de 2015— al estar acompañados de un pronombre átono, pasan a ser contabilizados como *no ambiguos*. Ahora bien, el estudio de Silva-Corvalán (1982) también considera esencial la distinción entre: a) verbo morfológicamente ambiguo, b) verbo morfológicamente no ambiguo (donde se han incorporado las formas ambiguas con pronombre átono), y c) verbo contextualmente no ambiguo. A la luz de las palabras de Silva-Corvalán (1982, p. 108), una forma verbal, a pesar de ser morfológicamente ambigua, se considera contextualmente no ambigua cuando el sujeto posee un rasgo [+humano] y la única opción interpretativa es la 1ª persona del singular, es decir, el hablante hace referencia a sí mismo. En estos casos, se entiende que

el discurso proporciona información suficiente para calificar a estas cláusulas como no ambiguas. A continuación, se muestran los ejemplos utilizados en el trabajo de Silva-Corvalán para la ilustrar el concepto de *no ambiguo contextualmente*:

- (29) *Así es de que yo hacia el plato y había un negrito que lo cortaba. Ese redondeaba. Y entonces me lo ø mandaba a mí y yo le ponía un anillo, que lleva arriba para, pa' librarlo del calor era el anillo ese.*
- (30) *Entonces, a mí me daban la formula y yo, yo todos los, las batía, luego ahí los ø mandaba a los tanques, ø ponía los tanques, los ø llenaba, la ø tapaba y luego lo, lo ø comenzaba ya pa' la hora de irme los ø comenzaba.*

En los primeros pasos del análisis, ya excluimos los casos en los que el referente presenta un rasgo diferente al [+humano]. Por lo tanto, siguiendo la clasificación propuesta por Silva-Corvalán, y adaptándola a la codificación de este trabajo, se seleccionan los casos con una forma verbal ambigua que son correferenciales con el sujeto previo y se encuentran en la oración inmediatamente anterior. En esta ocasión, en lugar de limitar el análisis a los casos de 1ª persona, optamos por incluir también los de 3ª persona.

La reordenación de los casos de ambigüedad morfológica produce cambios en la significatividad de las variables, ya que el valor de p es más bajo y la fuerza de asociación entre las variables aumenta en ambos corpus (V de Cramer: 0.141 en 1995 y V de Cramer: 0.113 en 2015). Además, la Tabla 15 muestra como los ejemplos contextualmente no ambiguos exhiben unas tasas de ausencia y presencia del pronombre personal similares a los no ambiguos. En cambio, al excluir las ocurrencias desambiguadas por el contexto, las formas morfológicamente ambiguas experimentan una gran variación en sus porcentajes, y alcanzan hasta un 47% de presencia del pronombre personal en el corpus de 1995 y un 50% en el corpus de 2015.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
No ambiguo	2443	81.03	572	18.97	3015	88.86	1926	75.50	625	24.50	2551	90.40
No ambiguo contextual	186	72.37	71	27.63	257	7.57	129	78.18	36	21.82	165	5.85
Ambiguo	63	52.07	58	47.93	121	3.57	53	50	53	50	106	3.76
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 67.76151, g.d.l: 2, p-valor: 1.930971e-15, V de Cramer: 0.1413187						Chi-cuadrado: 36.13904, g.d.l: 2, p-valor: 1.420717e-08, V de Cramer: 0.1131644					

Tabla 15. Presencia/ausencia de sujetos pronominales según la ambigüedad morfológica recodificada

Los resultados anteriores permiten afirmar que la ambigüedad morfológica no depende exclusivamente de la forma verbal que hace confluir a la 1ª y 3ª persona del singular en algunos tiempos verbales —como el imperfecto—, sino que esta puede estar condicionada por otros elementos presentes en el discurso, como los pronombres átonos o la ambigüedad contextual. Si bien la codificación original (formas morfológicamente ambiguas o no ambiguas) resulta de gran utilidad para describir la naturaleza del verbo que se está analizando, se requiere una observación más minuciosa y contextualizada si se pretende observar cómo esta la ambigüedad morfológica puede condicionar la presencia del pronombre. Estudios más recientes en el ámbito de PRESEEA también exploran la ambigüedad morfológica, aunque no incorporan en sus análisis la consideración de la desambiguación contextual y, por lo tanto, los resultados deberían ser tomados con cautela.

En algunos trabajos, como el de Manjón-Cabeza Cruz (2022, p. 7), se cuestiona si la mayor tasa pronominal viene determinada por la ambigüedad morfológica o por el tiempo verbal empleado y, de esta forma, si la ambigüedad es la que propicia una mayor explicitación del pronombre, debería observarse un aumento de la presencia del sujeto pronominal en todos los tiempos verbales que pudieran ser ambiguos morfológicamente. Además, Manjón-Cabeza Cruz, (2022) se cuestiona si la ambigüedad morfológica prevaleciera sobre el tiempo verbal, también en las formas del presente del subjuntivo debería aumentar de la tasa pronominal.

Para comprobar esta afirmación, presentamos en la Tabla 16 los resultados en función del modo y las formas verbales ambiguas, donde se han excluido los casos del modo imperativo por no presentar ambigüedad. Cabe destacar que, a pesar de las bajas

ocurrencias de los verbos en subjuntivo que provocan que los resultados deban ser interpretados con cautela (solo suman 110 ejemplos en el corpus de 1995 y 107 en el corpus de 2015). En la ciudad de Málaga se observa un patrón similar en el modo indicativo y subjuntivo en los dos cortes temporales estudiados. Las formas morfológicamente ambiguas son las que alcanzan mayores porcentajes de expresión pronominal tanto en indicativo (50% en el corpus de 1995 y 49% en el corpus de 2015) como en subjuntivo (33% en el corpus de 1995 y 37% en el corpus de 2015).

	Corpus 1995					Corpus 2015				
	Ø		Pro		Total	Ø		Pro		Total
	N	%	N	%	N	N	%	N	%	N
Indicativo ambiguo	55	49.54	54	50.45	109	50	51.03	48	48.97	98
Indicativo no ambiguo contextualmente	167	71.36	67	28.63	234	114	77.03	34	22.97	148
Indicativo no ambiguo	2339	80.76	557	19.23	2896	1815	74.94	607	25.06	2422
Subjuntivo ambiguo	8	66.66	4	33.33	12	3	62.5	5	37.5	8
Subjuntivo no ambiguo contextualmente	19	82.60	4	17.39	23	15	88.23	2	11.76	17
Subjuntivo no ambiguo	65	86.66	10	13.33	75	69	84.14	13	15.85	82
Total	2653	79.21	696	20.78	3349	2066	74.45	709	25.54	2775

Tabla 16. Presencia/ausencia de sujetos pronominales según el modo y la ambigüedad contextualizada

Uno de los argumentos² manejados en el trabajo de Manjón-Cabeza Cruz (2022, p. 7) es que el aumento de la presencia del pronombre se debe a la ambigüedad morfológica, tendría que esperarse una mayor explicitación en todos los tiempos verbales ambiguos y el pronombre sujeto debería aparecer por igual en la dos personas verbales donde se produce el sincretismo. Con el objetivo de corroborar si esto se cumple en la ciudad de Málaga, la

² Con el propósito de comprobar qué sucede en cada tiempo verbal, observamos solo las tasas de presencia del sujeto pronominal correspondientes a los tiempos verbales del modo indicativo, ya que, al desglosar por tiempos verbales y modo, los casos quedan muy repartidos y resultan insignificantes. Excluidos los ejemplos de presente y futuro de indicativo por no presentar formas morfológicamente ambiguas, observamos que los tiempos del imperfecto, pretérito y condicional, con mayor o menor número de ocurrencias, presentan la misma tendencia que, además, se repite de manera casi idéntica en las dos sincronías analizadas: las formas no ambiguas son las que menor tasa pronominal requieren y, al ascender el grado de ambigüedad el pronombre se hace más necesario en los tres tiempos verbales. Estos resultados no llegan a ser representativos debido a la baja ocurrencia del número de casos en nuestros corpus.

Tabla 17 recoge los datos de la interacción de la 1ª y 3ª persona del singular con la ambigüedad morfológica contextualizada.

	Corpus 1995					Corpus 2015				
	Ø		Pro		Total	Ø		Pro		Total
	N	%	N	%	N	N	%	N	%	N
1ª No ambiguo	760	69.98	326	30.18	1086	603	59.29	414	40.15	1017
1ª No ambiguo contextual	94	60.25	62	39.74	156	67	68.69	31	31.31	99
1ª Ambiguo	40	43.01	53	56.99	93	32	41.56	45	58.44	77
3ª No ambiguo	315	86.30	50	13.70	365	188	86.24	30	13.76	218
3ª No ambiguo contextual	92	91.08	9	8.91	101	61	92.54	5	7.46	67
3ª Ambiguo	23	82.14	5	17.85	28	18	69.23	8	30.77	26
Total	1324	72.29	505	27.71	1829	976	64.69	533	35.51	1509

Tabla 17. Presencia/ausencia de sujetos pronominales según la persona del sujeto y la ambigüedad contextualizada

En cuanto a la 1ª persona del singular, identificamos un aumento progresivo de la tasa de presencia del sujeto pronominal en los corpus de 1995 y 2015: desde un 30-40% respectivamente, acompañado de formas verbales no ambiguas, a un 39-31% con formas contextualmente no ambiguas, hasta alcanzar, finalmente, un 57-58% con formas morfológicamente ambiguas. A diferencia de los resultados obtenidos en la ciudad de Granada, los casos de 3ª persona del singular sí experimentan un aumento de la presencia del sujeto pronominal. En ambos corpus las formas no ambiguas obtienen un 13% de explicitación pronominal y esta asciende con formas morfológicamente ambiguas hasta el 17% en corpus de 1995 y alcanza el 30% de presencia en el corpus de 2015.

Las diferencias en el uso de la 1ª y 3ª persona podrían atribuirse a la opción de utilizar sintagmas nominales o nombres propios para referirse a la 3ª persona, pero siguen evidenciándose diferencias entre las formas verbales ambiguas y las no ambiguas. En cualquier caso, resulta indiscutible que el contexto de la ambigüedad parece ser clave para tratar adecuadamente los datos en la situación comunicativa analizada, ya que limitar el análisis a la forma verbal puede conducir a resultados poco esclarecedores.

Ambigüedad fonética

La elisión de la coda silábica de la morfología verbal puede ocasionar ambigüedad fonética, de manera similar a la morfológica, ya que varias personas gramaticales (2ª y 3ª persona del singular o 1ª, 2ª y 3ª persona del singular) pueden presentar una forma verbal idéntica. En el ejemplo (31), se manifiesta una elisión fonética de /-s/ que produce ambigüedad, ya que el sujeto de esa oración podría ser tanto una 2ª persona del singular como una 3ª persona del singular. Lo mismo sucedería en el ejemplo (32), donde la no aparición de la /-n/ en el verbo podría hacer confluir una 3ª persona del plural con una 2ª o 3ª persona del singular. Por el contrario, en el ejemplo (33), aunque el hablante no pronuncie la /-s/ de la coda silábica del verbo, podría seguir identificándose el sujeto sin problema alguno, ya que esta elisión no sería un rasgo distintivo de persona.

(31) *pero una rosa la otra celeste o rojo como ø quisieras que fuera, pero era un cuadro [MA_M21_105].*

(32) *ø dicen que en su casa ø están usando la variedad estándar// [MA_H13_028].*

(33) *en general, en general sí ø somos trabajadores, aunque nos tengan por, por raja[d]os [MA_H21_005].*

La Tabla 18 muestra los datos correspondientes a los dos cortes temporales en función de la variable ambigüedad fonética, donde los casos de 3ª persona del singular se consideran ambiguos por la posible elisión del segmento /-n/ en la coda silábica del verbo.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
No ambiguo	1559	79.62	399	20.38	1958	57.71	1160	70.43	487	29.57	1647	58.36
Ambiguo	1133	78.95	302	21.05	1435	42.29	948	80.68	227	19.32	1175	41.64
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 0.1861248, g.d.l: 1, p-valor: 0.6661617, Phi: 0.007406451						Chi-cuadrado: 37.57912, g.d.l: 1, p-valor: 8.777896e-10, Phi: 0.115397					

Tabla 18. Presencia/ausencia de sujetos pronominales según la ambigüedad fonética

Con el objetivo de identificar los factores que puedan estar afectando a los resultados obtenidos, desglosamos la ambigüedad fonética potencial según la persona del

sujeto. El análisis de la 3ª persona del plural en relación con la ambigüedad fonética puede que esté condicionado por el factor de la especificidad, ya que la 3ª persona del plural promueve un alto porcentaje de ausencia (100% en contextos inespecíficos).

Otro argumento que deber ser tenido en cuenta es que el estudio previo realizado en el mismo corpus de PRESEEA-Málaga por Martín Morales (2018, p. 244) ya indica que la tasa de elisión de la /-n/ en coda silábica es tan solo del 3.4%, ya que la realización velarizada se produce en el 95.5% de los casos. Por lo tanto, optamos por considerar los ejemplos de 3ª del plural como no ambiguos y, de esta forma, la variable queda reducida al sincretismo producido por la posible elisión de la /-s/ en la coda silábica.

Silva-Corvalán (2001, p. 159) recomienda que no solo se estudien los diferentes grados de elisión de /-s/ y /-n/, sino que también se observe la interacción de estos rasgos fonéticos con el nivel morfológico y sintáctico. De esta forma, seguimos el mismo criterio adoptado en la exploración de la ambigüedad morfológica: todas las unidades compuestas por un verbo pronominal (o dativo ético) no conforman el paradigma de la ambigüedad fonética porque, aunque la forma verbal pueda resultar ambigua fonéticamente, el pronombre átono actúa como desambiguador.

Para considerar en el análisis los elementos presentes en el contexto discursivo, recodificamos la variable de la ambigüedad fonética contextualizada, de modo que distinguimos entre formas no ambiguas fonéticamente, no ambiguas contextualmente y ambiguas fonéticamente. Al igual que hemos considerado relevante contextualizar la cláusula en el análisis de la ambigüedad morfológica, aplicamos en esta variable el mismo criterio propuesto por Silva-Corvalán (1982, p. 108): una unidad correferencial con la anterior deja de ser contextualmente ambigua, aunque la forma verbal pueda presentar ambigüedad morfológica o, en este caso, fonética. A continuación, la Tabla 19 presenta los resultados tras el reajuste de la variable:

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
No ambiguo	2026	82.06	443	17.94	2469	72.77	1599	74.75	540	25.25	2139	75.80
No ambiguo contextual	417	80.04	104	19.96	521	15.36	303	83.24	61	16.76	364	12.90
Ambiguo	249	61.79	154	38.21	403	11.88	206	64.58	113	35.42	319	11.30
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 87.03187, g.d.l: 2, p-valor: 1.26261e-19, V de Cramer: 0.1601575						Chi-cuadrado: 31.35248, g.d.l: 2, p-valor: 1.555588e-07, V de Cramer: 0.1054041					

Tabla 19. Presencia/ausencia de sujetos pronominales según la ambigüedad fonética contextualizada

Las formas fonéticamente no ambiguas obtienen una presencia del pronombre del 18% en el corpus de 1995 y un 25% en el corpus de 2015, lo que contrasta notablemente si lo comparamos con las formas contextualmente no ambiguas. En estas últimas, la tasa de explicitación (20%) en el corpus de 1995 muy similar a las formas no ambiguas (18%), pero en el corpus de 2015 la aparición del sujeto desciende hasta el 17%, siendo inferior que las formas no ambiguas (25%). La diferencia más notable reside en el aumento considerable de las formas potencialmente ambiguas que no pueden ser desambiguadas por contexto, ya que alcanzan entre el 36-38% de aparición del pronombre personal. Aunque estos resultados reflejan una gran necesidad de explicitación del pronombre ante la ambigüedad fonética, no logran alcanzar el 49% de presencia de las formas morfológicamente ambiguas.

Otro aspecto que parece interactuar con la ambigüedad fonética es la especificidad del referente, dado que esta última puede restringir la expresión del sujeto pronominal, especialmente cuando el referente es inespecífico. Trabajos previos, como el de Cameron (1993), exploran la relación entre las altas tasas pronominales obtenidas en sujetos específicos y no específicos con formas verbales ambiguas. El Gráfico 5 ilustra la distribución de los casos de 2ª persona del singular documentados en la ciudad de Málaga y, además, esta combinación de variables resulta estadísticamente significativa en los dos temporales examinados (valor de p : <0.001) con una fuerza de asociación más elevada en el corpus de 2015 (V de Cramer: 0.245) que el de 1995 (V de Cramer: 0.201).

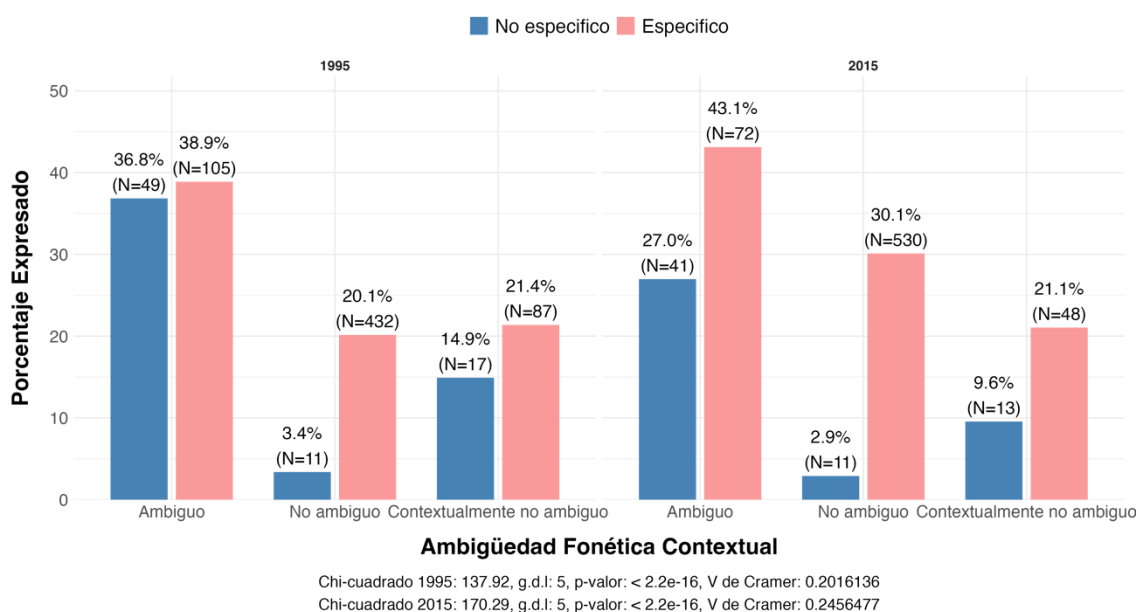


Gráfico 5. Presencia/ausencia de sujetos pronominales según la ambigüedad fonética contextualizada y la especificidad

En términos generales, observamos que la frecuencia de la explicitación del sujeto es considerablemente inferior cuando el referente es inespecífico en comparación con los referentes específicos, lo que se asemeja a los resultados obtenidos previamente. En cuanto a la ambigüedad, los casos se distribuyen según lo esperado: la presencia es más baja con formas verbales no ambiguas —para referentes inespecíficos 3% y para sujetos específicos entre un 20-30%—. Esta aparición pronominal se incrementa ligeramente con verbos contextualmente no ambiguos —entre un 10 y 15% en referentes no específicos y un 21% en referentes específicos— y las mayores tasas pronominales se alcanzan en contextos de ambigüedad fonética potencial.

Sin embargo, el comportamiento de los datos varía notablemente en cuanto a las formas ambiguas. En lo que respecta al corpus de 1995, las diferencias según el grado de especificidad del referente son poco evidentes (valor de p : 0.772), ya que los referentes específicos usan el pronombre expreso en un 39% de los casos y los no específicos en un 37%. Esto parece indicar que la necesidad de explicitación no dependía únicamente de la especificidad del sujeto, sino que la ambigüedad fonética que pudiera ocasionarse en el contexto comunicativo demanda una mayor explicitación.

No obstante, el corpus de 2015 revela un mayor contraste entre las tasas pronominales que se acentúa significativamente (valor de p : 0.003), de modo que la

presencia del pronombre con referentes específicos alcanza un 43% y con no específicos se reduce a un 27%. Los resultados de este corte temporal sugieren un cambio en el uso del pronombre personal con respecto al periodo anterior donde, además de confirmar que la ambigüedad fonética potencial influye en la aparición del sujeto, identificamos que el uso de referentes específicos tiende a adquirir una función más deíctica dentro del discurso. En cambio, los referentes no específicos se orientan hacia la omisión pronominal, pese a las posibles circunstancias de ambigüedad fonética.

2.3.2.3. Variables referentes al verbo de la oración

En cuanto a las variables relacionadas con el verbo de la oración, abordamos la clase semántica del verbo, y también otras que codifican características del verbo como tiempo verbal, modo, aspecto perfectivo, progredisidad, pronominalidad y si se trata o no de perífrasis verbal.

Clase semántica según la tipología ADESSE

La clasificación ADESSE (Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español) organiza los verbos en seis macrocategorías de procesos verbales:

- Verbos de procesos mentales.

(34) *yo creo que tampoco era solamente por, por el amor y por las cosas [MA_M31_025].*

- Verbos de procesos relacionales.

(35) *sólo ø tengo relación con/ con mi vecina [MA_M23_019].*

- Verbos de procesos materiales.

(36) *así que ya monté yo el taller de radio [MA_H31_050].*

- Verbos de procesos verbales.

(37) *ø no paraban de cuchichear, pum pum y hablar y chillando [MA_M11_057].*

- Verbos de procesos existenciales.

(38) *tú hoy en día vives en Benalmádena [MA_H33_716].*

- Verbos de modulación.

(39) *y yo también podría haber opta[d]o por/ estar de prácticas en algún departamento [MA_M13_036].*

La Tabla 20 evidencia que los verbos más empleados en ambos corpus son los verbos de procesos materiales (37% en el corpus de 1995 y 36% en el corpus de 2015). No obstante, los verbos que consiguen una mayor tasa de explicitación pronominal son los de procesos mentales con un 28% de presencia en el corpus de 1995 y un 37% en el corpus de 2015, donde se advierte que los verbos de procesos existenciales logran un 35% de aparición del sujeto.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Mental	517	71.41	207	28.59	724	21.34	415	62.78	246	37.22	661	23.42
Relacional	645	78.09	181	21.91	826	24.34	498	76.73	151	23.27	649	23.00
Material	1042	82.50	221	17.50	1263	37.22	797	78.99	212	21.01	1009	35.75
Verbal	230	82.73	48	17.27	278	8.19	226	84.01	43	15.99	269	9.53
Existencial	159	85.03	28	14.97	187	5.51	79	64.23	44	35.77	123	4.36
Modulación	99	86.09	16	13.91	115	3.39	93	83.78	18	16.22	111	3.93
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 45.11442, g.d.l: 5, p-valor: 1.375203e-08, V de Cramer: 0.1153097						Chi-cuadrado: 85.23723, g.d.l: 5, p-valor: 6.713157e-17, V de Cramer: 0.1737946					

Tabla 20. Presencia/ausencia de sujetos pronominales según la clasificación ADESSE

Al desglosar la clasificación anterior según las subcategorías verbales proporcionados por ADESSE (detallados en el apartado de la metodología), obtenemos los datos presentados en la Tabla 21 que nos permiten examinar con mayor detalle cómo se distribuye la presencia del pronombre.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Sensación	94	80.34	23	19.66	117	3.45	80	80	20	20	100	3.54
Percepción	132	75	44	25.00	176	5.19	119	68.39	55	31.61	174	6.17
Cognición	285	67.86	135	32.14	420	12.38	213	56.05	167	43.95	380	13.47
Elección	7	58.33	5	41.67	12	0.35	3	42.86	4	57.14	7	0.25
Existencia							7	63.64	4	36.36	11	0.39
Fase-Tiempo							57	73.08	21	26.92	78	2.76
Vida							15	44.12	19	55.88	34	1.20

Tabla 21. Presencia/ausencia de sujetos pronominales según la clasificación de subcategorías ADESSE

En cuanto a los verbos incluidos dentro de la macrocategoría verbos de procesos metales, los que logran una elevada presencia del pronombre, y a su vez cuentan con una representación considerable dentro del total del corpus, son los verbos de cognición con una tasa pronominal del 32% en el corpus de 1995 y del 44% en el corpus de 2015, lo que suponen aproximadamente el 13% de los casos totales del corpus. Al igual que los verbos de percepción, obtienen entre un 25-31% de presencia y representan el 5-6% del total.

Sin embargo, otros subtipos como elección alcanzan tasas pronominales elevadas, que rondan entre el 41% en el corpus de 1995 y el 57% en el de 2015, en el que su aparición resulta insignificante, ya que es inferior al 1%. Lo mismo sucede en el corpus de 2015 con los verbos de existencia y vida que llegan a obtener una presencia pronominal del 36% y el 55%, respectivamente, pero tiene una baja representatividad. A pesar de que todas las investigaciones previas no siguen los mismos criterios para clasificar semánticamente los verbos, a grandes rasgos los resultados parecen coincidir en que los verbos de percepción y los cognitivos son los que promueven una mayor la realización explícita del sujeto (Silva-Corvalán & Enrique Arias, 2017, p. 175).

Con el objetivo de estudiar con detalle cómo la interacción entre verbo empleado y la 1ª persona del singular influyen en la explicitación del sujeto pronominal, el Gráfico 6 muestra las frecuencias de uso y el porcentaje de presencia en los dos cortes temporales estudiados.

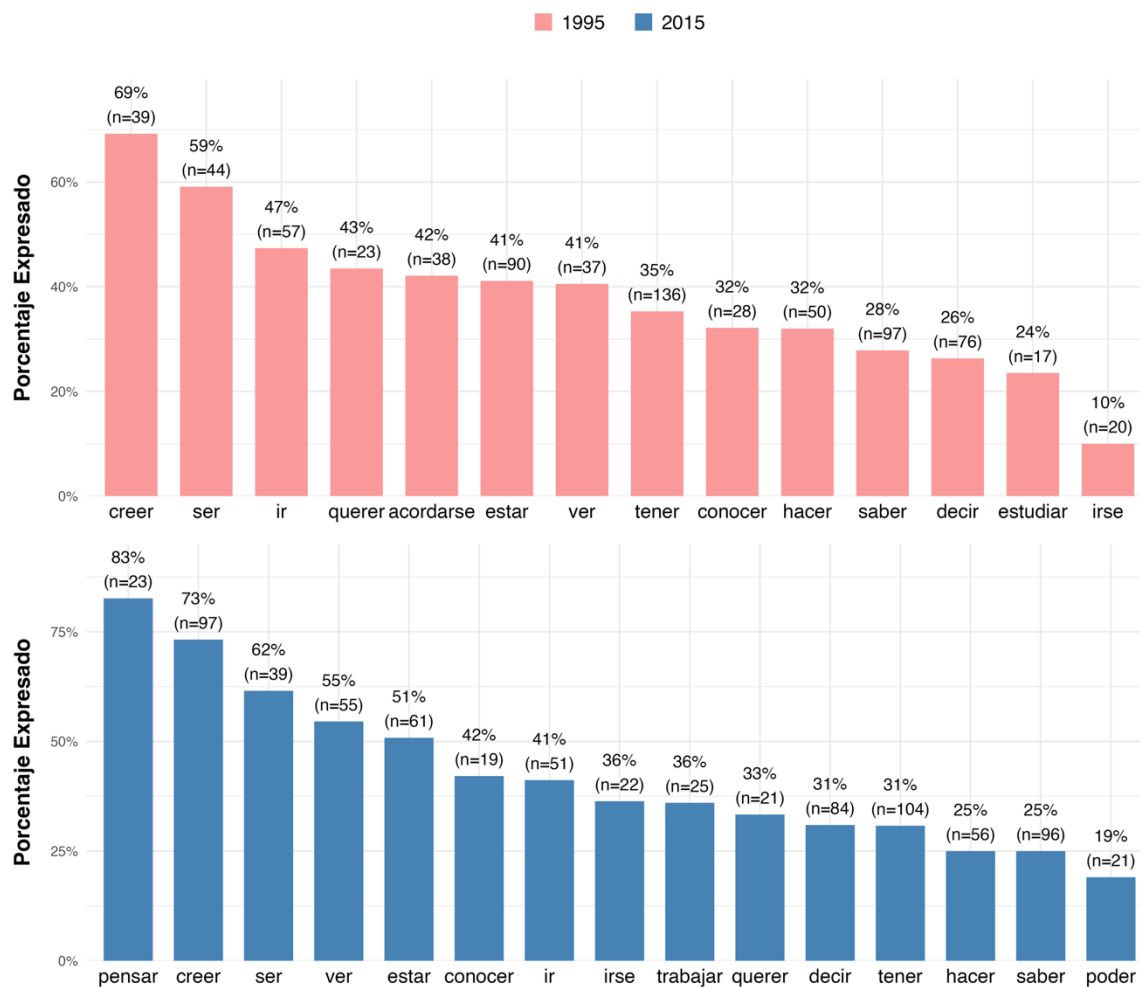


Gráfico 6. Presencia/ausencia de sujetos pronominales según el verbo y la persona del sujeto (porcentaje de presencia y frecuencia)

Cabe destacar una tendencia generalizada donde la presencia del sujeto con la 1ª persona del singular acompañada de ciertos verbos experimenta unas tasas de presencia considerablemente elevadas. En la parte izquierda del Gráfico 6, se presentan los verbos con mayor aparición del sujeto en el corpus de 1995, donde se sitúan los verbos relacionados con procesos mentales como *creer* (69%), *querer* (43%), *ver* (41%), *acordarse* (42%), *conocer* (32%) o *saber* (28%), y otros como *ser* (59%) o *estar* (41%). En cuanto al segundo corte temporal, los verbos que propician una mayor aparición del sujeto también están dentro de la macrocategoría de verbos de procesos mentales, pero especialmente los verbos *pensar* (83%) y *creer* (73%) son los que alcanzan unos porcentajes de presencia considerablemente elevados.

Además, se advierten cambios en su frecuencia de uso en cuanto al verbo *creer*: mientras que en el corpus de 1995 se registran 39 ejemplos, el número de casos experimenta un aumento considerable hasta 97 ejemplos en el corpus de 2015. Algunos trabajos previos ya manifiestan resultados similares a los aquí expuestos. Así Orozco y Hurtado (2020, p. 721) consideran que este patrón podría indicar un proceso de gramaticalización en el que ciertas colocaciones de sujeto y verbo tienden a constituirse como paradigmas prefabricados en español que se gramaticalizan en fórmulas discursivas.

Tiempos verbales

La última guía de codificación de PRESEEA (Guerrero González & Manjón-Cabeza Cruz, 2022) considera los tiempos verbales canónicos desglosados tanto en formas simples como compuestas.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Presente	1217	77.32	357	22.68	1574	46.39	1188	74.06	416	25.94	1604	56.84
Pretérito	491	84.08	93	15.92	584	17.21	313	77.67	90	22.33	403	14.28
Imperfecto	669	79.93	168	20.07	837	24.67	413	76.91	124	23.09	537	19.03
Futuro morfológico	26	96.30	1	3.70	27	0.80	9	100	0	0	9	0.32
Futuro perifrástico	47	90.38	5	9.62	52	1.53	48	80.00	12	20.00	60	2.13
Condicional	19	76.00	6	24.00	25	0.74	17	77.27	5	22.73	22	0.78
Pretérito perfecto compuesto	210	77.21	62	22.79	272	8.02	99	62.26	60	37.74	159	5.63
Pretérito Pluscuamperfecto	9	56.25	7	43.75	16	0.47	14	66.67	7	33.33	21	0.74
Futuro compuesto	3	60.00	2	40.00	5	0.15	4	100	0	0	4	0.14
Condicional compuesto	1	100	0	0	1	0.03	3	100	0	0	3	0.11
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
Chi-cuadrado: 28.22487, g.d.l: 9, p-valor: 0.0008747297, V de Cramer: 0.09120613							Chi-cuadrado: 23.72145, g.d.l: 9, p-valor: 0.004764117, V de Cramer: 0.09168369					

Tabla 22. Presencia/ausencia de sujetos pronominales según el tiempo verbal desglosado

En una primera observación de los datos, resulta llamativo que casi la totalidad de casos en ambas sincronías se encuentra distribuida entre los tiempos de presente (un 46%

en el corpus de 1995 y un 57% en el de 2015), imperfecto (17% en el corpus de 1995 y 14% en el de 2015) y pretérito (25% en el corpus de 1995 y un 19% en el de 2015), lo que reduce la presencia del resto de tiempos verbales a tasas que escasamente consiguen alcanzar el 4%. En cuanto a la explicitación del sujeto pronominal, los tiempos verbales que obtienen unas mayores tasas pronominales son: el pretérito perfecto logra un 22% en el corpus de 1995 y 37% en el corpus de 2015; el pretérito pluscuamperfecto alcanza un 43% en el corpus de 1995 y un 33% en el corpus de 2015; y, por último, el futuro compuesto solo obtiene un porcentaje elevado de presencia (40%) en el corpus de 1995.

Sin embargo, los tiempos compuestos solo suponen una proporción escasa del total del corpus y, como sugieren las indicaciones de la guía de codificación (Guerrero González & Manjón-Cabeza Cruz, 2022), los casos que estén infrarrepresentados pueden ser obviados o reagrupados. Por ello, optamos por recodificar esta variable y, de este modo, la Tabla 23 presenta los resultados obtenidos para los verbos de manera conjunta sin distinguir entre tiempos simples o compuestos.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Presente	1217	77.32	357	22.68	1574	46.39	1188	74.06	416	25.94	1604	56.84
Pretérito	710	81.42	162	18.58	872	25.70	426	73.07	157	26.93	583	20.66
Imperfecto	669	79.92	168	20.07	837	24.67	413	76.91	124	23.09	537	19.03
Futuro	76	90.48	8	9.52	84	2.48	61	83.56	12	16.44	73	2.59
Condicional	20	76.92	6	23.08	26	0.77	20	80.00	5	20.00	25	0.88
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 12.85287, g.d.l: 4, p-valor: 0.01201754, V de Cramer: 0.06154717						Chi-cuadrado: 5.95253, g.d.l: 4, p-valor: 0.2027215, V de Cramer: 0.04592745					

Tabla 23. Presencia/ausencia de sujetos pronominales según el tiempo verbal reagrupados

Con el fin de analizar los verbos más allá de su valor temporal, reclasificamos los verbos según la propuesta de Silva-Corvalán y Enrique Arias (2017) quien sostiene que, más que la ambigüedad morfológica, es el tiempo verbal, y en particular la función que este cumple en el contexto discursivo, lo que se correlaciona con la expresión del sujeto. Esto es conocido como la *hipótesis semántico-pragmática*. Para corroborar esta hipótesis,

empleamos la clasificación previa propuesta por Hochberg (1987), que vincula de la siguiente manera la relación semántica y pragmática de los tiempos verbales:

- Tipo A: Pretérito: factual, asertivo, dinámico, foco en el evento, *foregrounded*.
- Tipo B: Presente: factual y asertivo, pero no siempre dinámico y focal como en el pretérito.
- Tipo C: Imperfecto, condicional y tiempos del subjuntivo: *backgrounded*, *irrealis*, no-dinámico, no-asertivo, cortés.

Si aplicamos esta clasificación a los datos de nuestro trabajo, los resultados no coinciden con la tendencia esperada. En primer lugar, la variable no es significativa en ninguno de los dos corpus (valor de p: 0.128 para el corpus de 1995 y valor de p: 0.077 para el corpus de 2015). Aunque se espera que el Tipo A obtenga la menor presencia del pronombre personal sujeto (26.98% y 18.58%), y que dicha presencia aumente progresivamente en el Tipo B (25.75% y 22.31%) hasta alcanzar su punto máximo en el Tipo C (22.31% y 20.02%), los resultados obtenidos en nuestra investigación revelan lo tendencia contraria: el Tipo C es el que presenta la menor presencia del pronombre, con un 22.31% en el corpus de 1995 y un 20.02% en el corpus de 2015.

Modo

El modo verbal es otro de los factores que pueden condicionar la presencia del sujeto pronominal, ya que este se relaciona con los aspectos semántico-pragmáticos que reflejan la postura que adopta el hablante frente a la situación comunicativa (Hurtado, 2005).

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Indicativo	2561	79.07	678	20.93	3239	95.46	1980	74.21	688	25.79	2668	95.54
Subjuntivo	92	83.64	18	16.36	110	3.24	87	81.31	20	18.69	107	3.79
Imperativo	39	88.64	5	11.36	44	1.30	41	87.23	6	12.77	47	1.67
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 3.705133, g.d.l: 2, p-valor: 0.1568341, V de Cramer: 0.03304532						Chi-cuadrado: 6.588104, g.d.l: 2, p-valor: 0.0371032, V de Cramer: 0.0482233					

Tabla 24. Presencia/ausencia de sujetos pronominales según el modo

A pesar de que esta variable, como se muestra en la Tabla 24, solo sea significativa en el corpus de 2015, es posible que, al basar nuestro trabajo en corpus formados por entrevistas semidirigidas, la naturaleza de estas condicione una mayor frecuencia de uso del indicativo (95%) y en la tasa de presencia del pronombre personal (21% en el corpus de 1995 y 26% en corpus de 2015), en comparación con el resto de los modos.

Perfectividad

La variable de la perfectividad indica si el verbo en la cláusula analizada se presenta en un tiempo simple o compuesto. La Tabla 25 recoge los datos obtenidos en los dos corpus estudiados.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Simple	2469	79.67	630	20.33	3099	91.34	1988	75.45	647	24.55	2635	93.37
Compuesto	223	75.85	71	24.15	294	8.66	120	64.17	67	35.83	187	6.63
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
Chi-cuadrado: 2.163744, g.d.l: 1, p-valor: 0.1413001, Phi: 0.02525288							Chi-cuadrado: 11.15534, g.d.l: 1, p-valor: 0.0008379028, Phi: 0.0628728					

Tabla 25. Presencia/ausencia de sujetos pronominales según la perfectividad

La variable solo resulta ser significativa en el corpus de 2015, pero en ambos corpus los casos experimentan una distribución similar, donde las formas simples rondan entre el 91-93% y la presencia del pronombre personal sujeto es mayor en las formas compuestas (24% en el corpus de 1995 y 35% en el corpus de 2015).

Perífrasis verbal

La variable perífrasis verbal distingue entre tres tipos de verbos, ejemplificados a continuación: en el ejemplo (40), la forma verbal he jugado constituye una forma no perifrástica. Los siguientes ejemplos son considerados casos perifrásticos, donde diferenciamos dos tipos: formas perifrásticas modales como en el ejemplo (41) con la construcción no pude hacer y formas perifrásticas aspectuales como presenta el ejemplo (42) en la frase empecé a preparar.

(40) *es sí, ø he juga[d]o hasta en juveniles, juveniles es hasta 18*
[MA_M11_057].

(41) *no ø tenía la edad/ ø no pude hacer ni el examen de acceso*
[MA_M12_032]].

(42) *eh/ yo empecé a preparar oposiciones/ entonces eran estas*
oposiciones [MA_H33_040].

En la Tabla 26, podemos observar que las dos formas perifrásticas alcanzan tasas pronominales considerablemente bajas, 17,63% y 16,06% respectivamente.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
No perífrasis	2374	78.53	649	21.47	3023	89.10	1779	73.63	637	26.37	2416	85.61
Modal	122	84.14	23	15.86	145	4.27	107	80.45	26	19.55	133	4.71
Aspectual	196	87.11	29	12.89	225	6.63	222	81.32	51	18.68	273	9.67
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 11.53198, g.d.l: 2, p-valor: 0.003132296, V de Cramer: 0.05829885						Chi-cuadrado: 10.10769, g.d.l: 2, p-valor: 0.006384748, V de Cramer: 0.05984769					

Tabla 26. Presencia/ausencia de sujetos pronominales según la perífrasis verbal

Esta baja presencia del pronombre personal sujeto acompañado de formas perifrásticas puede estar relacionada con diversos factores, tal y como adelantaba Manjón-Cabeza Cruz et al. (2016). Una de las causas puede ser la progresividad, que no resulta ser significativa para la presencia del pronombre personal sujeto y esto podría afectar al bajo porcentaje de explicitación del sujeto pronominal cuando acompaña a las perífrasis aspectuales. Otro factor que puede estar enmascarando los porcentajes podría ser la clase semántica del verbo o la longitud de la forma verbal.

Progresividad

Los resultados obtenidos para la variable de la progresividad (Tabla 27) muestran una situación similar a la de la perfectividad, ya que, no resultar significativos para la variación de sujeto pronominal y la distribución de los casos está desbalanceada.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
No progresivo	2581	79.22	677	20.85	3258	96.02	1993	74.39	686	25.61	2679	94.93
Progresivo	111	82.22	24	17.78	135	3.98	115	80.42	28	19.58	143	5.07
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 0.5412432, g.d.l: 1, p-valor: 0.461918, Phi: 0.01263003						Chi-cuadrado: 2.299311, g.d.l: 1, p-valor: 0.1294314, Phi: 0.02854436					

Tabla 27. Presencia/ausencia de sujetos pronominales según la progresividad

Las formas no progresivas predominan en nuestros corpus y representan el 94-96% del total de casos. La presencia del pronombre personal sujeto en estos casos ronda entre el 21-25% en cada corpus. En cambio, las formas progresivas tienen una presencia muy escasa en los corpus en torno al 4-5% con, además, una tasa de explicitación del pronombre que varía entre el 18-20%.

Pronominalidad

La variable pronominalidad codifica la distinción entre verbos pronominales y no pronominales. La Tabla 28 muestra, tanto en el corpus de 1995 como en el de 2015, el uso de formas no pronominales (87% para el corpus de 1995 y 88% para el de 2015) es muy similar en los dos cortes temporales y los porcentajes de presencia no distan en gran medida entre las formas pronominales (19% en 1995 y 23% en 2015) y no pronominales (21% en 1995 y 25% en 2015).

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
No pronominal	2321	79.05	615	20.95	2936	86.53	1856	74.45	637	25.55	2493	88.34
Pronominal	371	81.18	86	18.82	457	13.51	252	76.60	77	23.40	329	11.66
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 0.9669664, g.d.l: 1, p-valor: 0.3254379, Phi: 0.01688161						Chi-cuadrado: 0.6000033, g.d.l: 1, p-valor: 0.4385768, Phi: 0.01458137					

Tabla 28. Presencia/ausencia de sujetos pronominales según la pronominalidad

Pese a que la variable no resulta significativa (Tabla 28), nos es de gran utilidad para complementar el estudio de otras variables, especialmente en aquellos casos donde

los verbos pronominales pueden desempeñar un papel desambiguador en situaciones de ambigüedad morfológica o fonética.

2.3.2.4. Variables sociales

Los factores de corte social explorados en este apartado se corresponden con las tres macrovariables empleadas en la estratificación de la muestra recogida en nuestros corpus. Por ello, consideramos el nivel de estudios, que distingue entre nivel bajo, medio y alto, la edad agrupada en tres grupos generacionales —Grupo 1 de 18 a 35 años, grupo 2 de 35 a 55 y el grupo 3 a partir de 55 años—, y el sexo que diferencia entre hombres y mujeres.

Nivel de estudios

La variable del nivel de estudios del informante resulta ser significativa en ambos corpus (valor de p : <0.01), a pesar de que la Tabla 29 muestra que la fuerza de asociación entre las variables es débil (V de Cramer: 0.048 en el corpus de 1995 y 0.055 en el corpus de 2015). Asimismo, tras la observación de los porcentajes de presencia obtenidos según el nivel de formación alcanzado, se precisa que las tasas de presencia del pronombre no siguen la misma tendencia en los dos corpus para cada nivel de estudios.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Nivel bajo	689	80.87	163	19.13	852	25.11	717	72.28	275	27.72	992	35.15
Nivel medio	1038	76.95	311	23.05	1349	39.76	771	77.88	219	22.12	990	35.08
Nivel alto	965	80.96	227	19.04	1192	35.13	620	73.81	220	26.19	840	29.77
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
	Chi-cuadrado: 7.831459, g.d.l: 2, p-valor: 0.01992601, V de Cramer: 0.04804292						Chi-cuadrado: 8.723838, g.d.l: 2, p-valor: 0.01275389, V de Cramer: 0.05560006					

Tabla 29. Presencia/ausencia de sujetos pronominales según el nivel de estudios

Por un lado, los informantes con un nivel de formación medio son los que alcanzan tasas pronominales más elevadas en el corpus de 1995 (23%). En cambio, en el corpus de 2015 se detecta la tendencia contraria: el grupo de informantes con estudios medios es el que tiende hacia una menor explicitación del pronombre y son los hablantes con un nivel

de instrucción bajo (28%) y alto (26%) los que más promueven un mayor uso del pronombre de manera expresa. Estos resultados concuerdan parcialmente con lo expuesto en trabajos previos, como el de Ávila-Jiménez (1996) que apunta que los hablantes con ocupaciones no especializadas y, por lo tanto, un menor nivel educativo, favorecen la presencia del pronombre, mientras que los que han cursado niveles educativos más elevados revelan una menor presencia del sujeto expreso.

Sexo

La variable sexo refleja una ligera variación en los porcentajes de presencia que obtienen las mujeres y los hombres de los dos cortes temporales analizados, a pesar de que estas diferencias no consigan ser estadísticamente significativas en ninguno de los corpus.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Mujer	1330	79.69	339	20.31	1669	49.19	1105	76.42	400	26.58	1505	53.33
Hombre	1362	79.00	362	21.00	1724	50.81	1003	76.16	314	23.84	1317	46.67
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
Chi-cuadrado: 0.2034862, g.d.l: 1, p-valor: 0.6519215, Phi: 0.007744183							Chi-cuadrado: 2.639042, g.d.l: 1, p-valor: 0.1042664, Phi: 0.0305805					

Tabla 30. Presencia/ausencia de sujetos pronominales según el sexo

A primera vista, los resultados reflejan un aumento generalizado de presencia en el corpus de 2015, tanto en hombres como en mujeres, respecto al corpus de 1995. Además, se detecta una tendencia inversa: mientras que en el corpus de 1995 son los hombres los que usan con mayor frecuencia el pronombre personal de forma expresa (21%), en 2015 son las mujeres las que alcanzan una tasa mayor con un 26% de presencia. Aunque esta variación de uso experimentada entre los dos cortes temporales pudiera interpretarse como un posible cambio lingüístico, habría que interaccionar esta variable social con otros factores lingüísticos como la coherencia discursiva o el estilo narrativo más empleado por hombres y mujeres (Bayley & Pease-Alvarez 1997).

En la misma línea, Silva-Corvalán (2003, p. 11) también considera que las variables sociales podrían ser estudiadas en un contexto donde los factores internos estén

controlados para poder así establecer diferencias o correlaciones entre los distintos grupos sociales.

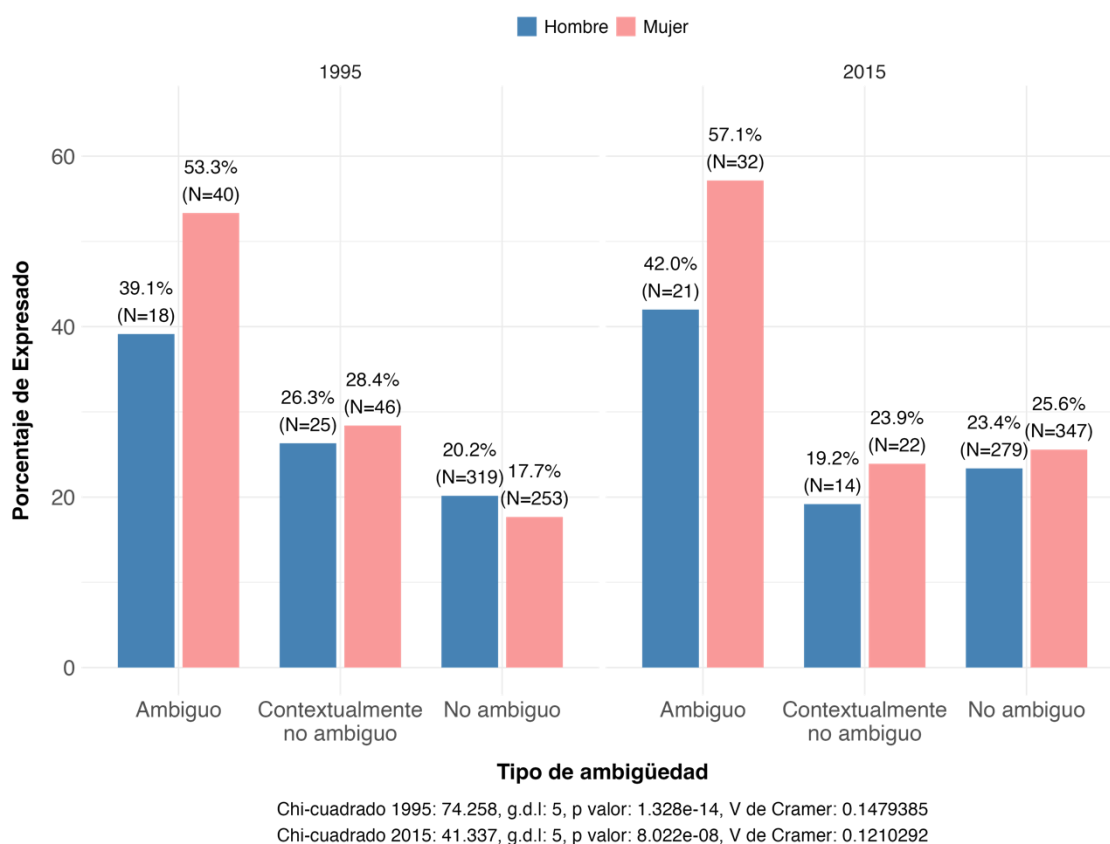


Gráfico 7. Presencia/ausencia de sujetos pronominales según el sexo y la ambigüedad morfológica contextualizada

Por ello, con el propósito de estudiar esta variable social ante un factor lingüístico específico, el Gráfico 7 representa la interacción entre las variables sexo y ambigüedad morfológica contextualizada, ya que la relación de estas dos variables puede darnos información sobre cómo actúan los hombres y las mujeres ante contextos que puedan resultar ambiguos. En una primera observación, detectamos que la interacción de estos factores resulta significativa para la presencia del sujeto. Aunque esta parece afectar ligeramente más al corpus de 1995, en ambos corpus la V de Cramer indica una fuerza de asociación moderada entre el 0.121 y 0.147.

En general, los porcentajes de presencia en los diferentes contextos son similares para hombres y mujeres, aunque se evidencian algunas disparidades. En el corpus de 1995, los hombres y mujeres siguen un mismo patrón de uso: una menor tasa pronominal con formas no ambiguas (20% de presencia los hombres y un 18% de presencia las mujeres);

un aumento progresivo con formas verbales cuya morfología puede ser ambigua, pero que se desambiguan contextualmente (26% y 28%, respectivamente) y, finalmente, ante formas ambiguas, la presencia del pronombre se hace más necesaria, de manera que los hombres explicitan el sujeto un 39% y, especialmente las mujeres alcanzan un 53%.

Sin embargo, las diferencias en función del sexo se ven más acentuadas en el corpus de 2015, donde las mujeres tienden a hacer un mayor empleo que los hombres del pronombre personal sujeto en todos los contextos. Las discrepancias son especialmente pronunciadas en las formas ambiguas, ya que los hombres obtienen un 42% de presencia y las mujeres alcanzan un 57%. Estos resultados pueden sugerir que las mujeres son más conscientes de la ambigüedad que se podría ocasionar en el discurso y, en consecuencia, recurren con mayor frecuencia a la explicitación del sujeto con el fin de asegurar la continuidad discursiva.

Otra forma de explorar los datos sociológicos es estudiar el comportamiento de los individuos. Para ello, consideramos necesario trabajar con proporciones, ya que, si nos basamos en el número de casos, no se tendría en cuenta la aportación dispar de cada individuo a la total de casos analizados. Con el objetivo de profundizar el comportamiento de perfiles sociales que surgen de la combinación de las variables *nivel de estudios* y *sexo*, mostramos en el Gráfico 8 las proporciones de presencia para cada grupo.

A simple vista, los resultados no parecen presentar la misma tendencia en las dos sincronías estudiadas: en el corpus de 1995, la dispersión es relativamente homogénea entre hombres y mujeres, aunque se percibe un aumento moderado en el porcentaje de expresado en el nivel de instrucción medio. Por otro lado, en el corpus de 2015, esta tendencia parece acentuarse, observándose un incremento generalizado en la variabilidad y en los valores medianos, lo que podría sugerir cambios en curso en las dinámicas sociolingüísticas del fenómeno estudiado.

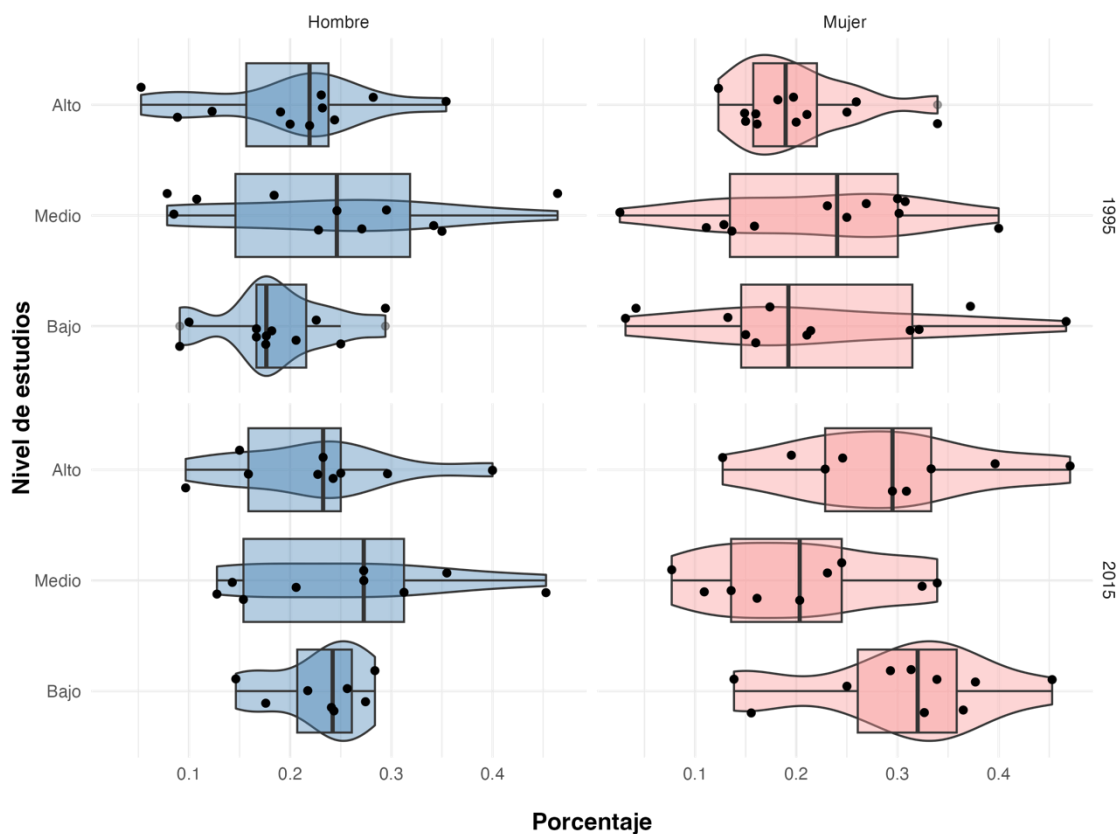


Gráfico 8. Proporción de presencia/ausencia de sujetos pronominales según el nivel educativo y el sexo por individuo

Al comparar las distribuciones presentadas por los hombres de ambos cortes temporales, detectamos que la presencia del pronombre es similar en los tres niveles de estudios de los dos corpus analizados, aunque se aprecia un aumento de la presencia en el corpus de 2015 en el nivel de estudios bajo y medio. Por otro lado, al examinar la distribución entre mujeres de las dos sincronías, detectamos una reducción en la variabilidad dentro de cada nivel de estudios en el corpus de 2015. Además, se observa como las mujeres del corpus de 2015, especialmente aquellas que tienen un nivel de instrucción alto y bajo, fomentan una explicitación pronominal en torno al 30% mayor que el resto perfiles sociológicos analizados en los dos cortes temporales.

Aunque sería demasiado pronto para considerar estos resultados como un cambio lingüístico, habría que explorar en futuros trabajos si la alta tendencia a la explicitación pronominal sigue dándose por parte de las mujeres y, además, si esta variación sintáctica pudiera llegar a ser un rasgo que se estuviera integrando dentro de la variedad intermedia malagueña, al igual que ha pasado previamente con otros factores fonéticos (Vida Castro

& Villena Ponsoda, 2012; Villena Ponsoda & Vida Castro, 2017; Villena Ponsoda & Vida Castro, 2017b).

Edad

El análisis presentado en la Tabla 31 revela que el factor de la edad no resulta significativo en ninguna de las dos sincronías estudiadas en la ciudad de Málaga. Sin embargo, el grupo etario 3 formado por informantes mayores de 55 años, exhibe la mayor tasa de explicitación pronominal (22% en el corpus de 1995 y un 27% en el corpus de 2015). Los grupos de las generaciones más jóvenes —tanto el Grupo 1 con hablantes entre 18 y 35 años, como el Grupo 2 con edades entre 35 y 55— emplean el pronombre personal sujeto expreso de manera muy similar y consiguen unos porcentajes de presencia en torno al 20% en el corpus de 1995 y sobre un 24% en el de 2015.

	Corpus 1995						Corpus 2015					
	Ø		Pro		Total		Ø		Pro		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Grupo 1	897	80.02	224	19.98	1121	33.04	686	76.14	215	23.86	901	31.93
Grupo 2	950	80.10	236	19.90	1186	34.95	736	75.41	240	24.59	976	34.59
Grupo 3	845	77.81	241	22.19	1086	32.01	686	72.59	259	27.41	945	33.49
Total	2692	79.34	701	20.66	3393	100	2108	74.70	714	25.30	2822	100
Chi-cuadrado: 2.287448, g.d.l: 2, p-valor: 0.3186303, V de Cramer: 0.02596472							Chi-cuadrado: 3.466102, g.d.l: 2, p-valor: 0.1767443, V de Cramer: 0.0350463					

Tabla 31. Presencia/ausencia de sujetos pronominales según la edad

Nuestros resultados se alinean con los obtenidos en otras ciudades (Manjón-Cabeza Cruz et al., 2016; Orozco & Guy, 2008; Prieto Muñoz, 2011) y respaldan la hipótesis de que los informantes de mayor edad tienden a ser más conservadores con los usos del pronombre personal sujeto con el fin de acercarse más a las reglas prescriptivas del español (Flores-Ferrán, 2007), aunque lo cierto es que no parece existir un patrón uniforme.

Expuestas las tres variables macrosociales que nos permiten estratificar en subgrupos el total de nuestra muestra, el Gráfico 9 plasma cómo se distribuyen los casos en función de los perfiles sociales derivados de la combinación de las tres variables sociales estudiadas. En términos estadísticos, esta interacción manifiesta su significatividad en ambas sincronías analizadas (valor de p : <0.000) y, asimismo, la fuerza de asociación entre

las variables consigue ser superior a la obtenida al analizarlas de forma individual (V de Cramer: 0.132 en el corpus de 1995 y 0.125 en el corpus de 2015).

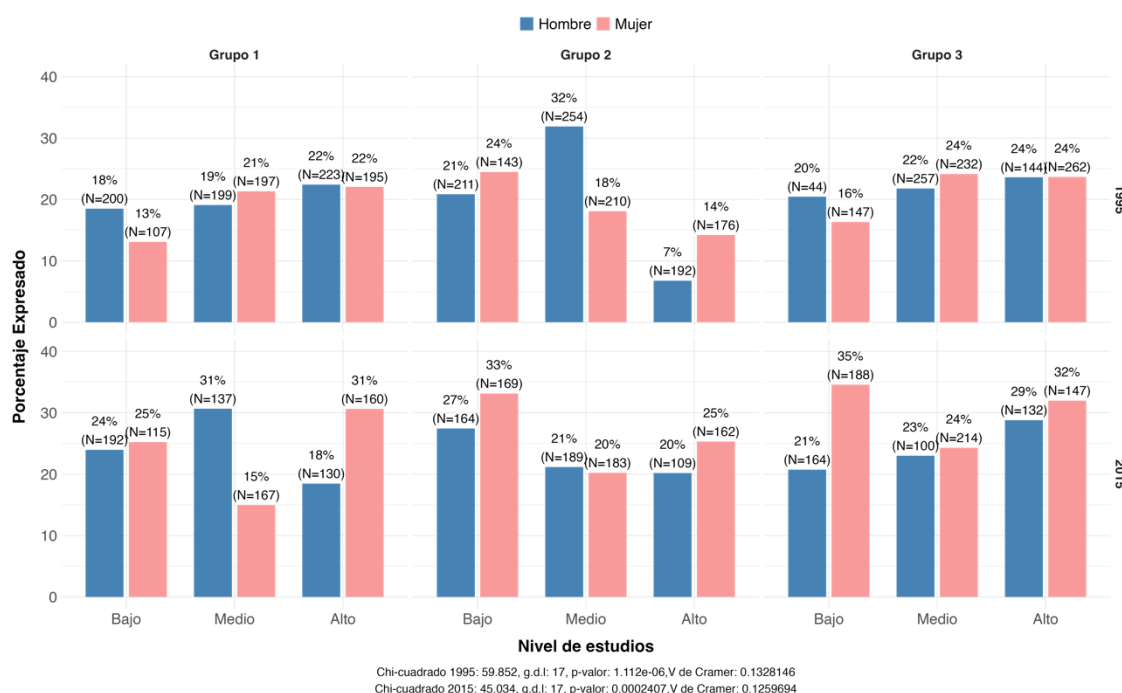


Gráfico 9. Presencia/ausencia de sujetos pronominales según la edad, el sexo y el nivel de estudios

En una primera aproximación, podemos identificar que el corpus de 1995 muestra una presencia del sujeto pronominal más moderada, donde la mayoría de los perfiles sociales obtienen frecuencias de uso entre el 20-22%, a excepción de algunas tasas inferiores en casos como el 7% de los hombres y el 14% de las mujeres del segundo grupo etario con nivel de estudios altos y las mujeres con nivel de estudios bajo tanto del primer grupo de edad (13%) como del grupo de mayor edad (16%). Además, es relevante destacar que los hombres del segundo grupo etario con estudios medios son los que promueven un mayor empleo del sujeto explícito de todo el corpus de 1995, cuyo porcentaje de sujeto expreso alcanza el 32%.

En cambio, el corpus de 2015 refleja una mayor divergencia en el uso del sujeto pronominal entre sus distintos perfiles sociales. En cuanto al grupo etario más joven, parece que las mujeres con estudios altos y los hombres con estudios medios incrementan la explicitación del pronombre hasta un 31%. De igual forma, los informantes del segundo grupo de edad con estudios elementales muestran una mayor la explicitación pronominal con un 27% de presencia en hombres y un 33% en mujeres. Por último, en el tercer grupo

de edad se observan diferencias más pronunciadas cuando los informantes tienen estudios básicos —los hombres registran una presencia del 21% y mujeres alcanzan un 35% de explicitación—, mientras que estas disparidades se reducen en los informantes con estudios superiores, ya que los hombres manifiestan un 29% de presencia y las mujeres un 32%.

Los resultados expuestos nos invitan a reflexionar sobre la posibilidad de un cambio lingüístico, que se ha puesto en marcha en el último corte temporal analizado, donde, principalmente las mujeres en las tres generaciones con estudios básicos y superiores estarían impulsando una mayor aparición del sujeto pronominal en el español hablado en Málaga.

2.3.3. Análisis multivariante

2.3.3.1. Análisis de modelos no paramétricos de regresión: árboles de inferencia y bosques aleatorios

Para una primera visualización general de las variables analizadas, recurrimos a los modelos no paramétricos de regresión, como los árboles de inferencia condicional y los bosques aleatorios, ya que nos permiten representar los resultados de manera visual y, a su vez, contribuyen a la interpretación de la importancia de factores (Martín Butragueño, 2020b). Por otra parte, los árboles de inferencia condicional seleccionan la variable independiente que más relación presenta con la dependiente y, de este modo, se generan particiones recursivas binarias con todas las variables independientes significativas restantes (Levshina, 2015; Scrivner & Díaz-Campos, 2016; Tagliamonte, 2012). Por otra parte, a partir de varios árboles es posible generar un bosque aleatorio capaz de medir la importancia de cada predictor respecto a las demás variables independientes (Levshina 2015, p. 292).

El Gráfico 10 representa el orden de importancia las variables del corpus de 1995, en el eje vertical se ubican las variables predictoras separadas por una línea discontinua roja para diferenciar aquellas que no logran influir en la presencia del pronombre como, en este caso, ocurre con la variable de la progresividad. Es necesario puntualizar que este tipo de análisis no considera los efectos aleatorios de la forma verbal ni del individuo y, de esta manera, los resultados de las variables relacionadas con ellos, como la clase semántica

ADESSE y el tiempo verbal o la edad, el sexo y el nivel de estudios, deben ser interpretados con cautela.

Excluidos estos factores, observamos que, entre las variables más importantes, se encuentra la variable persona del sujeto con un índice más elevado, esta es la variable más importante y se aleja considerablemente del resto, seguidamente se encuentran los factores de tipo de discurso, tipo de cláusula, turno de habla, correferencialidad y ambigüedad fonética contextual. Sin embargo, el resto de las variables aparecen en posiciones más periféricas con un índice de *gini* sobre 20.

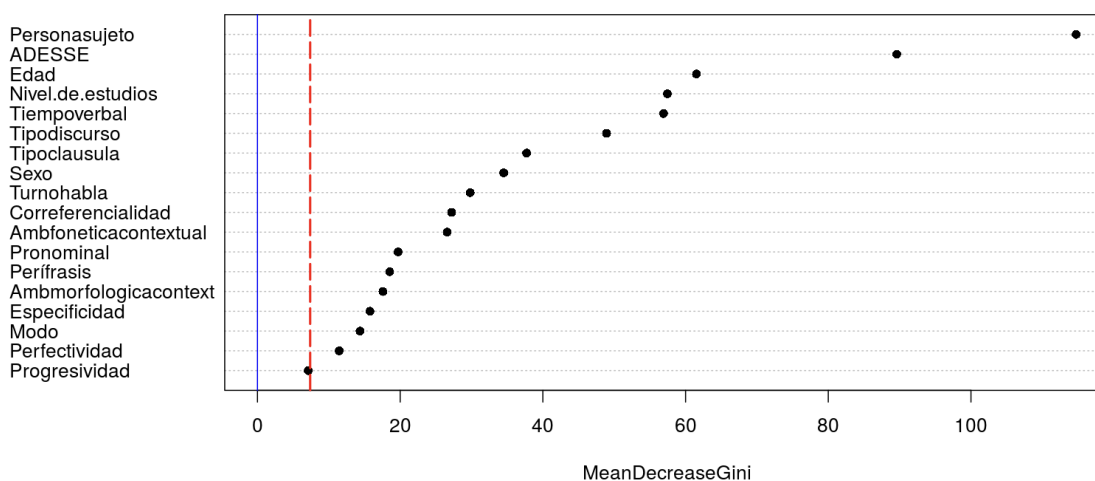


Gráfico 10. Importancia de variables lingüístico-sociales para el corpus 1995

Tras establecer la importancia de las variables, realizamos un árbol de decisión con el objetivo de identificar las variables que inciden con mayor fuerza en la partición del total de casos. El Gráfico 11 representa como, del 100% de casos, el 20% de ellos contienen sujetos expresos. La primera partición se realiza a partir de la variable persona del sujeto que separa, por un lado, los casos de la 3ª persona del singular y los plurales de la 1ª, 2ª y 3ª persona, cuya representación corresponde al 47% de los casos y estos tienen un 92% de ausencia. Por otro lado, la 1ª y 2ª persona del singular (53% de los casos) en contextos de no correferencialidad tienen una presencia del 38% y en situaciones de continuidad referencial solo logran un 24% de explicitación del sujeto. Entre los matices que aporta este gráfico, cabe destacar que, en esta situación las formas no ambiguas contextualmente suponen un 5% del total de casos analizados, pero su presencia ronda el 39%.

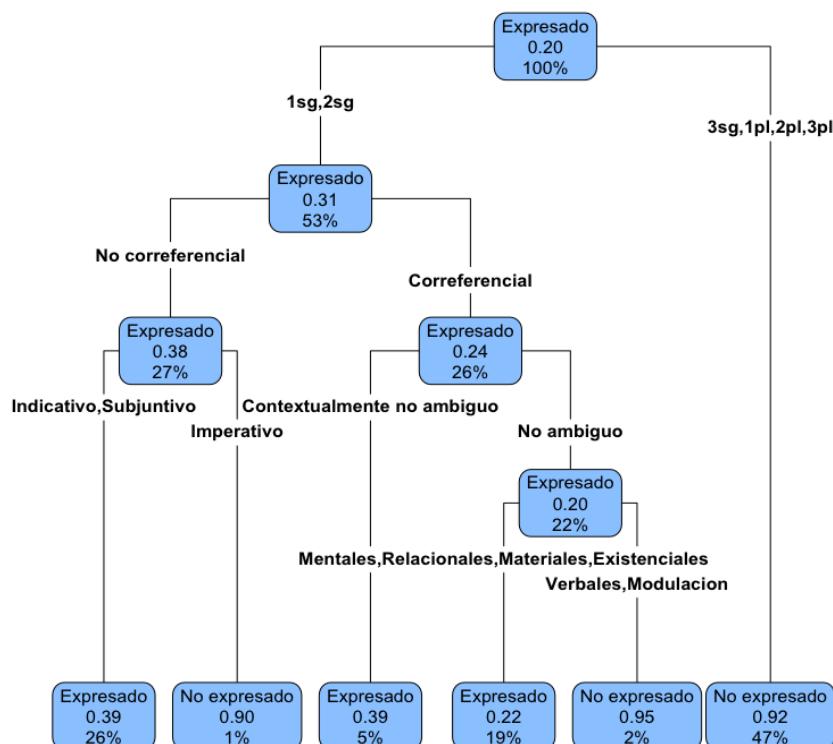


Gráfico 11. Árbol de decisión para el corpus de 1995

Al aplicar estos análisis a los datos referentes al corpus de 2015, visualizamos en el Gráfico 12 el orden de la importancia de las variables.

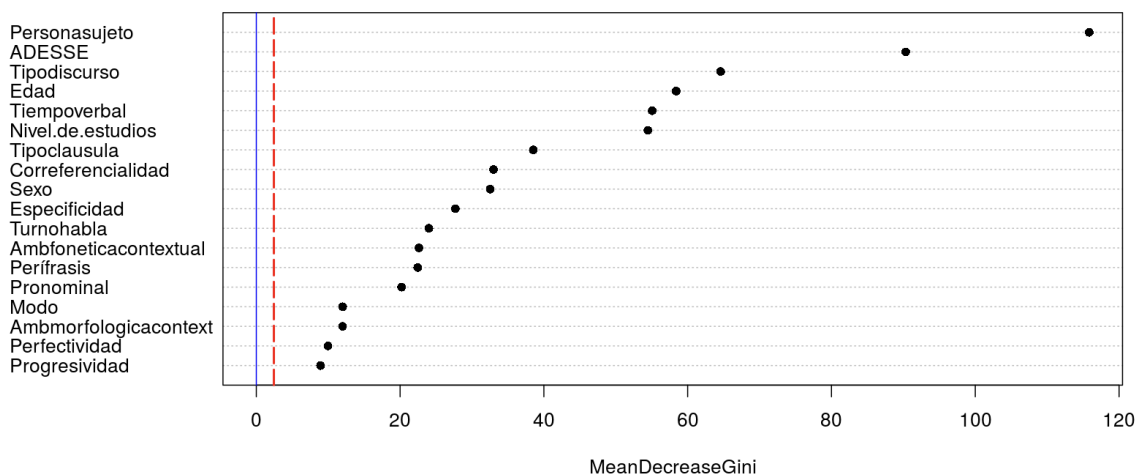


Gráfico 12. Importancia de variables lingüístico-sociales para el corpus 2015

Los factores que ocupan las cinco primeras posiciones coinciden con las variables que resultan ser las más relevantes en el corpus de 1995 (persona del sujeto, ADESSE, tipo de discurso, edad, tiempo verbal, nivel de estudios, tipo de cláusula), aunque se producen

cambios en el orden de aparición. Además, resulta conveniente señalar que en el corpus de 2015 todas las variables parecen tener influencia, ya que, en mayor o menor medida, todas logran superar el umbral de significatividad.

La observación de los datos en el árbol de decisión nos ayuda a identificar, además de las variables que más influyen de forma jerárquica en el corpus de 2015 (Gráfico 13), algunas divergencias entre los dos cortes temporales analizados.

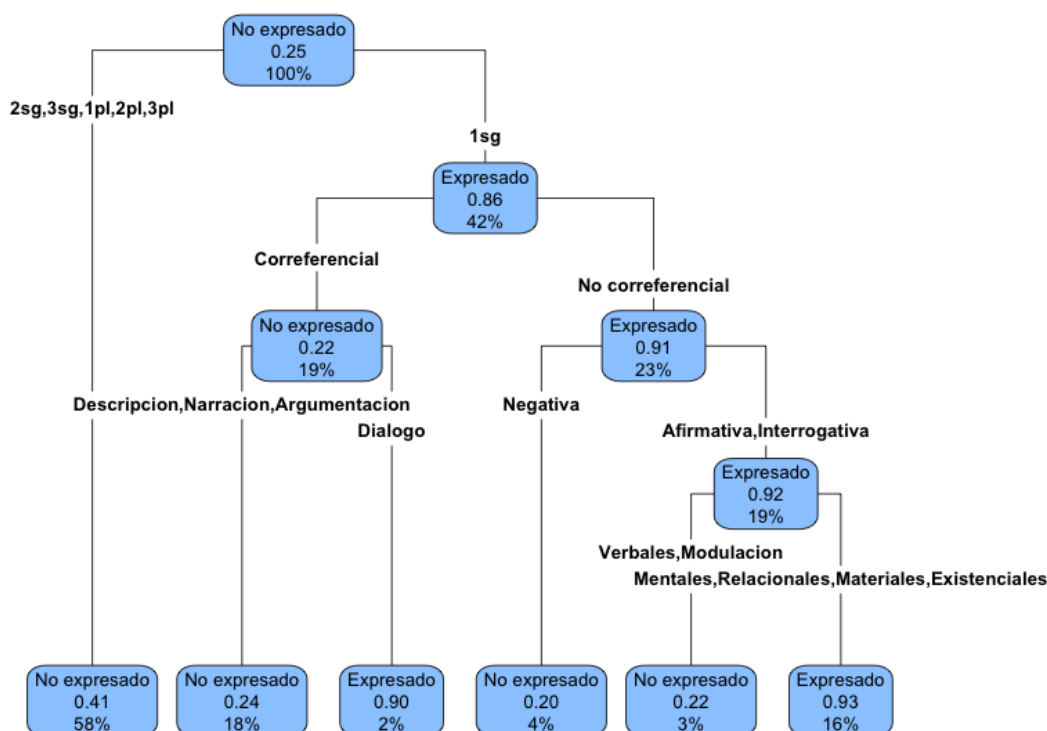


Gráfico 13. Árbol de decisión para el corpus de 2015

Una de las primeras diferencias que presenta el Gráfico 13 reside en el porcentaje de presencia, dado que la tasa de explicitación es más elevada en el corpus de 2015 (25%) que la observada en el corpus de 1995. A pesar de que las dos sincronías coinciden al realizar la primera partición mediante la variable persona del sujeto, en este corte temporal la 1ª persona del singular se separa del resto de personas gramaticales, lo que supone el 58% del total de casos analizados. Dentro de los casos de 1ª persona del singular (42% del total), observamos otras divisiones de la muestra como que la correferencialidad tiende hacia la no expresión del sujeto en situaciones discursivas de descripción, narración y argumentación, mientras que el diálogo rápido parece favorecer la presencia de este. En cambio, los pronombres de 1ª persona del singular incrementan la presencia del pronombre cuando se produce un cambio de referente (23% de los casos) y, especialmente, en

cláusulas afirmativas cuando van acompañadas de un verbo de proceso mental, relacional, material y existencial.

En definitiva, ambos árboles posibilitan la visualización de los puntos de partición sucesivos para las variables significativas y, de esta forma, detectamos que los factores más influyentes para el corpus de 1995 son persona del sujeto, correferencialidad, modo, ambigüedad morfológica y ADESSE, y en el corpus de 2015 resultan más determinantes para las particiones las variables de persona del sujeto, correferencialidad, tipo de discurso, tipo de cláusula y ADESSE.

2.3.3.2. Análisis de regresión logística con efectos mixtos

Con el fin de determinar el efecto de los predictores y de averiguar 1.^a) cuáles de ellos tienen una influencia estadísticamente relevante en el proceso de selección entre la variante pronominal o su ausencia, 2.^a) cuál es la jerarquía de su peso relativo y 3.^a) de qué forma el verbo o el individuo pueden influir en el proceso, llevamos a cabo un análisis multivariable de regresión logística de efectos mixtos mediante el programa *Rbrul* (Johnson, 2016).

El análisis de regresión lineal de efectos mixtos consta de factores fijos, constituidos por variables lingüísticas y sociales (Tabla 33 y Tabla 37), y efectos aleatorios que, en este caso, son los verbos en forma de infinitivo (Tabla 34 y Tabla 38) y los informantes (Tabla 35 y Tabla 39). Con el fin de facilitar la lectura de los análisis, mostramos los resultados en función del corpus en cuestión y, previamente resumimos el modelo estadístico de cada corpus en las Tabla 32 y Tabla 36. En ellas, detallan los efectos fijos y aleatorios estudiados y, además, presentan la coincidencia de los modelos de escalonamiento descendente y ascendente. La primera versión del modelo incluía la variable progresividad, pero debido a la colinealidad ocasionada con la variable perífrasis verbal (GVIF: 5.90), la excluimos del modelo final. A pesar de que el programa también advierte de una posible colinealidad entre las variables persona del sujeto y especificidad, esta es muy baja y no parece afectar a los resultados (GVIF: 1.50).

[a] Modelo para corpus 1995 (N=3393)	
Respuesta binaria	Presencia del sujeto: Expresado vs. No expresado
Efectos fijos	Persona del sujeto, Especificidad, Correferencialidad, Turno de habla, Tipo de discurso, Ambigüedad fonética contextualizada, Ambigüedad morfológica contextualizada, ADESSE, Tiempo verbal, Modo, Efectividad, Perífrasis, Pronominal, Tipo de cláusula, Edad, Sexo, Nivel de estudios
Efectos aleatorios	Informante, Verbo
[b Escalonamiento ascendente y descendente igualado	
Mejor escalonamiento ascendente: Informante (aleatorio) e Infinitivo (aleatorio) y Persona del sujeto (8.66e-64) + Correferencialidad (9.48e-17) + Ambigüedad morfológica contextualizada (0.00039) + Especificidad (0.000377) + Modo (6.93e-05) + Tipo de cláusula (4.29e-05) + Perífrasis (0.00316) + Tipo de discurso (0.037) [valor de p para la construcción del modelo nulo]	
Mejor escalonamiento descendente: Informante (aleatorio) e Infinitivo (aleatorio) y Persona del sujeto (1.14e-61) + Correferencialidad (9.41e-15) + Especificidad (4.74e-08) + Modo (1.67e-07) + Ambigüedad morfológica contextualizada (5.79e-06) + Tipo de cláusula (2.94e-05) + Perífrasis (0.00197) + Tipo de discurso (0.037) [valor de p para la construcción del modelo nulo]	

Tabla 32. Explicación del modelo de regresión logística de efectos mixtos para el corpus de 1995

Los siguientes análisis muestran, en la segunda columna, las estimaciones en forma de coeficiente de *log odds*, cuyos valores positivos indican que la variante favorece la presencia del sujeto, mientras que los negativos promueven su ausencia. Asimismo, en la quinta columna se presentan los pesos de los factores, que pueden interpretarse de la siguiente manera: cuando el peso obtenido tiene un valor superior a 0.5 señala que la variante contribuye a la aparición del sujeto, y aquellos con valores inferiores a 0.5 evidencian que la variante favorece la omisión. Las columnas centrales muestran un desglose del número de casos o *tokens* y la proporción de presencia asociada a cada variante, lo que proporciona una visión más completa de la distribución de las variantes en

el análisis. Tagliamonte (2012) considera que el rango y el orden de selección de los modelos de regresión son métodos alternativos.

A continuación, la Tabla 33 recoge los resultados de los predictores fijos, ordenados según el mejor modelo de escalonamiento descendente, en línea con la recomendación de *Rbrul* sobre la preferencia de uso de los valores de *p* asociados a la exclusión de variables para describir un modelo en un análisis escalonado. De esta forma, el criterio que seguimos para ordenar las variables en función de su importancia es la significatividad obtenida en el modelo descendente.

Variable	<i>log-odds</i>	Casos	P/P+A	Peso centrado
Persona sujeto	<i>p-valor: 1.14e-61</i>		Rango: 59	
2ª sg	1.743	459	0.279	.851
1ª sg	0.886	1335	0.330	.708
2ª pl	-0.259	10	0.100	.436
3ª sg	-0.365	494	0.130	.410
3ª pl	-0.988	534	0.047	.271
1ª pl	-1.018	561	0.075	.265
Correferencialidad	<i>p-valor: 9.41e-15</i>		Rango: 24	
No correferencial	0.608	1584	0.263	0.647
No aplicable	0.096	54	0.222	0.524
Corref. con sujeto	-0.316	1540	0.162	0.422
Corref. con objeto	-0.387	215	0.107	0.404
Especificidad	<i>p-valor: 4.74e-08</i>		Rango: 30	
Especificidad	0.630	2821	0.221	0.652
No específico	-0.630	572	0.135	0.348
Modo	<i>p-valor: 1.67e-07</i>		Rango: 55	
Indicativo	1.054	3239	0.209	0.742
Subjuntivo	0.355	110	0.164	0.588
Imperativo	-1.410	44	0.114	0.196
Ambigüedad morfológica	<i>p-valor: 5.79e-06</i>		Rango: 20	
Ambiguo	0.307	121	0.479	0.576

No ambiguo contextualmente.	0.198	257	0.276	0.549
No ambiguo	-0.504	3015	0.190	0.377
Tipo de cláusula	<i>p-valor: 2.94e-05</i>		Rango: 35	
Afirmativa	0.503	2909	0.202	0.623
Negativa	0.473	386	0.251	0.616
Interrogativa	-0.976	98	0.153	0.274
Perífrasis	<i>p-valor: 0.00197</i>		Rango: 16	
No perifrástico	0.425	3023	0.215	0.605
Modal	-0.207	145	0.159	0.448
Aspectual	-0.218	225	0.129	0.446
Tipo de discurso	<i>p-valor: 0.037</i>		Rango: 10	
Argumentación	0.253	1103	0.243	0.563
Diálogo	-0.014	101	0.228	0.497
Narración	-0.098	2023	0.188	0.476
Descripción	-0.141	166	0.181	0.465
N: 3393, g.d.l.: -4.033, Proporción: 0.207, Input: 0.017, <i>log. likelihood</i> : -1438.021, AIC: 2922.043, AICc: 2922.37, Dxy.fixed: 1 Dxy.total: 0.639, R ² fijo: 0.278, R ² aleatorio: 0.098, R ² total: 0.376				

Tabla 33. Regresión logística efectos fijos corpus 1995

En el modelo presentado en la Tabla 33, se observa que la aportación de los efectos fijos es considerable, con un R^2 : 0.278, lo que significa una explicación del 27.8%, mientras que los efectos aleatorios solo contribuyen en un 9.8%. En cuanto a las variables fijas que resultan más importantes para la aparición del sujeto pronominal tras el análisis de escalonamiento ascendente y descendente en el corpus de 1995, detectamos que la persona del sujeto es la que obtiene una mayor significatividad y la 2ª (0.85) y la 1ª (0.70) persona del singular son los factores que alcanzan un peso centrado superior a 0.5, lo que parece indicar que estos factores contribuyen a una mayor presencia del sujeto. De igual modo, las siguientes variables y variantes afectan positivamente a la explicitación pronominal: la correferencialidad (no correferencial y no aplicable), la especificidad (específico), el modo (indicativo y subjuntivo), la ambigüedad morfológica contextual (ambiguo y contextualmente no ambiguo), el tipo de cláusula (afirmativa y negativa), la perífrasis verbal (no perifrástica) y el tipo de discurso (narración).

Debido a que el listado completo de los efectos aleatorios (verbos en infinitivo e individuos) es muy extenso, presentamos en la Tabla 34 solo los primeros diez casos cada efecto aleatorio.

Variable	<i>log-odds</i>	Casos	P/P+A	Peso centrado
Creer	0.796	40	0.675	0.689
Salir	0.462	35	0.343	0.613
Nacer	0.410	10	0.500	0.601
Acordarse	0.378	49	0.429	0.593
Ir	0.312	151	0.258	0.577
Pensar	0.284	16	0.438	0.570
Vivir	0.282	36	0.306	0.570
Ser	0.270	182	0.209	0.567
Trabajar	0.262	36	0.278	0.565

Tabla 34. Regresión logística efectos aleatorios (Infinitivo) corpus 1995

En una primera visualización de los datos, se evidencia como el verbo *creer* alcanza un peso centrado de 0.68, lo que indica una fuerte inclinación hacia la presencia del sujeto y se corrobora con un 67.5% de explicitación pronominal, una tasa muy superior de la que obtienen el resto de los verbos. Además, otros verbos relacionados con los procesos mentales como *acordarse* o *pensar* también obtienen un peso que favorece a la presencia (0.59 y 0.57, respectivamente) y una tasa pronominal en torno al 43%.

En cuando al efecto aleatorio de informante, identificamos que los cuatro primeros sujetos obtienen un peso superior a 0.7, de los cuales dos de ellas son mujeres con una edad entre 35 y 55 años y con un nivel de estudios bajo y los otros dos son hombres con un nivel formativo medio y alto y pertenecientes a dos grupos de edad distintos (entre 18 y 35 años y entre 35 y 55 años).

Variable	<i>log-odds</i>	Casos	P/P+A	Peso centrado
MA_M21_008	1.098	43	0.372	0.745
MA_H13_715	1.057	48	0.354	0.737
MA_H22_731	1.050	84	0.464	0.736
MA_M21_069	0.917	28	0.321	0.709

MA_H32_705	0.807	105	0.295	0.686
MA_M21_906	0.785	32	0.312	0.681
MA_M32_722	0.700	65	0.308	0.662
MA_M22_709	0.639	63	0.302	0.649
MA_H32_214	0.629	40	0.350	0.646
MA_M12_729	0.614	80	0.300	0.643

Tabla 35. Regresión logística efectos aleatorios (Informante) corpus 1995

Para llevar a cabo el análisis del corpus de 2015, se han excluido los 20 casos de 2ª persona del plural, ya que presentan un 100% de ausencia y, por lo tanto, condicionan los resultados. Al igual que en el corpus anterior, la persona del sujeto y la especificidad ocasionan colinealidad, pero esta resulta ser tan baja que no parece afectar a los resultados (GVIF: 2.51). Por lo tanto, la Tabla 36 sintetiza el modelo para el corpus de 2015 y presenta la coincidencia del modelo de escalonamiento ascendente y descendente y, a partir de los valores de significatividad de este último, se organizan los predictores presentados en la Tabla 37.

[a] Modelo para corpus 2015 (N=2802)

Respuesta binaria	Presencia del sujeto: Expresado vs. No expresado
Efectos fijos	Persona del sujeto (excluida la 2ª persona del plural), Especificidad, Correferencialidad, Turno de habla, Tipo de discurso, Ambigüedad fonética contextualizada, Ambigüedad morfológica contextualizada, ADESSE, Tiempo verbal, Modo, Efectividad, Perífrasis, Pronominal, Tipo de cláusula, Edad, Sexo, Nivel de estudios
Efectos aleatorios	Informante, Verbo

[b] Escalonamiento ascendente y descendente igualado

Mejor escalonamiento ascendente: Informante (aleatorio) e Infinitivo (aleatorio) y Persona del sujeto ($5.89\text{e-}64$) + Correferencialidad ($2.01\text{e-}16$) + Especificidad ($1.88\text{e-}12$) + Modo ($5.58\text{e-}05$) + Tipo de cláusula (0.00374) + ADESSE (0.0168) [valor de p para la construcción del modelo nulo]

Mejor escalonamiento descendente: Informante (aleatorio) e Infinitivo (aleatorio) y Persona del sujeto ($2.59e-57$) + Correferencialidad ($3.6e-17$) + Especificidad ($2.05e-14$) + Modo ($1.84e-05$) + Tipo de cláusula (0.0026) + ADESSE (0.0168)

[valor de p para la construcción del modelo nulo]

Tabla 36. Explicación del modelo de regresión logística de efectos mixtos para el corpus de 2015

De la explicación del modelo, podemos extraer que en este corpus los efectos fijos aportan una mayor explicación (R^2 : 0.328) que en el corpus de 1995 y, por otro lado, los efectos aleatorios tienen una contribución ligeramente menor que en el corpus anterior (8%).

Variable	log-odds	Casos	P/P+A	Peso centrado
Persona sujeto	<i>p-valor</i> : $2.59e-57$		Rango: 58	
2ª sg	1.463	410	0.241	.812
1ª sg	0.993	1192	0.411	.730
3ª sg	-0.550	310	0.139	.366
3ª pl	-0.704	508	0.075	.331
1ª pl	-1.203	382	0.115	.231
Correferencialidad	<i>p-valor</i> : $3.6e-17$		Rango: 24	
No correferencial	0.312	1455	0.318	0.577
Corref. con objeto	0.305	192	0.271	0.576
No aplicable	0.052	46	0.283	0.513
Corref. con sujeto	-0.669	1109	0.169	0.339
Especificidad	<i>p-valor</i> : $2.05e-14$		Rango: 42	
Especificidad	0.907	2138	0.304	0.712
No específico	-0.907	664	0.098	0.288
Modo	<i>p-valor</i> : $1.84e-05$		Rango: 50	
Subjuntivo	0.756	105	0.190	0.680
Indicativo	0.736	2650	0.260	0.676
Imperativo	-1.492	47	0.128	0.184
Tipo de cláusula	<i>p-valor</i> : 0.0026		Rango: 14	

Afirmativa	0.340	2404	0.260	0.584
Interrogativa	-0.096	67	0.313	0.476
Negativa	-0.244	331	0.205	0.439
ADESSE	<i>p-valor</i> : 0.0168		Rango: 32	
Existenciales	0.645	123	0.358	0.656
Mentales	0.170	655	0.376	0.542
Relacionales	0.147	647	0.233	0.537
Materiales	0.117	1000	0.212	0.529
Modulación	-0.366	109	0.165	0.409
Verbales	-0.713	268	0.160	0.329
N: 2802, d.f.: -3.24, Proporción: 0.207, Input:0.038, log. likelihood: -1279.224, AIC: 2598.449, AICc: 2598.75, Dxy.fixed: 1 Dxy.total: 0.652, R ² fijos: 0.328, R ² aleatorios: 0.08, R ² total: 0.409				

Tabla 37. Regresión logística efectos fijos corpus 2015

Ordenadas según la jerarquización del modelo, las variables que favorecen la aparición del sujeto pronominal en el corpus de 2015 son: la persona del sujeto, en la que los factores de la 2ª (0.81) y la 1ª (0.73) persona del singular son los que promueven una mayor explicitación; la correferencialidad, donde los predictores de no correferencial (0.57) y correferencial con objeto (0.57); los referentes específicos (0.72); el modo subjuntivo (0.68) e indicativo (0.67); la cláusula afirmativa (0.58) y, como novedad respecto al corpus anterior, la clasificación ADESSE con verbos de procesos existenciales (0.65), mentales (0.54), relacionales (0.53) y materiales (0.52).

En cuanto al efecto aleatorio del verbo en infinitivo, el verbo *creer* vuelve a ser el que más patrocina la presencia del pronombre con un peso de 0.69 y una presencia del 74%, superior a la obtenida en el corpus anterior. Entre los verbos que alcanzan mayores tasas de explicitación pronominal, se encuentran los relacionados con procesos existenciales como *vivir* (0.63) con un 61% y *estar* (0.6) con un 31%. Asimismo, los verbos relacionados con procesos mentales obtienen porcentajes de aparición del sujeto muy elevados, como es el caso del verbo como *pensar* (0.61) que logra un 71% de presencia o *recordar* (0.52) que alcanza el 61%.

Variable	log-odds	Casos	P/P+A	Peso centrado
Creer	0.840	100	0.740	0.698
Vivir	0.546	26	0.615	0.633
Pensar	0.451	28	0.714	0.611
Estar	0.408	145	0.310	0.600
Fijarse	0.299	9	0.444	0.574
Montar	0.288	4	0.750	0.571
Apuntarse	0.283	4	0.750	0.570
Entrar	0.245	24	0.333	0.561
Poder	0.223	47	0.213	0.555
Pagar	0.218	17	0.235	0.554
Recordar	0.211	13	0.615	0.552

Tabla 38. Regresión logística efectos aleatorios (Infinitivo) corpus 2015

En cuanto al efecto aleatorio del individuo en el corpus de 2015, la Tabla 39 muestra cómo los cuatro individuos que más contribuyen a la presencia del pronombre son cuatro mujeres. Todas ellas pertenecen al tercer grupo de generacional con una edad comprendida entre 35 y 55 años, aunque tienen niveles educativos diferentes. Mientras que los hombres que más favorecen la presencia del sujeto pronominal son tres individuos pertenecientes al primer grupo de edad con un nivel de instrucción medio y bajo.

Variable	log-odds	Casos	P/P+A	Peso centrado
MA_M32_008	0.939	74	0.324	.715
MA_M31_025	0.755	51	0.471	.677
MA_M33_048	0.724	51	0.471	.670
MA_M31_015	0.612	63	0.365	.644
MA_H11_052	0.575	37	0.243	0.636
MA_H12_007	0.555	53	0.453	0.631
MA_M21_042	0.454	61	0.377	0.608
MA_M23_002	0.414	65	0.308	0.598
MA_M22_034	0.408	56	0.339	0.596
MA_H12_055	0.399	32	0.312	0.594

Tabla 39. Regresión logística efectos aleatorios (Informante) Corpus 2015

En definitiva, los análisis desglosados por corpus constituyen un paso fundamental poder establecer las posteriores comparaciones diacrónicas y diatópicas y, además, observar con mayor precisión el efecto que producen los factores aleatorios del verbo en infinitivo y el informante. Aunque en este apartado solo se han presentado los diez verbos y los diez informantes que más influyen en la explicitación del sujeto pronominal, el listado completo puede consultarse en el Anexo II.

2.4. Comparación de los análisis de los corpus PRESEEA

Una vez obtenidos los resultados de los dos corpus PRESEEA de la ciudad de Málaga, procedemos, en primer lugar, a comparar los resultados de los dos cortes temporales (1995 y 2015) y, posteriormente, cotejamos los factores significativos en los trabajos previos realizados dentro del marco de PRESEEA. Todas estas comparaciones se llevan a cabo con el propósito de trazar un patrón lingüístico que justifique el uso general del sujeto pronominal en español.

2.4.1. Comparación diacrónica: El sujeto pronominal a lo largo del tiempo en Málaga

Con el objetivo de determinar los factores que condicionan la aparición del sujeto en las dos sincronías estudiadas en la ciudad de Málaga, hemos analizado por separado los dos corpus de la ciudad y se han puesto de manifiesto diferencias reveladoras. En una primera observación exploratoria de los datos, ya se percibe una tendencia hacia un aumento progresivo de la explicitación del pronombre, desde un 20.7% en el corpus de 1995 hasta alcanzar un 25.3% en el de 2015, lo que puede reflejar que se estén produciendo cambios en el uso del pronombre por diversos factores lingüísticos y sociales. A pesar de que algunas variables presentan usos dispares en los dos corpus analizados, muchos otros factores se comportan de manera similar a pesar del transcurso del tiempo en la ciudad de Málaga. Ejemplo de ello es el caso de la variable referente del sujeto, donde la presencia del pronombre solo se manifiesta con el rasgo [+humano] y en el resto de los casos no se observa variación entre la ausencia y presencia el sujeto pronominal en ninguna de las dos sincronías. Esta continuidad gramatical podría evidenciar la existencia de una gramática

subyacente que permanece inalterada a lo largo del tiempo y que, probablemente, también se extienda a otras áreas geográficas.

El análisis de regresión logística multivariante con efectos mixtos nos ha permitido medir la importancia de las variables, el peso de los factores en ambos corpus y, a su vez, hemos identificado patrones de convergencia y divergencia entre los dos cortes temporales estudiados. Con el objetivo de explorar detalladamente cada predictor, presentamos, en primer lugar, las variables que han resultado significativas en los dos corpus en los análisis de regresión, ordenadas de mayor a menor significancia según su valor de *p*.

La variable que ejerce mayor influencia en la variación del sujeto pronominal es la persona del sujeto, ya que obtiene los valores de *p* más bajos en ambos corpus. Asimismo, ambas sincronías coinciden al mostrar pesos más altos en la 2ª (0.85 en el corpus de 1995 y 0.81 en el corpus de 2015) y 1ª persona del singular (0.70 en 1995 y 0.73 en 2015). En términos de uso, la 1ª persona del singular mantiene sus elevadas tasas de presencia 33% en 1995 y 41% en 2015; en cambio, la 2ª persona del singular sufre un descenso de las tasas pronominales en el corpus de 2015 que podría estar relacionado con el aumento del carácter inespecífico. Asimismo, la 3ª persona del singular acumula un mayor número de casos en que la segunda del singular en el corpus de 1995, lo que indica que el informante tiende a mencionar otros referentes que no son el entrevistador; sin embargo, en el corpus de 2015 se observa un uso diferente, donde se produce un mayor empleo de la 2ª persona singular (14%) que de la 3ª persona singular (10%).

La segunda variable más importante en los análisis de regresión es la variable pragmático-discursiva de la correferencialidad. Los factores que promueven una mayor presencia del sujeto pronominal son la no correferencialidad (0.64 en 1995, 0.57 en 2015), no aplicable (0.52 en 1995 y 0.51 en 2015) y en el corpus de 2015 también la correferencialidad con objeto (0.57). Estos resultados presentan un patrón de uso distinto, ya que en el corpus de 1995 la explicitación del pronombre cuando es correferencial con objeto es de un 10% y cuando es correferencial con sujeto es de 16%, pero en el corpus de 2015 la tendencia mostrada es la contraria: existe una mayor presencia cuando la correferencialidad se mantiene con la misma función (26%) que cuando es correferencial, pero con función de objeto (16%).

La tercera variable más importante está asociada a una característica propia del sujeto de la oración: la especificidad. A pesar de que los referentes específicos favorecen

la explicitación del pronombre personal sujeto en los dos corpus con un peso de 0.65 en el corpus de 1995 y de 0.71 en el de corpus de 2015, un análisis más pormenorizado según la persona gramatical (Tabla 7) refleja un cambio en el uso del pronombre con carácter inespecífico. En cuanto al corpus de 1995, la aparición del sujeto con referentes específicos y no específicos está muy próxima (27% no específicos y 30% específicos), al igual que su representación en el total de la muestra (5% y 8%, respectivamente). Por su parte, en el corpus de 2015 estas diferencias se acentúan: aumenta el número de casos inespecíficos hasta un 11% y disminuyen los específicos hasta un 3, así como la presencia de los inespecíficos desciende al 26% y se incrementa en los específicos hasta el 35%.

El modo verbal resulta ser la cuarta variable en el modelo de regresión en ambos corpus, a pesar de que los factores tengan diferente importancia: en el corpus de 1995 el indicativo (0.72) contribuye notablemente a la expresión del pronombre personal y, con una influencia considerablemente menor, el subjuntivo (0.58). En cambio, en el corpus de 2015, el modo subjuntivo (0.68) influye en la presencia del pronombre de manera muy similar al indicativo (0.67).

Por último, ambas sincronías coinciden al considerar el tipo de cláusula como una variable significativa dentro de su modelo de regresión, pero no lo hacen con la misma importancia, ya que en el corpus de 1995 esta ocupa la sexta posición por debajo de la ambigüedad morfológica. En los dos corpus, las cláusulas afirmativas son las que obtienen un mayor peso 0.62 en el corpus de 1995 y un 0.58 en el corpus de 2015 y, además, las cláusulas negativas también favorecen la presencia en el corpus de 1995 (0.61).

Además de las similitudes presentadas en los dos cortes temporales, también se encuentran divergencias. Por un lado, el corpus de 1995 incluye en su modelo de regresión las siguientes variables: la ambigüedad morfológica contextualizada, cuyas formas ambiguas (0.57) y formas contextualmente no ambiguas (0.54) obtienen un peso que favorece la presencia del pronombre; las formas no perifrásticas también incrementan la explicitación del sujeto pronominal (0.60); y, por último, el tipo de discurso en secuencias argumentativas logran un peso de 0.56. Cabe señalar que en el corpus de 2015 es significativa la clasificación semántica del verbo según la tipología ADESSE y, especialmente los verbos de procesos existenciales (0.65), mentales (0.54), relacionales (0.53) y materiales (0.52) son los que promueven una mayor aparición del sujeto pronominal.

A pesar de no haber sido significativas en los modelos de regresión, consideramos que algunas variables revelan diferencias en las dos sincronías que merecen ser mencionadas. Por ejemplo, la variable turno conversacional no resulta significativa en la regresión en ninguno de los dos corpus, pero los dos cortes temporales presentan una tendencia similar en cuanto a la discontinuidad discursiva, ya que la no correferencialidad propicia la aparición del sujeto pronominal —un 26% en el corpus de 1995 y un 31% en el de 2015—, al igual que el cambio de turno conversacional (25% en el corpus de 1995 y un 34% en el corpus de 2015). No obstante, el corpus de 1995 parece presentar un mayor dinamismo entre el entrevistador y el entrevistado (un 72% en el mismo turno, frente a un 27% en el cambio de turno), mientras que en el corpus de 2015 el entrevistado tiende a monopolizar más la conversación con un 82% en el mismo turno.

Dada su relevancia, la interacción entre el turno conversacional y la correferencialidad pone de manifiesto que, aunque el patrón de uso en ambos corpus sea similar, las diferencias resultan más abruptas en el corpus de 2015. Gracias a esta representación visual de los datos, podemos extraer dos conclusiones. Por un lado, la presencia del pronombre es menor cuando hay continuidad referencial (con sujeto y con objeto) y esta tiene lugar en la misma intervención del informante, pero se incrementa levemente cuando se produce el cambio de turno conversacional. Por otro lado, la diferencia más relevante reside en el cambio referencial, el cual logra tasas de explicitación del sujeto pronominal más elevadas cuando la cláusula se encuentra en el mismo turno conversacional (25-29%) y, especialmente, cuando se produce cambio de turno y el cambio de referente (27-36%). Por lo tanto, parece que la correferencialidad afecta en mayor medida a la presencia del pronombre y, además, ciertas situaciones discursivas —como el cambio de turno de habla y la no correferencialidad— favorecen notablemente la presencia del pronombre.

Otro aspecto relevante en nuestro análisis es la consideración del contexto en las variables de ambigüedad verbal, tanto fonética como morfológica, ya que, en una primera exploración de los datos, las formas morfológicamente ambiguas parecen promover una mayor presencia del pronombre (aproximadamente un 33%). Sin embargo, excluidos posteriormente los elementos que desambiguan la forma verbal (pronombres átonos, mantenimiento del referente), los resultados muestran un incremento de las tasas pronominales en las formas ambiguas que alcanza entre el 48-50% de presencia del sujeto.

Al analizar la ambigüedad fonética sucede algo similar, aunque en un primer momento esta variable no resulta significativa en el corpus de 1995 y los porcentajes de presencia obtenidos son considerablemente bajos (entre un 19-20%). Cuando se estudia la variable de manera contextualizada, la presencia aumenta hasta un 38% en el corpus de 1995 y un 35% en el de 2015 y obtiene un valor de *p* más bajo en los análisis descriptivos.

Además, la interacción entre la variable de ambigüedad fonética potencial y especificidad ha mostrado un cambio en el uso de la 2ª persona del singular en los dos cortes temporales. De esta manera, en el corpus de 1995, la presencia del sujeto no muestra grandes diferencias por el rasgo de especificidad del referente, lo que parece indicar que la explicitación de este varía en función de la ambigüedad fonética. Sin embargo, en el corpus de 2015, las diferencias respecto a la especificidad del referente sí son relevantes, ya que la presencia del pronombre con referentes específicos logra un 43% y con no específicos se reduce a un 27%. Por tanto, podría tratarse de un cambio en el empleo del pronombre personal de 2ª persona en comparación con el corpus anterior, de modo que, aunque la ambigüedad fonética potencial condiciona la presencia del sujeto, detectamos que en los referentes específicos predomina una función más deíctica y en los referentes no específicos se mantiene la omisión pronominal a pesar de las situaciones de ambigüedad fonética.

En la variable tipo de discurso, la frecuencia de la tipología discursiva empleada parece haber experimentado cambios en estos 20 años. Mientras que en el corpus de 1995 se hacía un mayor uso de las secuencias narrativas y argumentativas, en el corpus de 2015 la narración sigue siendo la más empleada la narración y, como novedad, la descripción obtiene un alto porcentaje de empleo. En términos de presencia del sujeto pronominal, parece encontrarse la misma tendencia en ambos corpus, ya que la argumentación y el diálogo rápido son los tipos discursivos que favorecen notablemente la aparición.

Dentro de las variables referidas al verbo, localizamos un uso divergente en cada corpus según la clase semántica del verbo. Por un lado, encontramos similitudes como el mayor uso de los verbos de procesos materiales y la presencia del pronombre más elevada en los verbos de procesos mentales; pero, además, en el corpus de 2015 también se detecta que los verbos existenciales aumentan considerablemente la aparición del sujeto (35%). Entre las diferencias más notables detectadas el transcurso de 20 años en la ciudad cabe destacar la combinación de la 1ª persona del singular y el verbo creer, ya que no solo ha

duplicado la frecuencia en cuanto a su uso, sino que también su tasa de presencia del sujeto alcanza un 73.2%. Estos resultados podrían indicar una tendencia hacia el uso gramaticalizado de la forma verbal acompañada de la presencia del sujeto.

Por otro lado, los tiempos verbales presentan un patrón bastante similar en las dos sincronías: el indicativo promueve un mayor uso y una tasa de presencia más elevada, mientras que el presente, el imperfecto y el pretérito acumulan mayor número de casos y la presencia del pronombre es mayor en tiempo presente e imperfecto. En cambio, algunas variables solo han sido significativas en uno de los corpus, como es el caso de la perfectividad en el corpus de 2015 y la perífrasis verbal en el corpus de 1995.

Por último, las variables sociales no parecen incidir con la misma fuerza que las variables lingüísticas en la explicación del fenómeno analizado. De las tres variables sociales contempladas, solo se ha mostrado ser significativa el nivel de estudios en ambos cortes temporales. Sin embargo, no presentan la misma tendencia en el uso de la aparición del sujeto pronominal, puesto que son los informantes con un nivel de formación medio los que alcanzan tasas pronominales más elevadas en el corpus de 1995 (23%), pero en el corpus de 2015 la tendencia es justo la contraria: el grupo de nivel medio es que logra menores tasas pronominales y los de nivel bajo y alto son los que más patrocinan el uso del pronombre (28% y 26%, respectivamente). A pesar de no ser significativas, los resultados según la variable edad sí parecen ser más consecuentes, ya que el grupo de mayor edad coincide al ser el que promueve una mayor aparición el pronombre en ambos corpus, y el sexo tampoco parece presentar el mismo patrón en los dos cortes temporales.

Las interacciones entre las variables sociales han sido muy reveladoras a la hora de identificar las tendencias de explicitación pronominal según el perfil social analizado. La diferencia más notable detectada entre los dos cortes temporales es que, mientras que los hombres permanecen con unos porcentajes de presencia estables, las mujeres con nivel de estudios bajo y alto han experimentado un aumento considerable en la presencia del pronombre, lo que podría indicar un cambio lingüístico en curso. Sería conveniente replicar el análisis en el corpus que se está recogiendo actualmente en la ciudad de Málaga (Cruces Rodríguez, en preparación) con el fin de observar si la tendencia detectada en 2015 se acentúa o, por el contrario, se atenúa en datos más recientes.

En definitiva, el estudio comparativo de los dos corpus en la ciudad de Málaga ha subrayado muchos matices diferenciadores entre las dos sincronías, pero, a su vez, se han

señalado similitudes en ciertas variables que promueven la aparición del pronombre, como la persona del sujeto, la especificidad, la correferencialidad, el tipo de cláusula y el modo. Esta continuidad en cuanto a los factores condicionantes, en este caso de naturaleza lingüística, podría indicar la existencia de una gramática estable que se mantiene a pesar del transcurso del tiempo. Además de estas variables comunes, otros factores particulares condicionan la presencia en cada corpus, como la ambigüedad morfológica contextualizada, el tipo de discurso o la perífrasis verbal han sido incluidas en el modelo de regresión en el corpus de 1995 y, en el caso del corpus de 2015, la clasificación semántica de los verbos (ADESSE).

2.4.2. Comparación diatópica: El sujeto pronominal en la geografía hispana

Uno de los objetivos propuestos en este trabajo reside en la posibilidad de trazar un patrón general sobre los usos del pronombre personal que sea extrapolable a las distintas zonas de habla hispana. Para ello, procedemos a comparar los resultados obtenidos en la ciudad de Málaga con trabajos previos desarrollados en otras comunidades de habla hispana. Gracias a los estudios sobre la expresión del sujeto pronominal dentro del ámbito de PRESEEA, disponemos de un gran conjunto de datos de distintas zonas de hispanohablantes con una metodología parcialmente homogénea y, de este modo, las conclusiones alcanzadas podrían ser considerablemente sólidas y transversales.

En primer lugar, la Tabla 40 presenta las ciudades que han sido estudiadas hasta la fecha, detallando la comunidad de habla, la muestra utilizada, el número de casos analizados y el porcentaje de presencial pronominal. Ahora bien, las similitudes y diferencias encontradas deben interpretarse con cautela, puesto que algunos de los trabajos previos presentan ciertas diferencias metodológicas en cuanto a las personas del sujeto tratadas, las diferencias entre las muestras y los niveles de estudios considerados.

Ciudad	Estudio	Muestra	N	%
Caracas, Venezuela	Martínez Lara (2013)	36	1766	48.3
Barranquilla, Colombia	Pérez Córdoba y Gomes-Camacho (2019)	18	2700	45.6

Cartagena, Colombia	Pérez Córdoba y Gomes Camacho (2019)	18	2700	45.3
Valledupar, Colombia	Pérez Córdoba y Gomes Camacho (2019)	18	2700	41.8
Caracas, Venezuela	Martínez Lara (2016)	12	552	41
Medellín, Colombia	Martínez Lara (2016)	12	176	41
Montevideo, Uruguay	Martínez Lara (2016)	12	396	41
San Juan de Puerto Rico	Claes (2011)	18	2700	38.8
San Juan de Puerto Rico	Prieto Muñoz (2011)	12	4322	34.2
Caracas, Venezuela	Pérez Brabandere (2010)	36	3600	30
Cali, Colombia	Orozco y Orozco (2022)	--	2679	29.3
Cali, Colombia	Orozco et. al. (2024)	72	6409	28.6
Medellín, Colombia	Orozco y Hurtado (2021)	119	4623	28
Bogotá, Colombia	Prieto Muñoz (2011)	12	4906	26.3
Santiago de Chile, Chile	Martínez Lara et. al (2021)	54	9191	25.3
Málaga (2015), España	Segura Lores	54	2833	25.2
Sevilla, España	Repede (2019)	24	879	24.8
Valencia, España	Roselló Verdeguer (2021)	72	3261	21.8
Ciudad México, México	Lastra y Martín Butragueño (2015)	18	2040	21.7

Málaga (1995), España	Segura Lores	72	3382	20.7
Montevideo, Uruguay	Martínez Lara (2022)	12	446	20
Granada, España	Manjón-Cabeza Cruz et. al (2016), Manjón-Cabeza Cruz (2022)	54	3415	17.7
Valencia, España	Prieto Muñoz (2011)	12	3977	16.2

Tabla 40. Tasas de presencia del sujeto en distintas zonas de habla hispana

Si atendemos a la clasificación propuesta por Martín-Butragueño (2020), la cual ordena las diferentes áreas hispanas en tres grandes bloques en función del porcentaje de presencia obtenido, las ciudades que aquí comparamos se organizarían de la siguiente manera:

-Presencia alta ($\approx 40\%$): Este grupo incluye a Caracas, Barranquilla, Cartagena, Valledupar, Medellín y Montevideo. Sin embargo, debemos considerar que los trabajos de Caracas, Montevideo y Medellín (Martínez Lara 2013, 2016) solo contemplan la 3ª persona del singular en sus análisis.

-Presencia moderada ($\approx 30\%$): En esta categoría se encuentran las ciudades de San Juan de Puerto Rico, Caracas, Medellín y Cali.

-Presencia baja ($< 30\%$): En este bloque se enmarcan ciudades tanto de Iberoamérica —Medellín, Bogotá, Santiago de Chile, Ciudad de México y Montevideo (solo 3ª persona del singular)— como de España —Málaga, Granada, Valencia y Sevilla (nivel alto)—, lo que refleja una mayor preferencia por la omisión del sujeto en España en comparación con las demás zonas geográficas.

Los siguientes mapas dibujados en el Gráfico 14 muestran comparativamente los porcentajes de presencia de sujeto pronominal obtenidos en distintas variedades del español. Los tres bloques establecidos por Martín Butragueño (2020) se distinguen en el gráfico mediante los tres colores especificados en la leyenda.

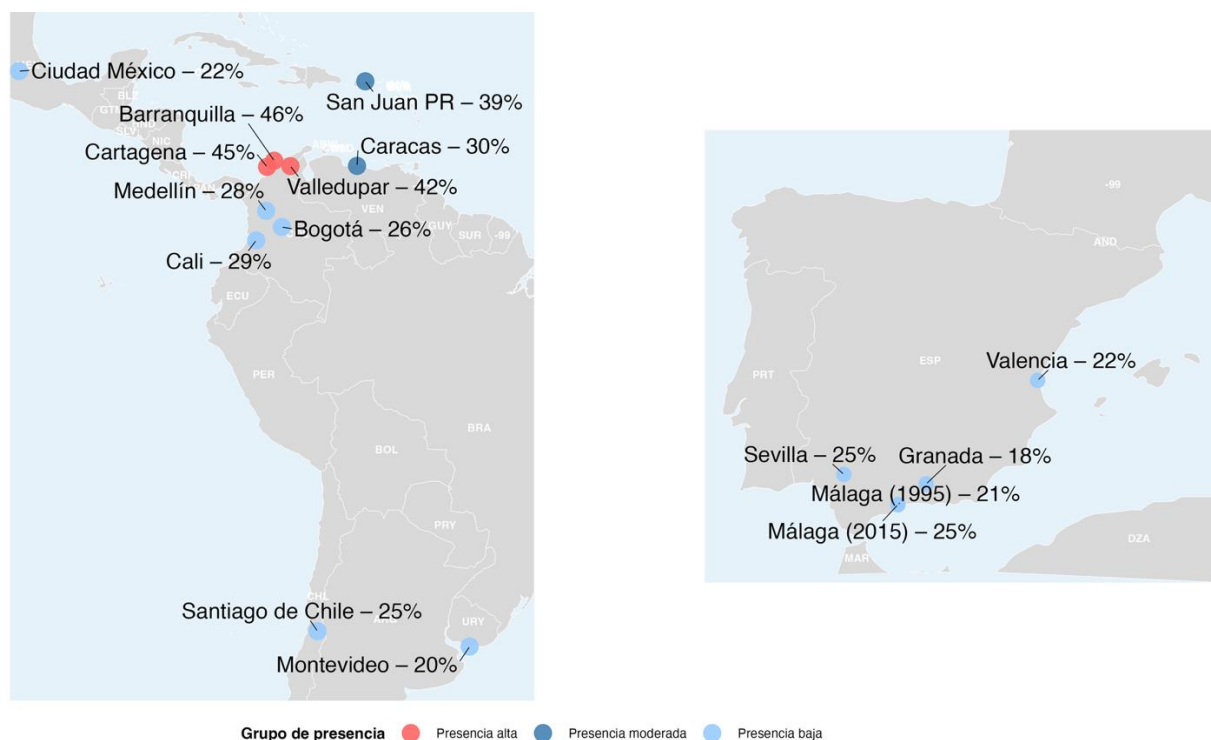


Gráfico 14. Mapa de distribución de trabajos dentro de PRESEEA

A pesar de que las diferencias entre las tasas pronominales pueden aportar una visión general de la aparición del sujeto, Claes (2011, p. 209) subraya que las comparaciones deben hacerse teniendo en cuenta las variables que resultan significativas, la jerarquización de las restricciones, el peso relativo de las variables y la jerarquización de las subvariables. Dadas estas consideraciones y con el objetivo de establecer comparaciones lo más equilibradas posible, a partir de ahora solo confrontamos los estudios que han tratado todas las personas del sujeto y todos los niveles de estudios, así como los que han realizado análisis estadísticos más complejos que permitan evaluar el peso de los factores. No obstante, cabe recordar que los trabajos no han abordado todas variables y tampoco han recibido el mismo tratamiento estadístico en cada estudio, ya que algunos sí incorporan como factores aleatorios el individuo y el verbo.

Para establecer una comparación progresiva y detallada, el Gráfico 15 presenta las variables que han resultado ser significativas en los trabajos previos y su orden de importancia, aunque el criterio de jerarquización de las variables no ha sido el mismo en todas las investigaciones, ya que algunas se han basado en el rango y otros en la incorporación progresiva de las variables al modelo estadístico.

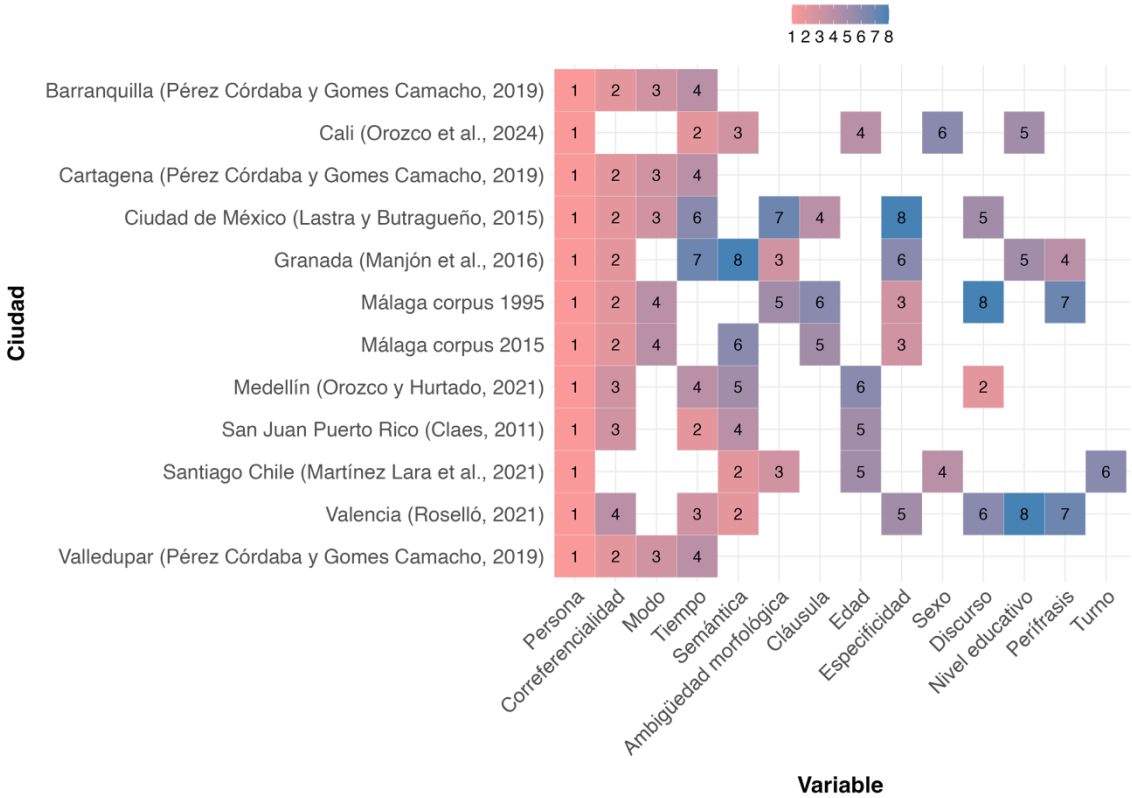


Gráfico 15. Variables significativas en trabajos previos ordenadas según importancia

En una primera visualización del Gráfico 15, se detecta una convergencia en todos los trabajos, ya que coinciden al considerar la persona del sujeto como la variable que más influyente en la explicitación del sujeto. Además, al identificar los factores que más promueven esta aparición, se evidencia un consenso en la mayoría de los estudios: la 1ª persona del singular es la que alcanza pesos probabilísticos más elevados y en algunos trabajos también la 2ª persona del singular. De hecho, en ciertas ciudades, como Málaga 1995, Cali y Medellín, el peso probabilístico de la 2ª persona del singular supera al de la 1ª persona, como se muestra en la Tabla 41.

Ciudad	1ª singular		2ª singular	
	Peso	%	Peso	%
Málaga 1995	0.70	33%	0.73	27%
Málaga 2015	0.85	41%	0.81	24%
Granada (Manjón- Cabeza Cruz et. al, 2016)	0.64	25%		
Valencia (Roselló Verdeguer, 2021)	0.64	30%		
Santiago Chile (Martínez Lara et. al, 2021)	0.81	29%	0.54	55%
Ciudad de México (Lastra & Martín Butragueño, 2015)	0.58	25%		
Cali (Orozco et. al, 2024)	0.58	33%	0.71	47%
Medellín (Orozco y Hurtado, 2021)	0.51	32%	0.52	33%
San Juan Puerto Rico (Claes, 2011)	0.64	45%		
Barranquilla (Pérez Córdoba y Gomes Camacho, 2019)	0.55	50%		
Valledupar (Pérez Córdoba y Gomes Camacho, 2019)	0.59	50%		
Cartagena (Pérez Córdoba y Gomes Camacho, 2019)	0.51	46%		

Tabla 41. Pesos probabilísticos y porcentaje de presencia de la 1ª y 2ª persona del singular

Por otro lado, la variable que ocupa la segunda posición en la mayoría de los trabajos es la correferencialidad, a excepción de Orozco et al. (2024) y Martínez Lara et al., (2021) que no incluyen esta variable en sus análisis, el resto de trabajos coincide al obtener pesos probabilísticos más elevados en el factor no correferencial: Málaga 1995 (0.64), Málaga 2015 (0.57), Granada (0.33), Valencia (0.64), Ciudad de México (0.70), Medellín (0.62), San Juan de Puerto Rico (0.68), Valledupar (0.66) y Cartagena (0.68). Asimismo, la correferencialidad parcial favorece la aparición en las ciudades de Málaga 2015 (0.57), Valencia (0.77), México (0.54) y Medellín (0.56) y también destaca el factor no aplicable en Málaga 1995 (0.52), Valledupar (0.71), Cartagena (0.58). Por lo tanto, estos resultados sugieren que la no correferencialidad promueve la presencia en todos los corpus, mientras que en algunos inciden también la correferencialidad parcial con objeto y el no aplicable.

Otra variable pragmático-discursiva que también ha resultado significativa en cuatro trabajos previos es el tipo de discurso, como se detalla en el Gráfico 15. Al examinar el factor con más peso, podemos comprobar que la argumentación facilita la presencia del

sujeto pronominal, con valores elevados en Málaga en 1995 (0.56), Valencia (0.60), Ciudad de México (0.66) y Medellín (0.59).

Respecto a las variables que afectan a la conjugación de las formas verbales, en todas las ciudades han resultado significativas aquellas relacionadas con el tiempo y/o el modo, a excepción de Santiago de Chile, que no las incluye en sus análisis. Los resultados parecen mostrar una continuidad, ya que el presente ha sido un factor influyente en Granada (0.443), Ciudad de México (0.50), Barranquilla (0.55), Cartagena (0.53), Valledupar (0.52), Cali (0.50), Medellín (0.50); el imperfecto en Valencia (0.68), Valledupar (0.60), San Juan de Puerto Rico (0.64), Cali (0.55) y Medellín (0.61) y el pretérito en las ciudades de Valencia (0.51) y Ciudad de México (0.60). El modo ha sido estudiado con diferentes enfoques, ya que en algunos trabajos se integran con los tiempos verbales y en otros se han explorado de manera individual. Las ciudades que lo han abordado individualmente presentan coherencia en sus resultados al considerar que el indicativo posee un peso más elevado en Málaga en 1995 (0.74), Málaga 2015 (0.67), México (0.52), Cartagena (0.51) y Valledupar (0.51), y, además, en los dos cortes temporales de la ciudad de Málaga también lo hace el subjuntivo (0.58 en el corpus de 1995 y 0.68 en el corpus de 2015).

En cuanto a la clase semántica del verbo, esta no ha sido analizada en el trabajo de Pérez Córdoba y Gomes Camacho (2019), así como no ha resultado ser condicionante para la presencia del sujeto en Ciudad de México ni en Málaga en el corpus de 1995. Sin embargo, a pesar de que no se han usado las mismas clasificaciones para estudiar la clase semántica del verbo, aquellos relacionados con la cognición o los procesos mentales son los que suscitan un mayor peso probabilístico, lo cual se observa en los resultados de Málaga en 2015 (0.54), Valencia (0.57), Granada (0.39), Santiago de Chile (0.63), San Juan de Puerto Rico (0.60), Cali (sin el peso total debido al desglose por verbos específicos) y Medellín (0.57).

Además, la comparativa de los trabajos en el Gráfico 15 revela que no todas las variables están presentes en todas las ciudades analizadas, ya sea porque no han resultado ser significativas o porque no han sido incluidas en los estudios. Por lo tanto, las comparaciones con otras variables como la especificidad, la ambigüedad morfológica, la perífrasis, el tipo de cláusula y el tipo de discurso deben establecerse de manera parcial. En primer lugar, la especificidad solo se ha considerado en cuatro ciudades, pero en todas

ellas los referentes específicos promueven una mayor manifestación del pronombre. Este efecto se ha observado en Málaga 1995 (0.65) y 2015 (0.71), Valencia (0.53) y Granada (0.42), mientras que en la Ciudad de México la especificidad no logra ser significativa.

La ambigüedad morfológica se ha abordado en las ciudades de Málaga 1995 (0.57), Granada (0.36) y Ciudad de México (0.58), los datos parecen revelar un incremento de la expresión del sujeto con formas ambiguas. La perífrasis verbal solo ha sido estudiada en España, y los resultados indican una mayor explicitación del sujeto a formas no perifrásticas: Málaga (0.60), Valencia (0.51) y Granada (0.41). Por último, el tipo de cláusula ha sido examinado únicamente en Málaga y en Ciudad de México, donde se observa una concordancia al considerar que las oraciones afirmativas favorecen la presencia, con unos pesos probabilísticos positivos (0.62 y 0.54, respectivamente).

Las variables sociales solo han logrado ser relevantes en algunas ciudades, aunque no presentan una tendencia uniforme. Por un lado, la edad, especialmente los informantes del 3ª grupo generacional favorecen la presencia del pronombre en Medellín (0.58), Cali (0.55) y Santiago de Chile (0.50), y también los del 2ª grupo generacional en San Juan de Puerto Rico (0.59), Cali (0.52) y Santiago de Chile (0.52). Por otro lado, el nivel educativo del informante condiciona la explicitación pronominal del sujeto: en Cali (0.53) el nivel bajo propicia una mayor aparición, en Valencia (0.54) y Granada (0.46) es el nivel medio el que promueve la expresión del sujeto. Por último, la variable del sexo del entrevistado solo ha resultado ser significativa en dos ciudades, Santiago de Chile (0.53) y Cali (0.53), y ambas coinciden al ser las mujeres las que presentan un mayor peso probabilístico.

2.4.2.1. El patrón de uso del pronombre personal sujeto

A partir las comparaciones expuestas, trazamos el posible patrón de uso del pronombre personal sujeto en el Gráfico 16. No todos los factores tienen la misma importancia ni afectan en todas las ciudades analizadas, de modo que distinguimos entre unas reglas generales compartidas por la gran mayoría de los trabajos aquí abordados (presentadas en **negrita**) y unas reglas parciales, que requieren una mayor evidencia empírica para confirmar su vigencia en todas áreas hispanohablantes.

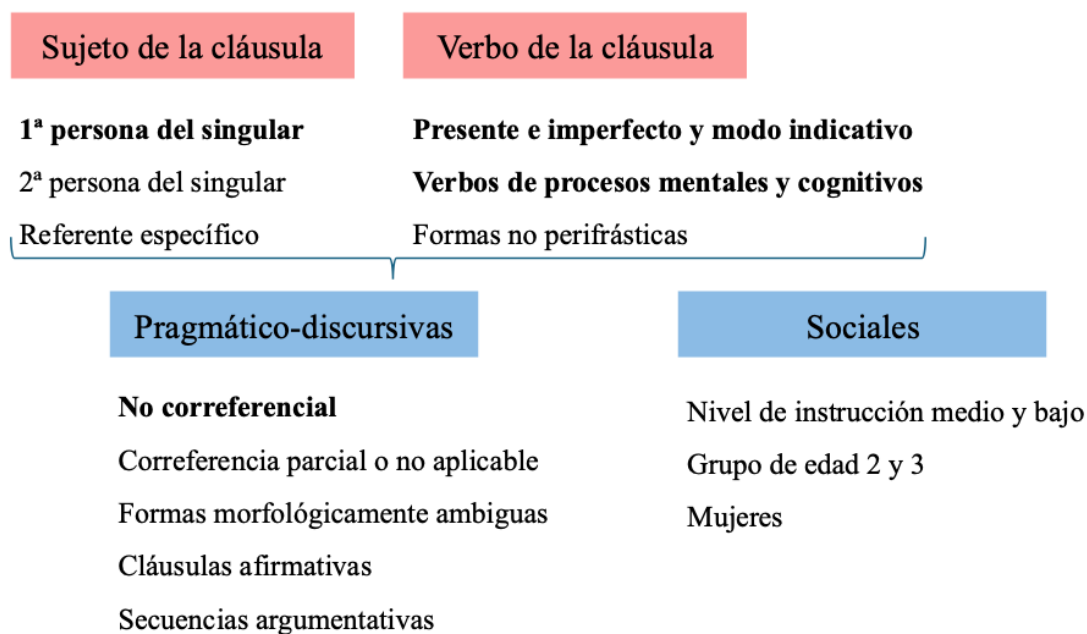


Gráfico 16. Patrón de uso del sujeto pronominal

De este modo, los factores que parecen incidir en mayor medida en la presencia del pronombre personal sujeto son persona del sujeto, correferencialidad, tiempo, modo y clase semántica del verbo. La explicitación del pronombre tiende a ser más frecuente cuando:

- El sujeto es 1ª persona del singular, y en algunos casos, también 2ª persona del singular.
- En contextos donde no hay correferencialidad el pronombre sujeto suele aparecer con mayor frecuencia y, aunque en menor medida, en situaciones donde la correferencialidad no es aplicable o es de tipo parcial.
- Los tiempos verbales presente e imperfecto, así como el modo indicativo.
- Los verbos cuya clase semántica denota procesos mentales o cognitivos.

Además, se detectan coincidencias que deben ser interpretadas de manera parcial, ya que no se dan en todos los trabajos, bien porque no se han considerado en el estudio o bien porque no han resultado significativas en los análisis estadísticos. Por lo tanto, estas reglas parciales promueven una mayor explicitación del sujeto pronominal en los siguientes contextos: cuando el referente es específico; cuando el sujeto acompaña a formas morfológicamente ambiguas; ante formas verbales no perifrásticas; en los casos con cláusulas afirmativas; y en las secuencias discursivas de carácter argumentativo. En

cuanto a las variables sociales, los informantes con niveles de instrucción medio y bajo, las generaciones 2ª y 3ª, y las mujeres tienden a explicitar más el sujeto pronominal.

A pesar de que estas comparaciones deban interpretarse con cautela —dado que los procedimientos metodológicos no son completamente idénticos y existen variaciones en los modelos estadísticos según la consideración (o no) de los efectos aleatorios, entre otros aspectos—, el patrón de uso del sujeto revela tendencias transversales a las diferentes zonas de habla hispana. De esta manera, se puede comprobar que, pese a que las tasas de explicitación sean diferentes según la zona geográfica, operan mecanismos lingüísticos subyacentes que determinan la aparición del sujeto pronominal en español.

2.5. Discusión y conclusiones parciales

Una vez establecidas las comparaciones diacrónicas y diatópicas dentro de los corpus PRESEEA y perfilado un posible patrón de uso, convendría adoptar una visión más panorámica para identificar los aspectos generales que condicionan sustancialmente la variación de la presencia y ausencia del sujeto pronominal en el discurso. En términos generales, esta variación sintáctica podría explicarse a partir de dos condicionantes: la percepción del individuo como agente articulador del discurso y la dimensión pragmático-discursiva del contexto comunicativo.

Numerosas investigaciones en distintas áreas han documentado un mayor protagonismo de la 1ª persona del singular —y, en ocasiones, también de la 2ª— en comparación con el resto de los pronombres del paradigma. Entre las explicaciones sobre esa prevalencia encontramos la propuesta por Givón (1986), fundamentada en la jerarquía de las implicaturas, así como la teoría sobre el efecto del punto de vista planteada por van Hoek, (2007). Ambas perspectivas convergen con los postulados de la sociolingüística cognitiva al considerar la influencia del individuo sobre el discurso y, a su vez, al situar el análisis del discurso dentro su contexto comunicativo inmediato.

Givón (1986) establece a partir de diferentes criterios —el rasgo humano, la especificidad, la implicación del participante en el discurso y la persona (1ª, 2ª y 3ª)—, la jerarquización de las implicaturas. Al aplicar estos criterios de jerarquización al pronombre personal de 1ª persona del singular, se observa que este reúne en grado máximo todas las

características mencionadas: posee necesariamente el rasgo de animación humano y de especificidad, corresponde a la 1ª persona gramatical y manifiesta una alta implicación discursiva al ser el propio emisor. Por su parte, el pronombre personal de 2ª persona en función de sujeto presenta unas características similares al de 1ª persona, ya que ambos comparten el rasgo humano y, en algunas ocasiones, también el de especificidad; aunque su grado de implicación en el discurso, al referirse al interlocutor, es inferior al de primera persona y superior al de 3ª persona.

En consonancia con lo anterior, van Hoek (2007, p. 911-914) subraya la relevancia del efecto el punto de vista sobre la organización del discurso, donde los pronombres *yo* y *tú* acaparan la atención de manera predominante al actuar como participantes del propio discurso. Además, pone en valor el doble rol que juega el pronombre *yo*, ya que actúa simultáneamente como observador y como objeto observado, un papel que también puede atribuirse a la 2ª persona del singular cuando esta posee un matiz inespecífico y se refiere al propio hablante de manera indirecta.

Desde esta perspectiva, podríamos considerar que la 1ª persona del singular en forma de sujeto, al remitir al propio hablante, es el argumento presupuesto con mayor fuerza dentro del paradigma de persona gramatical. El reflejo del carácter egocéntrico del hablante en el discurso también podría trasladarse a la prevalencia de construcciones en las que el sujeto aparece acompañado de verbos de entendimiento, creencia y opinión, puesto que es el propio hablante quien experimenta la acción cognitivamente y se sitúa así mismo como eje central alrededor del cual se articula el discurso.

En consecuencia, el alto grado de implicación y la posición central del hablante en el discurso constituyen algunas de las posibles causas detrás del elevado uso de los pronombres de 1ª y 2ª persona del singular documentado en los trabajos previos. Resulta conveniente destacar que los estudios de corte variacionista analizados se basan en entrevistas semidirigidas. De este modo, la naturaleza de este tipo de entrevistas podría acrecentar la aparición de estos pronombres debido a la interacción predominante del entrevistador y del informante.

Otro factor considerablemente influyente en el estudio de la variación de la ausencia y presencia del sujeto pronominal es el contexto discursivo. La dimensión pragmático-discursiva es fundamental para examinar minuciosamente la cláusula analizada, pero no debe descuidarse el papel que esta desempeña en el contexto

comunicativo previo. Como ya señalaba Silva-Corvalán (1982), la integración en el análisis de elementos discursivos próximos a la cláusula analizada es esencial para analizar, especialmente, las situaciones donde pueda ocasionarse ambigüedad por el efecto de competición de varios referentes, como ocurre con formas del imperfecto donde una misma forma verbal puede referirse indistintamente a la 1ª y 3ª persona del singular. Por ello, los análisis deberían considerar la presencia de elementos en el discurso que resuelvan dicha ambigüedad como, por ejemplo, los pronombres átonos, la referencia humana o la continuidad referencial. De lo contrario, estudiar estos casos –los verbos morfológica o fonéticamente ambiguos– de manera aislada sin atender al contexto puede conducir a resultados superficiales que analicen meramente la naturaleza de la forma verbal.

El contexto suele ofrecer tantas marcas de redundancia que, a pesar de la omisión del sujeto, la interpretación discursiva se puede mantener con facilidad. No obstante, otros factores como las intervenciones del entrevistador o la distancia con la que se ha mencionado previamente al referente en el discurso pueden llevar al hablante a explicitar más el pronombre. En la mayoría de los casos, la predictibilidad del sistema lingüístico y las estrategias que el interlocutor tiene a su alcance para continuar con el mensaje, como el conocimiento compartido por ambos (Lastra & Martín Butragueño, 2015) o la retención cognitiva del referente en cuestión, podrían ser suficientes para garantizar la coherencia y continuidad del mensaje.

Dada la revisión teórica y metodológica, recuperamos las hipótesis específicas planteadas al inicio de este capítulo para contrastarlas con los resultados obtenidos. Comprobamos a continuación la hipótesis según estén relacionadas con el sujeto o el verbo de la oración, con los factores pragmático-discursivo o los sociales.

En primer lugar, los resultados parecen respaldar la *hipótesis del egocentrismo directo a indirecto* del hablante ha sido verificada (Tabla 3 y Tabla 4), ya que la 1ª persona del singular es la que mayor aparición y presencia del sujeto pronominal alcanza, lo cual refleja que el hablante se posiciona en el centro de la situación comunicativa. El carácter egocéntrico se refleja a su vez en el empleo de otros recursos indirectos como el uso de *uno/a* o el *tú* inespecífico, que en ciertas ocasiones funciona como un *yo* encubierto. Esta hipótesis también se ha podido comprobar en el patrón lingüístico trazado, donde la 1ª persona es una gran impulsora de la aparición del sujeto.

Además, la *hipótesis de la especificidad de referente* también parece comprobarse, puesto que se esperaba que en oraciones donde el referente fuese inespecífico hubiera una menor presencia del sujeto. Además, se ha demostrado en la Tabla 6 y la Tabla 7 que la explicitación del sujeto comprometería el carácter inespecífico del referente en los casos de 3ª persona del plural, por lo que las variables de persona del sujeto y especificidad están estrechamente vinculadas.

En cuanto a las hipótesis relacionadas con las variables pragmático-discursivas, se ha demostrado que la pérdida del referente puede verse afectada por varios factores: el cambio de referente respecto a la oración anterior (Tabla 8 y Tabla 9), las situaciones de ambigüedad en las que existen varias entidades posibles para un mismo sujeto (Tabla 14 y Tabla 17) y la interrupción del entrevistador (Tabla 10). Además, los contextos donde confluyen el cambio en el turno conversacional y también se produce un cambio de referente son más proclives a la explicitación del sujeto (Gráfico 2). De manera que se corrobora la *hipótesis de la pérdida referencial contextual*.

Asimismo, la *hipótesis pragmático-discursiva* (Tabla 11) también se ha cumplido, ya que las secuencias argumentativas, tanto en Málaga como en el resto de las ciudades estudiadas, tienden a una mayor explicitación del sujeto. Se anticipaba que la correferencialidad tuviera una mayor influencia que el turno de habla y el tipo de discurso, lo que ha podido comprobarse (Gráfico 4), ya que el turno de habla no ha sido incluido en el modelo estadístico como una variable significativa y el tipo de discurso solo lo resultó relevante en el corpus de 1995. Además, la interacción de estas tres variables ha evidenciado que el contexto comunicativo marcado por el cambio de referente y las interrupciones del entrevistador no afecta de igual manera a todos los tipos de discurso.

La recodificación de algunas variables, como la ambigüedad morfológica (Tabla 14 y Tabla 15) y fonética (Tabla 18 y Tabla 19), ha contribuido a poner un énfasis particular en el contexto discursivo, donde se encuentran marcas que ejercen un papel desambiguador ante las formas verbales idénticas. En consecuencia, en aquellos casos donde la ambigüedad no puede ser resuelta mediante el contexto, la explicitación del pronombre es más necesaria y se corrobora así la *hipótesis de la ambigüedad descontextualizada*.

En cuanto a las variables que hacen referencia a los verbos de la oración, la *hipótesis de la semántica verbal* anticipaba que los verbos de procesos mentales, percepción sensorial, opinión y entendimiento motivarían una tasa pronominal más

elevada. Esta hipótesis se ha comprobado no solo en la ciudad de Málaga, como muestran la Tabla 20 y la Tabla 21, sino que también se integra en el patrón general de uso del sujeto.

El estudio del pronombre personal sujeto en la ciudad de Málaga se ha demostrado que, a pesar de ser una variación compleja motivada por factores de diversa naturaleza, existe cierta continuidad entre lo encontrado en la ciudad de Málaga y el resto de las ciudades de habla hispana. Si bien hemos podido corroborar nuestras hipótesis de partida, a partir de ahora, surgen otras preguntas como, por ejemplo, si estamos ante un cambio lingüístico impulsado por las mujeres que tiende hacia un incremento de la presencia del pronombre; o si del mismo modo que se ha producido con la gramaticalización del verbo *creer* en 1ª persona del singular, ocurrirá con otros verbos.



UNIVERSIDAD
DE MÁLAGA

CAPÍTULO 3. LA PERCEPCIÓN DE LA AMBIGÜEDAD MORFOLÓGICA Y LA CORREFERENCIALIDAD. ANÁLISIS EXPERIMENTAL MEDIANTE *EYE-TRACKING*

3.1. Introducción al análisis experimental mediante *eye-tracking*

La combinación metodológica del análisis de corpus y la experimentación dota a la investigación de pruebas complementarias que aportan una mayor solidez y fiabilidad a los datos al tratar un fenómeno desde distintas perspectivas (Recio Fernández et al., 2021, p. 1007). Los estudios de corpus se centran en mostrar qué relación existe entre los diversos factores lingüísticos y sociales, cuál es la frecuencia de uso de la variable analizada y qué función realiza esta dentro del discurso. En cambio, en la experimentación, el discurso se aborda como una actividad, de la cual se extraen datos sobre las reacciones (por ejemplo, el esfuerzo cognitivo medido en milisegundos) ante ciertos estímulos (fragmentos de texto o ciertas estructuras gramaticales). La función de estos trabajos es clarificar las posibles correlaciones entre las características específicas de las expresiones que funcionan a nivel del discurso y los patrones de procesamiento de los usuarios de la lengua (Loureda Lamas et al., 2019, p. 2-3).

En nuestro trabajo, tras haber examinado los datos de producción oral dentro de los corpus PRESEEA, hemos observado que entre los factores lingüísticos que condicionan la presencia del pronombre, las variables pragmático-discursivas ejercen una influencia significativa en la explicitación del sujeto pronominal. Con el objetivo de complementar el estudio de estos usos desde la óptica de la comprensión discursiva, analizamos el impacto cognitivo que el procesamiento de la variación pronominal puede tener en estos contextos. De esta forma, pretendemos evaluar en particular qué función cumple el sujeto pronominal durante el procesamiento de determinadas estructuras sintácticas. En concreto, nos interesa determinar si, en situaciones de ambigüedad morfológica y/o contextual, el pronombre explícito actúa como desambiguador con el fin de facilitar la interpretación o si, por el contrario, la redundancia pronominal incrementa los costes del procesamiento lingüístico.

Con el objetivo de llevar a cabo este análisis complementario, recurrimos al método del *self-paced reading* mediante *eye-tracking*. Este enfoque proporciona evidencia experimental sobre el proceso cognitivo que realizan los informantes al leer ciertas estructuras gramaticales y cómo su comprensión se refleja en los tiempos de lectura y en las regresiones. De este modo, el presente análisis experimental pretende correlacionar las variables lingüísticas —más concretamente aquellas relacionadas con el mantenimiento o cambio de referente con verbos morfológicamente ambiguos— con los parámetros de los tiempos de lectura, cuyo fin es identificar patrones sobre los esfuerzos cognitivos que desencadenan las estructuras sintácticas analizadas.

El análisis experimental propuesto parte de las siguientes hipótesis específicas, orientadas a evaluar cómo los diferentes factores lingüísticos influyen en los patrones de procesamiento cognitivo implicados en la percepción de las distintas estructuras.

1. *Hipótesis de la penalización por repetición.* Se anticipa una penalización en los tiempos de lectura ante aquellos estímulos donde el sujeto se repita de manera expresa en ambas cláusulas indicando un mismo referente, lo que producirá una sobrecarga informativa (Gernsbacher, 1989).
2. *Hipótesis de la ventaja.* Los enunciados en los que se explicita el sujeto pronominal en la primera cláusula y se omite en la segunda presentarán menores tiempos de lectura y/o menos regresiones, lo que podría indicar que la presencia

parcial del sujeto facilitaría su procesamiento al entenderse el enunciado completo como continuidad referencial (Ariel, 2001).

3. *Hipótesis del cambio referencial*. Se espera un mayor coste de procesamiento en los enunciados donde se produce un cambio de referente, ya que la introducción de un nuevo sujeto implicaría una mayor carga cognitiva (Hyönä, 1995; Khan, 1999).
4. *Hipótesis de la ambigüedad*. En las oraciones que sean ambiguas debido a la ausencia del pronombre personal sujeto, el informante necesitará mayor tiempo para procesar la oración y resolver la ambigüedad. Este incremento en el tiempo de lectura puede deberse al efecto de competición o efecto de múltiples candidatos (Gernsbacher, 1989; Badecker & Straub, 2002).

Antes de proceder a detallar el diseño experimental y los resultados obtenidos, consideramos necesario contextualizar brevemente las teorías relacionadas con los procesos de comprensión y procesamiento del discurso. Además, resulta necesario detallar cómo la herramienta de *eye-tracking* permite evaluar estos procesamientos cognitivos mediante los patrones de fijación durante la lectura.

3.2. Comprensión y procesamiento del discurso

La comunicación humana se considera un proceso cognitivo donde se combinan la ostensión y la inferencia; es decir, los interlocutores (re)crean una representación mental de lo que se ha dicho, mediante la decodificación del significado del material lingüístico emitido, y recurren a la inferencia necesaria para conectarlo con el contexto (Sperber & Wilson, 1995). Además, el ser humano posee una capacidad intrínseca para procesar cualquier *input* (una percepción visual, un sonido, un enunciado, un recuerdo) relevante de manera eficiente durante la comprensión del discurso, bajo la premisa de la eficiencia cognitiva. Esta eficiencia se basa en gran medida en el aprovechamiento de nuestros recursos de procesamiento con la finalidad de maximizar el rendimiento cognitivo y, además, debe producirse un equilibrio entre el esfuerzo invertido y su grado de consecución (Wilson & Sperber, 2004, 2012).

En este contexto, la Teoría de la Relevancia de Sperber y Wilson (1995) proporciona dos principios claves para explicar esta capacidad humana. El principio

cognitivo afirma que la cognición humana tiende a usar los dispositivos de procesamiento, de la manera más eficaz posible, para escoger de modo automático los estímulos potencialmente más relevantes. En cuanto al principio comunicativo, este considera que cada acto de comunicación ostensiva tiene su propia presunción de relevancia óptima en la que el receptor solo presta atención a un estímulo considerado relevante y el emisor, por su parte, anima al receptor a sospechar que el estímulo es tan relevante que merece la pena procesarlo (Wilson & Sperber, 2004, p. 254). En consecuencia, esta teoría sostiene que, durante el proceso de la comprensión, el oyente infiere el significado oracional codificado a partir de la forma lingüística y, siguiendo la ley del mínimo esfuerzo, calcula los posibles efectos cognitivos —resolución de la ambigüedad, identificación del referente, etc.— por orden de accesibilidad y se detiene cuando cumple sus expectativas de relevancia.

Dentro del marco de la comunicación ostensiva, se distinguen dos tipos de intenciones: la intención informativa y la intención comunicativa. La primera se refiere al deseo del emisor de modificar la representación mental del oyente y pone de manifiesto una serie de supuestos subyacentes. Por otro lado, la intención comunicativa pretende que el receptor reconozca, a través del estímulo producido, la intención informativa (Sperber & Wilson, 1995; Grice, 1975). Este doble nivel de la intencionalidad es esencial para la interpretación adecuada del mensaje, puesto que permite al destinatario recibir la información relevante y comprender el propósito del hablante al comunicarla.

Otro factor importante dentro de la comunicación es el contexto concebido como un elemento cognitivo y dinámico compuesto por información extraída de enunciados anteriores, del entorno o de cualquier suposición mantenida en la memoria a corto o largo plazo. En cuanto a su denominación, se prefiere el uso del término *supuestos probabilísticos mutuos*, en lugar de *conocimiento mutuo* para hacer referencia al entorno cognitivo compartido. La información compartida no está determinada *a priori*, sino que, durante el proceso de interpretación del enunciado, se va seleccionando progresivamente, ya que la relevancia está relacionada con la noción de efecto contextual (Moeschler, 1989, p. 329). En esta misma línea, Recio Fernández (2019, p. 11-12) subraya la importancia de los efectos contextuales desencadenados durante el proceso de inferencia, los cuales provocan cambios en el entorno cognitivo mutuo de los interlocutores. Estos pueden

clasificarse en función de tres posibles tipologías recogidas por Sperber y Wilson (1995, p. 107-108) y por Blakemore (2002, p. 61):

- La información nueva se contextualiza mediante información antigua ya almacenada en la mente. En este caso, los dos tipos de información se utilizan como premisas para llegar a una conclusión que no podría haber sido inferida recurriendo únicamente a la información nueva o a la antigua, y se derivan sintéticamente implicaciones contextuales o nuevos supuestos.
- La información nueva presenta un efecto reforzador del grado de mantenimiento una suposición ya almacenada.
- Por último, la nueva información conduce al abandono o a la eliminación de una suposición sostenida.

La relevancia no es el único principio presente en el discurso: la coherencia también ejerce un papel fundamental en la estructuración y comprensión de los enunciados (Recio, 2019, p. 14; Giora, 1997). Para Almor y Nair (2007), la coherencia discursiva implica la repetición de elementos mencionados previamente y esas referencias pueden hacerse mediante diferentes recursos como el empleo de un nombre propio, una descripción o un pronombre. Esto también podría estar asociado al grado de accesibilidad del referente en la memoria (Ariel, 1990) o a su estado cognitivo (Gundel et al., 1993).

Por un lado, la Teoría de la Accesibilidad propuesta por Ariel (1990, 2001) relaciona el lenguaje humano, más concretamente el sistema referencial, con la activación de nodos en la memoria humana. Estos nodos no se accionan todos por igual, ya que algunos presentan activaciones más rápidas o lentas. Con la intención de ordenar los mecanismos, esta Teoría ofrece una jerarquización de las expresiones referenciales ordenadas de menor a mayor grado de accesibilidad. En esta clasificación, los pronombres personales, las flexiones verbales y el cero ocupan una posición de máxima accesibilidad (Ariel, 1990, p. 73). La accesibilidad del referente también estaría ligada de manera intrínseca a la situación contextual, la cual se vería afectada por cuatro factores: la distancia que separa al referente, la competencia con otros posibles referentes potenciales, el grado de prominencia y el tamaño de la unidad en la que se integra la mención que puede ser más pequeña, como en la cláusula anterior, o en una unidad más grande como un párrafo.

Además, la Teoría de la Accesibilidad (Ariel, 2001) considera que los marcadores de accesibilidad menos informativos, menos rígidos y más atenuados tienden a recuperar representaciones mentales altamente accesibles. En cambio, los marcadores más informativos, más rígidos y menos atenuados se asocian con representaciones de baja accesibilidad. Otra de las predicciones de esta Teoría es que los marcadores de alta accesibilidad deben ser usados cuando la conexión entre el antecedente y la anáfora es estrecha, mientras que en conexiones más laxas es más apropiado el uso de marcadores de baja accesibilidad. Para corroborar esta hipótesis, Ariel (2001) cita el trabajo de Khan (1999) donde se evidencia que el uso del pronombre de sujeto opcional se emplea para marcar una separación del contexto anterior por algún tipo de discontinuidad o disyunción. Esta tendencia es particularmente notable cuando se produce un cambio de sujeto, se modifica el contexto, los eventos se perciben como separados o se inicia una intervención discursiva.

Por otro lado, las expresiones referenciales están condicionadas por su estado cognitivo, de forma que estas podrían depender de las suposiciones que un hablante haga según su conocimiento y el estado de atención en el contexto en el que se utiliza la expresión (Gundel et al., 1993, p. 281). Además, Gundel et al. (1993, p. 283) proponen *The Universality Givenness Hierarchy*, que organiza en seis estados cognitivos las formas de referencia en el discurso del lenguaje natural de la más a la menos restrictiva: (1) foco > (2) activado > (3) familiar > (4) únicamente identificable > (5) referencial > (6) identificable por tipo.

Al aplicar esta jerarquización al español, el estado de (1) *foco* correspondería a la presencia o ausencia del pronombre utilizada cuando el referente no solo está en la memoria a corto plazo, sino que también se encuentra en el centro de la atención actual (*Nosotros/ Ø viajamos mucho*); en nivel (2) *activado* se ubicaría el pronombre personal carácter enfático y los demostrativos (*ELLA viaja mucho, no él o Ese chico viaja mucho*); en (3) *familiar* estarían los referentes que pueden ser identificados de manera inequívoca porque ya se encuentra en la memoria una representación de ellos (*-¿Te acuerdas de Juan? -Sí, ese hombre es muy inteligente*); en (4) *únicamente identificable*, el referente puede ser identificado con el contenido nominal (*El niño ya va al colegio*); y, por último, en español se unifican en una única categoría el (5) *referencial* y el (6) *identificable por*

tipo, donde el referente puede ser identificado, pero no de manera única, sino que se refiere a un individuo específico dentro de esa categoría (*Un niño jugaba en el parque*).

Al centrarnos exclusivamente en los pronombres, estos son considerados mecanismos de continuidad y coherencia discursiva que deben ser interpretados en relación a la información presentada previamente en el discurso (Gordon et al., 1993. p. 313-314). Los pronombres, como centros de atención, pueden establecer diferentes tipos de transiciones entre oraciones compuestas por dos o más cláusulas. De este modo, Gordon et al. 1993 consideran que la continuación referencial es más coherente que la retención, y esta última, a su vez, es más coherente que el cambio referencia. Además, sostienen que estas preferencias deberían reflejarse en la carga de procesamiento que implica la comprensión de segmentos discursivos con diferentes tipos de transición.

3.3. Lectura con *eye-tracking*: de la fijación ocular a la comprensión

La experimentación resulta fundamental a la hora de observar cómo las características y los diferentes elementos de un discurso influyen en la recreación y comprensión de las representaciones mentales desencadenadas. En la actualidad, disponemos de técnicas de investigación experimentales capaces de medir la actividad mental clasificadas en dos grandes modalidades: los métodos *offline* y los *online* (Recio Fernández et al., 2021, p. 1008-1009). A continuación, nos centramos en el método *online*, ya que consideramos que es el más adecuado para la consecución de nuestro objetivo. Esta técnica posibilita la recogida de datos en el instante en que el informante percibe los estímulos y, de esta forma, se reduce la influencia que genera el conocimiento explícitamente aprendido y acorta el tiempo de resolución consciente de problemas lingüísticos (Keating & Jegerski, 2015, p. 2).

Las diferentes técnicas *online*, más concretamente las de seguimiento ocular, han experimentado un notable desarrollo en la investigación del lenguaje durante las últimas décadas. Los estudios de lectura con *eye-tracking* ofrecen datos sobre los patrones de procesamiento relacionados con la mirada y proporcionan información sobre cuánto tiempo necesita un informante en el procesamiento discursivo y cómo ciertas características discursivas requieren un mayor o menor esfuerzo cognitivo en la resolución de una tarea (Conklin et al., 2018; Rayner, 2009)

Estos estudios se sustentan principalmente en dos supuestos básicos: el supuesto de la inmediatez y el supuesto ojo-mente (Just & Carpenter, 1980, p. 330). El primero (*the immediacy assumption*) considera que el lector interpreta cada palabra y, a su vez, procesa los diferentes niveles del sistema a medida que aparecen en el texto, a pesar de que no posponerlos pueda llevar a conjeturas erróneas que demanden una aclaración más adelante. El segundo (*the eye-mind assumption*) postula que el ojo permanece fijo en una palabra mientras que está siendo procesada. De este modo, la duración de la fijación se correspondería con el tiempo que tarda en procesarse dicha palabra, sin desfase entre lo que se observa y lo que se procesa. Entonces, las palabras que requieren un mayor procesamiento generarían fijaciones más prolongadas.

Estos estudios experimentales conllevan un gran control metodológico por parte del investigador sobre las diferentes partes del experimento, desde la selección rigurosa de los participantes, la adecuación de la técnica experimental elegida o la casuística del diseño experimental hasta el desarrollo en directo de la prueba y el posterior tratamiento estadístico de los datos (Keating & Jegerski, 2015; Recio Fernández et al., 2021). No obstante, las investigaciones con seguimiento ocular describen con gran exactitud la naturaleza, las características y las funciones de las fijaciones oculares, refijaciones, movimientos sacádicos y de regresión durante la lectura. Además, se han definido parámetros sobre el campo visual humano y su relación con el proceso de lectura, así como varias métricas para analizar los fenómenos asociados a la lectura.

En primer lugar, detallamos los diferentes comportamientos oculares ocasionados durante la lectura (Figura 1), los cuales son indicadores del procesamiento cognitivo (Guillén Jiménez, 2023):

- Las fijaciones se definen como aquellos periodos en los que la retina se mantiene relativamente estable sobre un punto específico del estímulo, ya sea un texto o una imagen (Duchowski, 2007, p. 44). Los lectores habitualmente se fijan en el centro de las palabras durante su lectura, lo cual proporciona una posición óptima de visualización para la extracción de información (McConkie et al., 1988; O'Regan, 1990; O'Regan & Jacobs, 1992; O'Regan & Lévy-Schoen, 1983; Rayner, 1979).
- Las refijaciones se producen cuando los ojos regresan a un elemento visitado previamente dentro del estímulo. Este comportamiento es relativamente poco frecuente durante la lectura, ya que solo el 15% de las palabras en un texto son

releídas (Rayner, 1998, p. 387). Además, se considera que las refijaciones están motivadas por un procesamiento léxico incompleto (Pollatsek & Rayner, 1990) o por la necesidad de un análisis más profundo debido a la complejidad del contenido (Just & Carpenter, 1980, p. 337).

- Las sacadas son movimientos bruscos de los ojos entre fijaciones, voluntarios o reflejos, cuya función principal es trasladar la mirada de una zona de interés a otra para extraer información. Durante las sacadas no se obtiene nueva información, aunque sí es posible continuar procesando la información obtenida anteriormente (Rayner, 2009).
- Las regresiones son movimientos sacádicos hacia atrás que representan entre el 15-25% de los movimientos oculares realizados durante la lectura normal (Booth & Weger, 2013). Holmqvist et al. (2011) distinguen entre regresiones dentro de una palabra, asociadas a procesos de activación léxica, y regresiones entre palabras, que reflejan procesos de integración de la oración. En cuanto a las regresiones de mayor alcance, es decir, aquellas que van más allá de la palabra previa o que regresan incluso a segmentos anteriores del texto, su función aún no está completamente esclarecida (Conklin et al., 2018, p. 67). Sin embargo, todo apunta a que estas regresiones podrían ser indicativas de dificultades en procesamiento de la información (Rayner, 1998, p. 375).

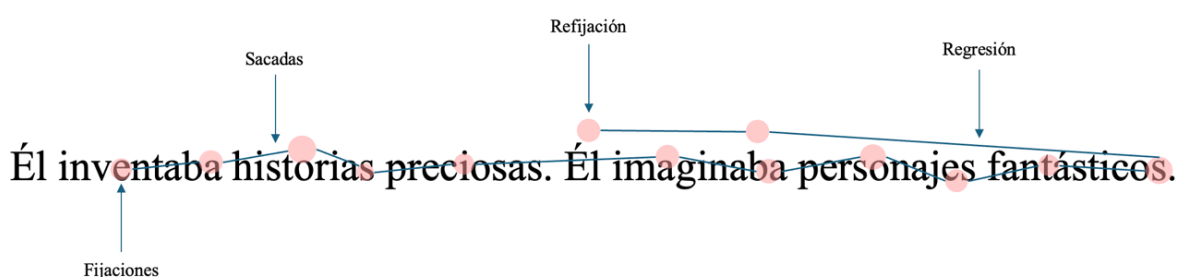


Figura 1. Tipos de movimientos oculares en lectura

Una vez identificados los comportamientos oculares, los investigadores han implementado diversas métricas por cuantificar los distintos comportamientos oculares, asociando cada uno a un tipo de procesamiento cognitivo determinado (Pulvermüller et al., 2009). Estos parámetros acumulados del seguimiento ocular permiten recopilar datos a partir de áreas de interés (AOI) dibujadas de acuerdo con los objetivos y el diseño del experimento (Holmqvist et al., 2011, p. 187).

Las medidas de fijación se pueden clasificar en dos tipos: medidas tempranas y medidas tardías (Conklin et al., 2018, p. 66). Las primeras reflejan un procesamiento automático, como el reconocimiento de palabras y el acceso léxico, mientras que las segundas están vinculadas a un procesamiento más controlado y consciente (Staub & Rayner, 2007). Estas métricas de fijación también se conocen como medidas de duración de posición. Para la investigación desarrollada en este capítulo solo resultan relevantes las siguientes tres medidas: el tiempo de primera pasada (en rojo) y el tiempo de relectura (en azul) representado en la Figura 2 y el tiempo total de lectura en la Figura 3.

- Tiempo de primera pasada (*First Reading Time*, FPT): Esta medida captura el proceso de construcción de una interpretación inicial a partir de un enunciado, donde se produce el procesamiento sintáctico, semántico y pragmático (Cruz & Loureda Lamas, 2019; Loureda Lamas et al., 2021; Nadal et al., 2016, 2017). Se define como la suma de todas las duraciones de fijación en un estímulo desde que la mirada entra en una AOI hasta que se dirige hacia la derecha (Holmquist et al., 2011; Rayner, 1998).

Tiempo de relectura (*Re-reading Time*, RRT): Esta métrica —conocida también como tiempo de segunda pasada o tiempo de fijación de segunda pasada— corresponde a la suma de todas las fijaciones en una AOI después de que esta haya sido visitada y abandonada previamente (Recio Fernández et al., 2018). El RRT refleja un procesamiento de orden superior que implica la confirmación, el enriquecimiento o la revisión de la información previamente procesada (Loureda Lamas et al., 2021; Recio Fernández, 2019, p.125).

Él inventaba historias preciosas. Él imaginaba personajes fantásticos.

Figura 2. Fijaciones para la primera lectura (FRT) y la relectura (RRT)

- Tiempo total de lectura (*Total Reading Time*, TRT): Esta métrica también se denomina tiempo total de fijación, duración de la mirada, tiempo acumulativo de fijación o tiempo total de observación, entre otros términos (Holmqvist et al., 2011, p. 389) y se refiere a la suma de todos los periodos en los que la mirada se fija en una AOI durante una prueba (Conklin et al., 2018, p. 67). El TRT

proporciona una visión general del procesamiento cognitivo de alto nivel y abarca tanto la recuperación inicial como el procesamiento posterior de palabras.



Figura 3. Fijaciones para la lectura total (TRT)

Además de los parámetros presentados en este capítulo, existen más variables referentes a las medidas de lectura, pero limitamos la explicación a las que se implementan en nuestro experimento.

3.4. Diseño experimental

3.4.1. Variables, materiales y áreas de interés

Este experimento pretende observar cómo la variación del sujeto pronominal influye en el procesamiento cognitivo de los estímulos, concretamente en las métricas de lectura seleccionadas como variables dependientes: *Total Reading Time*, *First Reading Time* y *Re-reading Time*. En cuanto a las variables independientes, contemplamos la variación ausencia-presencia del pronombre personal sujeto, la ambigüedad de ciertas estructuras sintácticas con verbos morfológicamente ambiguos, y la correferencialidad, entendida como continuidad o cambio referencial.

Con el propósito de variación la variación del sujeto pronominal, construimos cuatro condiciones a partir de la combinación de las tres variables lingüísticas mencionadas. Cada condición está formada por dos enunciados para poder establecer las relaciones de correferencialidad con la cláusula inmediatamente anterior. En la Tabla 42 se muestra la descripción y el esquema de cómo son las estructuras de cada ítem según la condición que corresponda.

Condición	Estructura	Pronombre	Ambigüedad	Referente
C1				
<i>Inventaba historias preciosas. Imaginaba personajes fantásticos.</i>	Ø..., Ø ...	Ausencia total	Ambiguo	Continuidad
C2				
<i>Él inventaba historias preciosas. Imaginaba personajes fantásticos.</i>	Él ..., Ø...	Presencia parcial	Ambiguo	Continuidad
C3				
<i>Él inventaba historias preciosas. Él imaginaba personajes fantásticos.</i>	Él ..., Él ...	Presencia total	No ambiguo	Continuidad
C4				
<i>Él inventaba historias preciosas. Yo imaginaba personajes fantásticos.</i>	Él..., Yo...	Presencia total	No ambiguo	Cambio

Tabla 42. Descripción de los ítems según la estructura y las variables lingüísticas

Para entender el diseño experimental es necesario aclarar algunos términos fundamentales empleados en el experimento. El *ítem crítico* es la unidad o la condición crítica que nos interesa analizar, en él se encuentran implícitamente las variables lingüísticas y sobre él se miden las variables dependientes relacionadas con los tiempos de lectura. Los ítems críticos se integran en las *historias* que suelen ser relatos sencillos sobre dos personajes humanos. Estas historias se distribuyen en *listas* que recogen qué condición y de qué historias proviene cada ítem crítico. Cada lista es leída por un determinado número de informantes.

En estos términos, nuestro diseño experimental está compuesto por cuatro condiciones (C1, C2, C3 y C4), donde cada ítem crítico está adaptado dentro de diferentes historias (historia 1 a historia 8). Todos los ítems críticos que forman parte de la misma historia son oraciones léxicamente iguales, las cuales solo se difieren las unas de las otras por la ausencia o presencia del sujeto pronominal en sus cláusulas. De esta manera, se garantiza la consistencia interna de los ítems y la comparabilidad del conjunto experimental.

Como muestra la Tabla 43, creamos cuatro listas experimentales (listas A, B, C y D), cuya distribución sigue el diseño del cuadro latino contrabalanceado. En este diseño, las ocho historias se distribuyen de manera homogénea en las cuatro listas, pero en cada lista se presenta una condición diferente de cada historia. Para cada lista, asignamos un mínimo de 20 participantes (un total de 80 participantes), quienes leen dos veces cada condición en historias diferentes. En los estímulos críticos, se controlan la frecuencia y la longitud de las palabras con el objetivo de evitar posibles variables de confusión (Rayner, 2009).

	Lista A	Lista B	Lista C	Lista D
Historia 1	C1	C2	C3	C4
Historia 2	C2	C3	C4	C1
Historia 3	C3	C4	C1	C2
Historia 4	C4	C1	C2	C3
Historia 5	C1	C2	C3	C4
Historia 6	C2	C3	C4	C1
Historia 7	C3	C4	C1	C2
Historia 8	C4	C1	C2	C3
N= 80	20	20	20	20

Tabla 43. Diseño experimental según cuadro latino

Las historias sirven como contextualización de los ítems críticos de manera lingüística y visual, antes y después del estímulo que pretendemos analizar. Antes de mostrar las historias que contienen los ítems críticos, se hace una introducción, donde se presenta el personaje principal. Cada historia está compuesta por varias pantallas, de modo que la Figura 4 representa cómo se vería un pasaje completo de una historia con el ítem crítico de la condición 2.

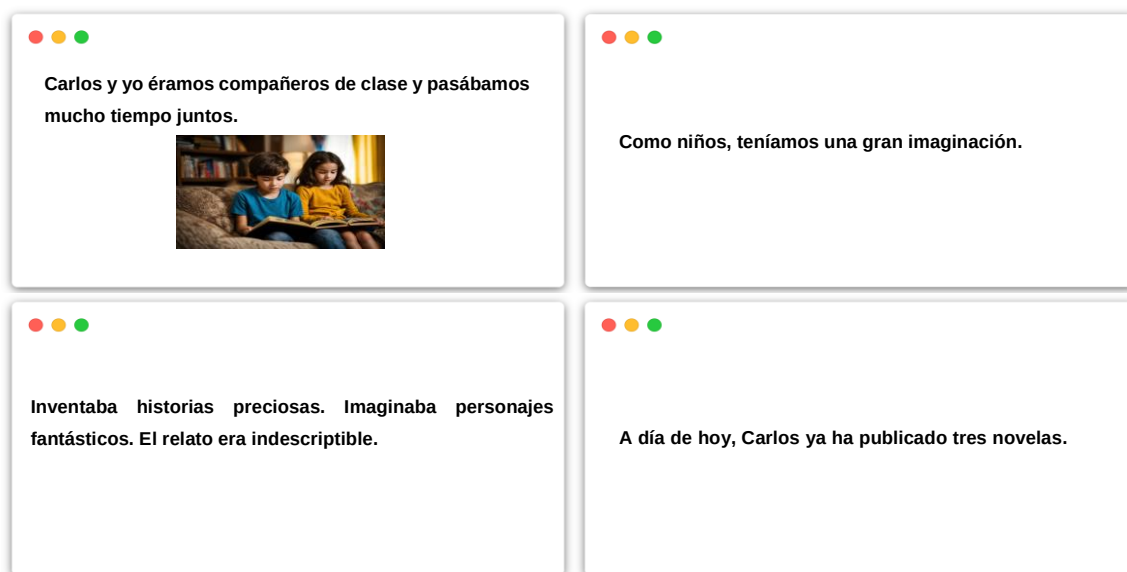


Figura 4. Secuencia completa de la historia 1 con la condición 1

En la primera pantalla se presenta un estímulo visual —una imagen creada con inteligencia artificial— donde aparecen las dos personas que protagonizan la historia, y una contextualización para presentar a los participantes y la situación en cuestión. En la siguiente pantalla, el primer relleno contiene un verbo en 1ª persona del plural. A continuación, se presenta el ítem crítico compuesto por dos cláusulas, donde se pueden establecer relaciones de correferencialidad, y una oración final encabezada por un sintagma nominal. Esta última oración no se considera en el análisis con el objetivo de minimizar el efecto *wrap-up*³. Por último, se agrega un relleno final que sirve para cerrar la historia.

Por lo tanto, los estímulos presentan una proporción de 2:1 —dos rellenos por un ítem crítico (Conklin et al., 2018)— y, además, cada lista está compuesta por ocho historias críticas (Anexo II) y tres distractoras, —en una proporción de 3 historias críticas por 1 historia distractora (Keating y Jegerski, 2015)—.

Para medir las variables dependientes referentes al tiempo de procesamiento, los ítems críticos se fragmentan en las llamadas *áreas de interés* (AOI) dibujadas según los

³ El efecto *wrap-up* hace referencia al fenómeno que ocurre al final de oración o de cláusula durante el cual tradicionalmente se ha considerado que se lleva a cabo el procesamiento integrativo destinado a relacionar de las oraciones y actualizar el modelo discursivo. (Just y Carpenter, 1980; Warren et al., 2009)

diferentes criterios lingüísticos dependiendo de lo que se pretenda analizar. Para este trabajo, consideramos relevante la división de las áreas de interés presentadas en las siguientes figuras: las áreas mayores (Figura 5) que comprenden la totalidad de los dos enunciados o los dos enunciados por separado, sin incluir la frase final (*El relato era indescriptible*), y las áreas más específicas (Figura 6) que inciden en los componentes de la oración como el verbo, el objeto o el sujeto.

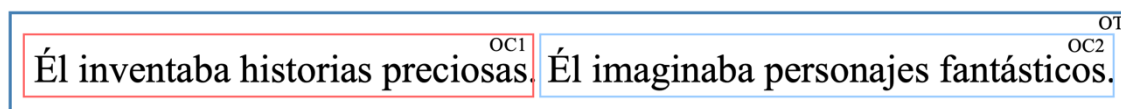


Figura 5. Áreas de interés mayores

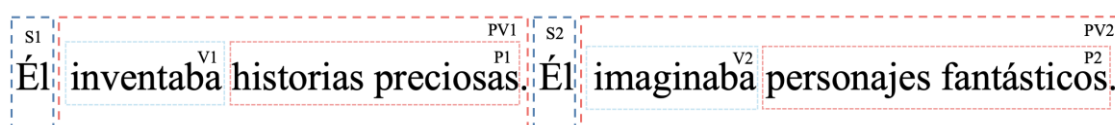


Figura 6. Áreas de interés menores

De este modo, cada área de interés recibe una nomenclatura que se emplea en los análisis posteriores para referirse a ciertas partes del estímulo, como muestra la Tabla 44.

AOI	Definición	Ejemplo
OT	Oración Total	Él inventaba historias preciosas. Él imaginaba personajes fantásticos.
OC1	Oración coordinada 1	Él inventaba historias preciosas.
OC2	Oración coordinada 2	Él imaginaba personajes fantásticos.
PV1	Verbo 1 y Objeto 1	inventaba historias preciosas.
PV2	Verbo 2 y Objeto 2	imaginaba personajes fantásticos
S1	Sujeto 1 de la oración 1	Él
S2	Sujeto 2 de la oración 2	Él
V1	Verbo 1	inventaba
V2	Verbo 2	imaginaba
P1	Objeto 1	historias preciosas
P2	Objeto 2	personajes fantásticos

Tabla 44. Descripción y nomenclatura de las áreas de interés seleccionadas

3.4.2. Participantes, procedimiento, equipamiento y tratamiento estadístico

Los participantes del experimento presentan un perfil homogéneo, descrito detalladamente en la Tabla 45. Todos ellos son hablantes nativos de español, con una formación académica similar y características visuales adecuadas para la realización de la tarea experimental. La muestra inicial recogida consta de 124 participantes, si bien, tras aplicar los criterios de exclusión de valores atípicos o tras detectar errores en la grabación, el análisis se realiza sobre un total de 99 participantes. A continuación, se resumen las principales características de la muestra.

Participantes	Descripción
Sexo	84% Mujeres y 14% hombres
Edad	18-40 años
Lengua	Español nativo
Formación	Estudiantes universitarios
Visión	Normal o corregida
Muestra recogida	124 participantes
Muestra empleada	99 participantes

Tabla 45. Descripción de la muestra analizada mediante *eye-tracking*

A la llegada al laboratorio, los participantes firmaban un consentimiento previo para autorizar el tratamiento de información personal y cumplimentaban un cuestionario sociológico de manera anónima. Antes de empezar, el investigador explicaba al participante las instrucciones del procedimiento experimental, sin revelar el objetivo específico del estudio. Además, se advertía a los participantes que debían leer en silencio el conjunto de oraciones que aparecerían en pantalla y que tendrían que permanecer quietos y en silencio durante el transcurso del experimento.

Tras recibir las instrucciones por parte del investigador, los participantes leían nuevamente las indicaciones directamente desde el equipo para que durante ese tiempo pudieran acomodarse a la posición requerida para la realización del experimento. Posteriormente, se procedía al ajuste del equipo mediante un sistema de calibración de nueve puntos y una validación posterior. Antes del inicio del experimento, se presentaba un estímulo de prueba con el fin de aclarar cualquier duda relativa al procedimiento. Durante el transcurso de la prueba, los participantes se sentaban a 65 cm del monitor con

la cabeza apoyada en la mentonera para minimizar los movimientos no deseados que pudieran influir en el desarrollo de la prueba. La duración aproximada del experimento oscilaba entre 10 y 12 minutos por individuo.

El equipo usado para la recogida el experimento es el *EyeLink 1000 Plus*, que permite una frecuencia de muestreo de hasta 2000 Hz, apoyado por el software especializado *SR Research* de *Experiment Builder* para la implementación del diseño experimental propuesto y la programación de la base datos que contiene los estímulos presentados. Tras la obtención de los datos, se revisan las grabaciones mediante el programa *Data Viewer*, integrado también en el software de *SR Research*. Este programa permite definir las áreas de interés, además visualizar, limpiar y extraer los datos finales con el propósito de analizarlos estadísticamente.

En cuanto al tratamiento estadístico de los datos obtenidos, los análisis se llevaron a cabo en colaboración con el laboratorio StabLab del Institut für Statistik der Ludwig-Maximilians-Universität München. Los modelos se calculan en R implementando la función *gam* del paquete *mgcv* para modelos mixtos aditivos generalizados (Wood, 2017). Debido al elevado número de pruebas de hipótesis en todos los modelos, los valores de *p* en cada modelo se corrigen utilizando el método de Holm (Holm, 1979). El modelo basado en efectos mixtos considera como efectos fijos las áreas de interés para cada condición y los tiempos acumulativos por palabra en cada AOI, y como efectos mixtos los participantes y los ítems críticos (Holmquist et al., 2011).

De todas las observaciones registradas durante la grabación, se excluyen los valores atípicos en cada área de interés según los siguientes criterios:

- Exclusión de valores vacíos: tiempos de lectura de la OT igual a 0 ms. En esta categoría se detectan 9 valores atípicos, que afectan el 1.8% del total de las observaciones.
- Exclusión de lecturas excesivamente rápidas: valores inferiores a 80 ms en la lectura y relectura de la OT. Este criterio excluye 10 valores atípicos, lo que supone el 2% de las observaciones.
- Exclusión de lecturas excesivamente lentas: tiempo de lectura superior a 800ms en la OT. Bajo este criterio se excluyen 78 valores atípicos, equivalentes al 16.7% del total.

Una vez aplicados estos criterios, se eliminan del análisis 396 observaciones, lo que supone un 18.4% del total de datos grabados.

El modelo estadístico resultante genera los valores estimados y las predicciones medias por letra, junto con las desviaciones estándar de los tiempos de lectura.

Diferencia de porcentajes en la media de los tiempos de lectura	Magnitud del efecto
$\leq 3.99\%$	Trivial
4.00% a 4.99%	Efecto pequeño
5.00% a 9.99%	Efecto mediano
10.00% a 19.99%	Efecto grande
$\geq 20\%$	Efecto muy grande

Tabla 46. Escala de magnitud de efectos

Para interpretar los resultados proporcionados por el modelo estadísticos, en el siguiente apartado aplicamos la escala de la magnitud de efectos (Loureda Lamas et al., 2021). Estas diferencias entre los tiempos de lectura se expresan en porcentajes, como muestran la Tabla 46.

3.5. Análisis

Con el propósito de corroborar las hipótesis planteadas previamente, analizamos los resultados referentes la resolución de la ambigüedad y, posteriormente, los vinculados al cambio referencial. Para poder percibir visualmente las diferencias en las comparativas realizadas, los gráficos presentan en el eje izquierdo el tiempo en milisegundos y en el eje derecho la magnitud porcentual de estas diferencias de manera normalizada, es decir, todas las representaciones mantienen el mismo rango proporcional. Los resultados presentados se extraen de un modelo estadístico más amplio, que se puede consultar en el Anexo III.

3.5.1. Ambigüedad referencial

En esta sección se muestran los resultados referidos a los tiempos de lectura de las condiciones C1, C2 y C3, dado que en estas se puede examinar si la ausencia o presencia

del sujeto ejerce algún efecto en la percepción cognitiva de la ambigüedad referencial. Por lo tanto, en este apartado comparamos las siguientes estructuras ejemplificadas a continuación:

C1: Inventaba historias preciosas. Imaginaba personajes fantásticos.

C2: Él inventaba historias preciosas. Imaginaba personajes fantásticos.

C3: Él inventaba historias preciosas. Él imaginaba personajes fantásticos.

Como puede apreciarse en estos ejemplos, la primera condición (C1) presenta una doble omisión del sujeto pronominal en ambas cláusulas; la segunda condición (C2) manifiesta una presencia parcial del sujeto, explícito en la primera cláusula y omitido en la segunda; finalmente, la tercera condición (C3) presenta la explicitación pronominal en las dos oraciones. Inicialmente, evaluamos los efectos cognitivos en el conjunto de los enunciados (OT), por ejemplo: *Inventaba historias preciosas. Imaginaba personajes fantásticos*. La Tabla 47 recoge los resultados obtenidos según el tiempo total de lectura (TRT) de los enunciados completos y toma como ítem de control la condición 1, que ocupa la primera posición en dicha comparativa.

OT	Comparación	Ítem 1 (control)	Ítem 2	Diferencia (ms)	Diferencia (%)
TRT	C1 vs C2	432.76	416.36	16.4	-3.78
	C1 vs C3	432.76	427.99	4.77	-1.10
	C2 vs C3	416.36	427.99	-11.63	2.79

Tabla 47. Comparativa de C1, C2 y C3 según tiempo de lectura TRT en la OT

A pesar de que las condiciones analizadas revelan diferencias en los tiempos totales de lectura, estas no parecen ser muy relevantes, ya que sus diferencias porcentuales no superan el 4%. Por lo tanto, se considera que los efectos cognitivos derivados de estas variaciones son triviales. La condición que implica mayor carga cognitiva es la ausencia total de sujeto (432.76 ms), mientras que la condición óptima es aquella con presencia parcial de sujeto pronominal (416.36 ms). En cambio, los parámetros de primera lectura (FRT) y de relectura (RRT) ponen de manifiesto diferencias considerables entre las condiciones examinadas, como presenta el Gráfico 17.

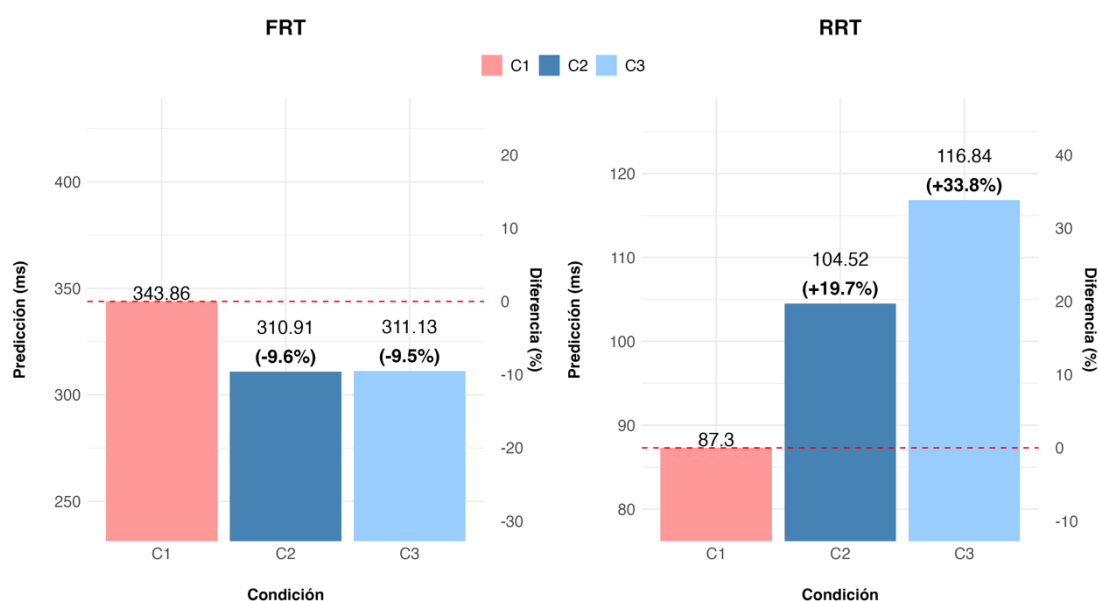


Gráfico 17. Comparativa de C1, C2 y C3 según tiempos de lectura de FRT y RRT en la OT

Los resultados de la primera lectura (FRT) indican que la condición con ausencia total de sujeto (C1) requiere un mayor tiempo de procesamiento para la construcción del primer supuesto, concretamente 343.86 ms, frente a condiciones con el sujeto explícito parcialmente (310.91ms en la C2) o duplicado en ambas oraciones (311.13 ms en la C3). De estas diferencias se extrae que la condición C1 conlleva un esfuerzo cognitivo notablemente mayor, con una magnitud mediana del efecto en torno al 9.5%, respecto a las condiciones C2 y C3. No obstante, las diferencias entre la presencia parcial (C2) y la duplicación del sujeto (C3) resultan irrelevantes en cuanto a la primera lectura se refiere.

Respecto a la fase de relectura de las oraciones (RRT), los efectos parecen ser más pronunciados y se observa un patrón de procesamiento diferente: cuando se ha invertido mayor tiempo procesando la primera lectura (como ocurre en la C1), se experimenta una reducción considerable en la duración de las relecturas (87.3 ms). Por el contrario, la presencia parcial del sujeto experimenta una mayor relectura que la ausencia total con un esfuerzo cognitivo grande (19%), aunque la condición con duplicación de referente exige aún más tiempo de relectura respecto a la ausencia de sujeto (efecto muy grande del 33%). Además, aunque de forma más atenuada, se advierten diferencias entre las condiciones C2 y la C3, con un efecto grande del 11%.

Una vez analizada la oración completa, desglosamos los resultados en función de la primera cláusula (OC1: *Inventaba historias preciosas*) y de la segunda (OC2: *Imaginaba personajes fantásticos*). La Tabla 48 recoge la comparativa de las tres condiciones según el tiempo total de lectura (TRT), tomando como ítem de control la C1.

TRT	Comparación	Ítem 1 (control)	Ítem 2	Diferencia (ms)	Diferencia (%)
OC1	C1 vs C2	469.59	438.19	31.40	-6.68
	C1 vs C3	469.59	438.28	31.31	-6.66
	C2 vs C3	438.19	438.28	-0.09	0.02
OC2	C1 vs C2	408.79	413.87	-5.08	1.24
	C1 vs C3	408.79	423.83	-15.04	3.67
	C2 vs C3	413.87	423.83	-9.96	2.40

Tabla 48. Comparativa de C1, C2 y C3 según tiempo de lectura de TRT en la OC1 y OC2

En una primera observación de los datos, los resultados indican que la primera parte de la oración es la que experimenta mayores diferencias en su procesamiento, mientras que la segunda parte de la oración parece no verse afectada por las condiciones. En el tiempo total de lectura (TRT), los efectos de la ausencia o presencia del sujeto parecen ser triviales en la segunda oración (OC2), mientras que la primera cláusula (OC1), al comparar la C1 con la C2 y C3, revela que los efectos son medianos entorno al 6%, siendo la condición de ausencia total la que demanda un mayor coste en su procesamiento (469.59 ms). En cambio, la primera cláusula de las condiciones de presencia parcial y duplicación pronominal no desencadena efectos cognitivos diferentes en el tiempo de lectura total. En cuanto a la segunda cláusula, los tiempos de lectura son similares en las tres condiciones y no consiguen tener un efecto cognitivo diferenciado.

Si desglosamos el tiempo total según la primera lectura y las relecturas, el Gráfico 18 exhibe las disparidades detectadas en la lectura de los dos enunciados en las tres condiciones estudiadas. Una primera visualización de los datos es suficiente para identificar que las diferencias de procesamiento recaen sobre la primera parte de la oración (OC1), tanto en la primera lectura como en las relecturas, ya que la segunda oración presenta tiempos de lectura similares en las tres condiciones.

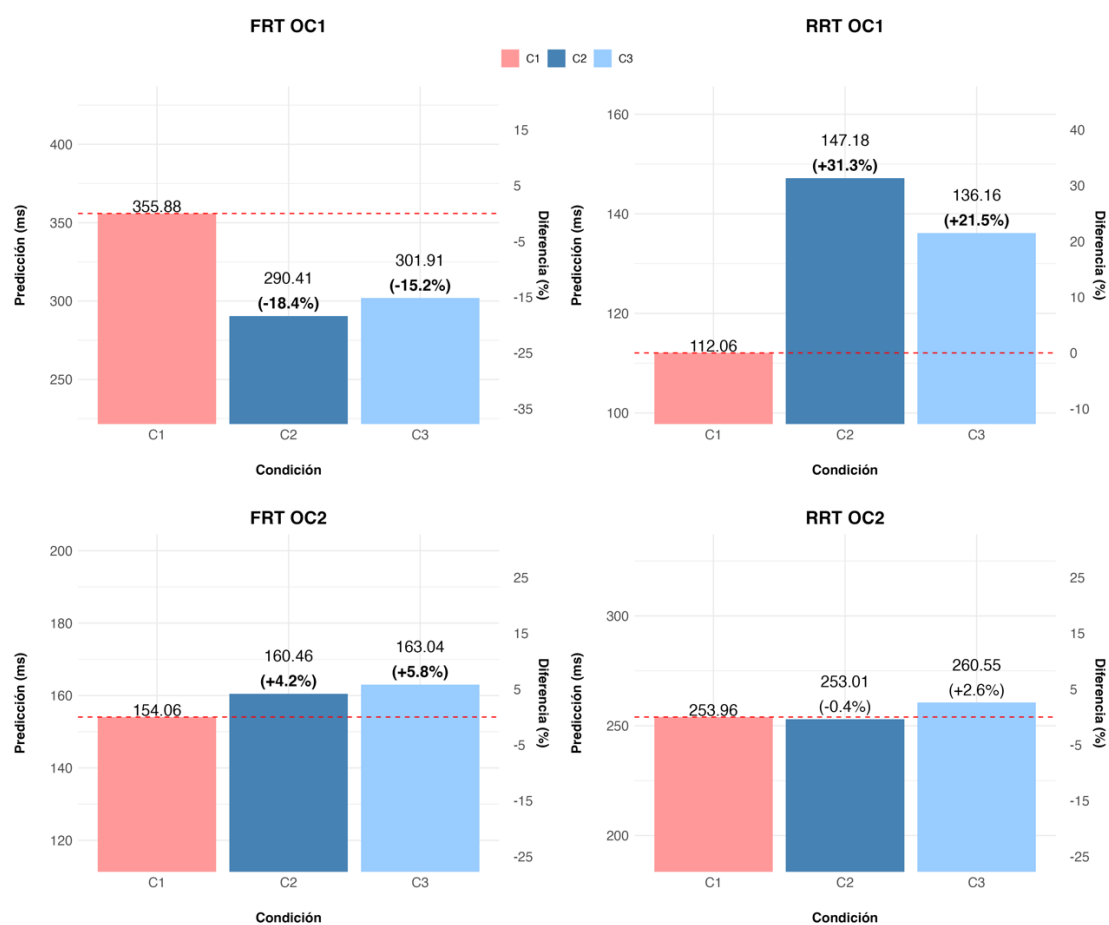


Gráfico 18. Comparativa de C1, C2 y C3 según tiempos de lectura de FRT y RRT en la OC1 y OC2

En cuanto a la primera lectura (FRT) del primer enunciado (OC1), en la condición de ausencia total de sujeto exige un esfuerzo cognitivo superior (355.88 ms) que el resto de las condiciones. Estas disparidades son considerablemente acentuadas, ya que la presencia parcial (C2) y la duplicación (C3) requieren una menor carga cognitiva con unos efectos grandes (-18% y -15%, respectivamente). No obstante, el tiempo invertido en la relectura de las condiciones parece ser diferente en la primera parte del estímulo, donde la condición con ausencia total es la que requiere menor tiempo en la relectura (112.06 ms). De modo que la necesidad de relectura manifestada es mayor en las condiciones presencia parcial (147.18 ms) y duplicación (136.16 ms) —aquellas que en las primeras lecturas necesitaban menos tiempo— con unos efectos cognitivos muy grandes respecto a la ausencia total (31% para la C2 y 21% para la C3).

Si comparamos la condición de presencia parcial (C2) y la duplicación referencial (C3), se observan efectos triviales en la primera lectura, pero estas diferencias se acentúan

en las relecturas: la ausencia parcial (147.18 ms) requiere un mayor tiempo en el procesamiento de la primera cláusula frente a la duplicación referencial (136.16 ms), lo que supone un efecto mediano (7.49%).

Tras identificar las disimilitudes en las áreas de interés mayores, comparamos áreas de interés más específicas, es decir, aquellas que sean idénticas excepto por un elemento. De este modo, las diferencias entre la C1 y la C2 se encuentran en la primera cláusula y las divergencias entre la C2 y la C3 se localizan en la segunda cláusula. Con el objetivo de corroborar si la presencia del pronombre logra resolver la ambigüedad, comparamos el área de interés OC1 de la condición de ausencia total (*Inventaba historias preciosas*) y el PV1 de la condición de presencia parcial (*inventaba historias preciosas*), donde ya se ha excluido el sujeto pronominal. Cabe destacar que en el Gráfico 19, para una correcta visualización del RRT, se ha aumentado el rango del eje derecho, correspondiente a la diferencia porcentual de C2 respecto a C1.

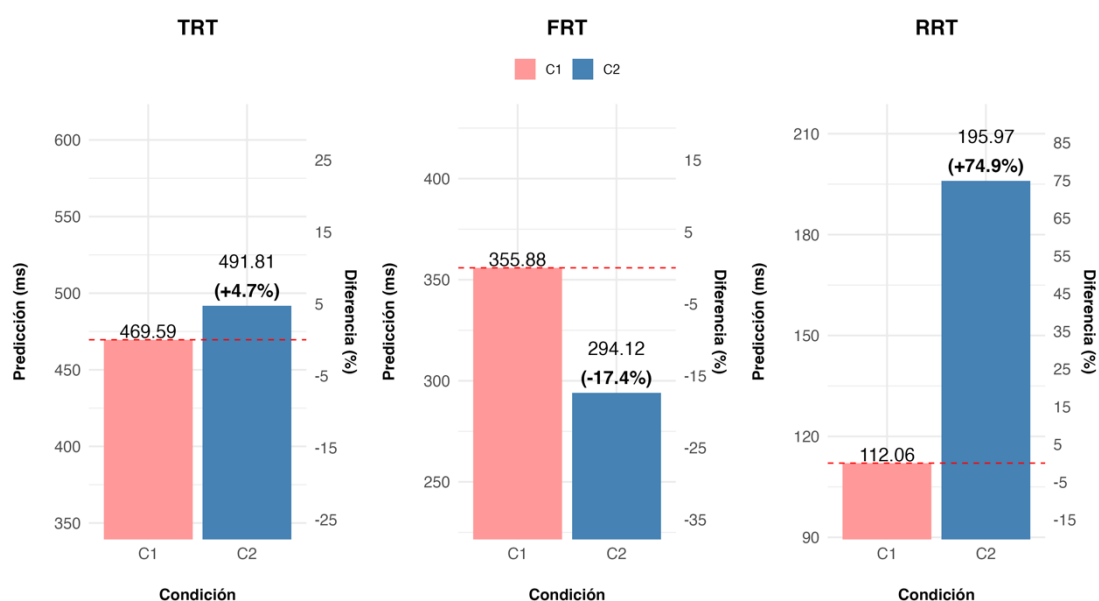


Gráfico 19. Comparativa de C1 y C2 según tiempos de lectura de TRT, FRT y RRT en la OC1 y PV1

Las diferencias en cuanto al tiempo de lectura total (TRT) reflejan un efecto pequeño de un 4% más de carga cognitiva en la condición con presencia parcial de sujeto (C2) respecto a la ausencia total (C1). Sin embargo, si prestamos atención a la primera lectura, el FRT muestra que la primera condición exige un mayor tiempo de procesamiento en la primera cláusula (355.88 ms) frente a la segunda condición (294.12

ms) con un efecto grande del 17%. Sin embargo, los tiempos de relectura (RRT) penalizan más a la condición de presencia parcial, ya que puede que el individuo vuelva a leer la primera cláusula con la intención de confirmar la primera explicitación referencial. De este modo, la lectura de la C2 supone un mayor coste en el procesamiento cognitivo (195.97 ms frente a 112.06 ms en la C2) con un efecto de gran magnitud del 74%.

Para contrastar los resultados de la duplicación referencial (C3) frente a la presencia parcial (C2), enfocamos la comparación en la segunda cláusula, donde se ha excluido el sujeto pronominal en la OC2 de la tercera condición. De este modo, el Gráfico 20 compara los tiempos de lectura en función de los tres parámetros analizados (TRT, FRT y RRT) en la OC2 de la condición con presencia parcial del sujeto (*Imaginaba personajes fantásticos*) y el PV2 de la condición con duplicación referencial (*Imaginaba personajes fantásticos*).

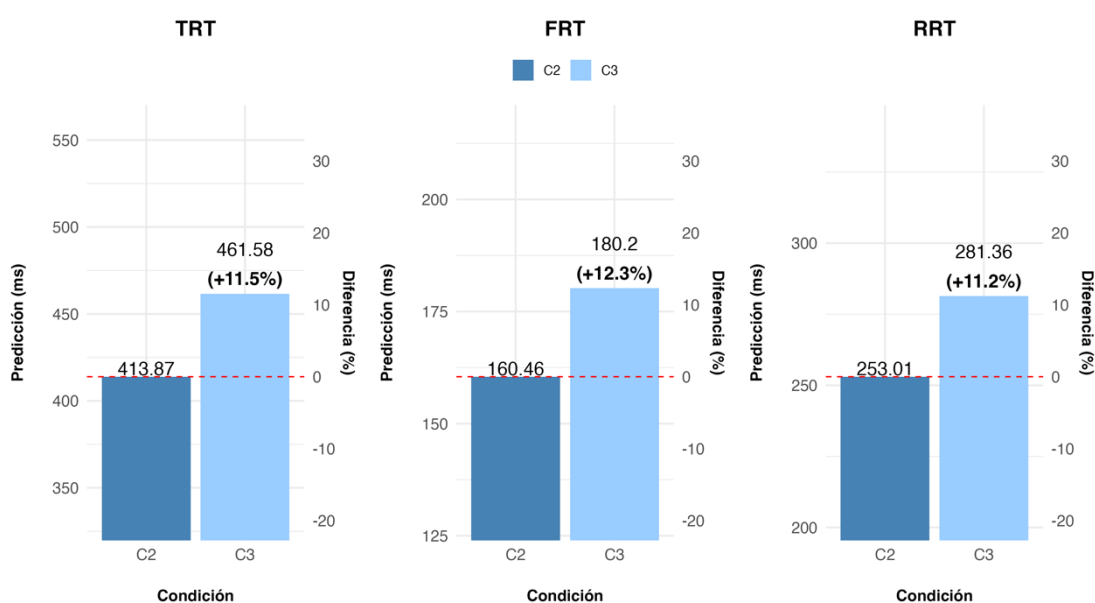


Gráfico 20. Comparativa de C2 y C3 según tiempos de lectura de TRT, FRT y RRT en la PV2

Las diferencias entre ambas condiciones son muy evidentes en los tres parámetros de lectura y, además, presentan el mismo patrón para la segunda cláusula: la duplicación del sujeto incrementa el tiempo de lectura del enunciado y, por lo tanto, la comprensión de dichas estructuras supone un esfuerzo cognitivo del 12% aproximadamente respecto a las oraciones con presencia parcial. Estas disparidades cognitivas, representadas en el Gráfico 21, afectan especialmente al segundo verbo del enunciado (*imaginaba*), cuyos

tiempos de lectura total son considerablemente superiores en la C3 (544.67 ms) respecto a la C2 (483.18 ms) suponiendo una diferencia cognitiva con un efecto grande (12%). Si bien en una primera lectura de estos verbos, las diferencias suponen un efecto mediado del 8%, lo que indica una mayor carga cognitiva en la C3, la principal diferencia reside en el tiempo de relectura. En estos casos, la duplicación referencial evidencia un mayor costo en el (re)procesamiento de la información, incrementando la carga mental en un 17% respecto a la presencia parcial del sujeto.

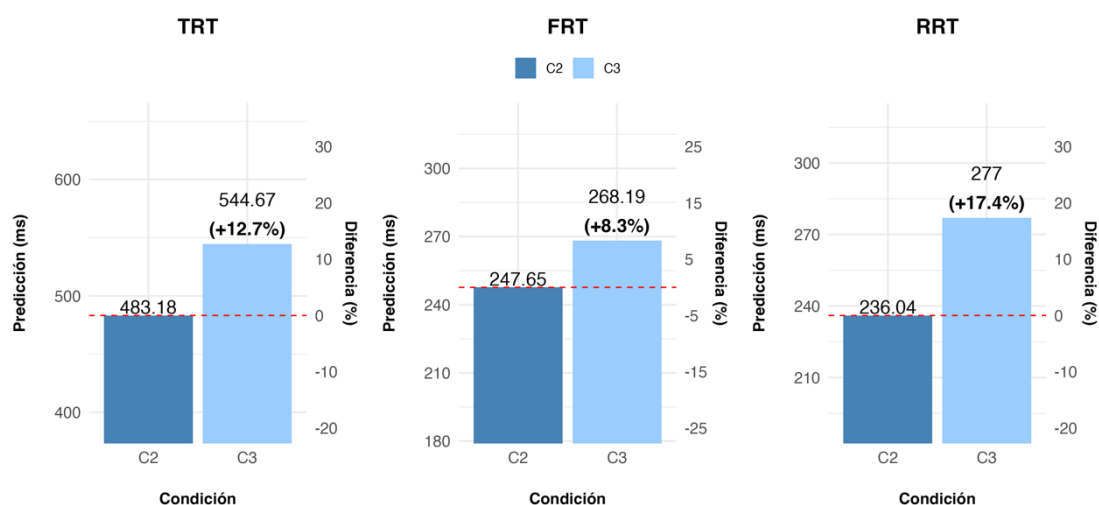


Gráfico 21. Comparativa de C2 y C3 según tiempos de lectura de TRT, FRT y RRT en la V2

3.5.2. Continuidad y cambio referencial

Para observar los efectos cognitivos del cambio referencial, procedemos a comparar la C4, tomándola como ítem de control, respecto al resto de condiciones que podrían considerarse como continuidad referencial en las oraciones con ausencia total de sujetos pronominales (C1), con presencia parcial (C2) y con duplicación referencial (C3). De este modo, las estructuras aquí estudiadas son las siguientes:

C4: Él inventaba historias preciosas. Yo imaginaba personajes fantásticos.

C1: Inventaba historias preciosas. Imaginaba personajes fantásticos.

C2: Él inventaba historias preciosas. Imaginaba personajes fantásticos.

C3: Él inventaba historias preciosas. Él imaginaba personajes fantásticos.

En un primer análisis de los estímulos, los tiempos obtenidos en la lectura total de la oración completa, comprendiendo los dos enunciados —*Él inventaba historias preciosas. Yo imaginaba personajes fantásticos*—, se muestran en la Tabla 49 para explorar el procesamiento de las cuatro condiciones.

OT	Comparación	Ítem 1 (control)	Ítem 2	Diferencia (ms)	Diferencia (%)
TRT	C4 vs C3	435.92	427.99	7.93	-1.81
	C4 vs C2	435.92	416.36	19.56	-4.48
	C4 vs C1	435.92	432.76	3.16	-0.72

Tabla 49. Comparativa de C4, C1, C2 y C3 según tiempo de lectura TRT en la OT

Al observar el TRT, las únicas diferencias leves se encuentran entre el cambio referencial (435.92 ms) y la presencia parcial (416.36 ms), donde la C4 conlleva una mayor carga cognitiva con una diferencia de pequeño efecto del 4.48%. Si diferenciamos entre el tiempo de la primera lectura y la relectura, percibimos en el Gráfico 22 una dinámica diferenciada durante el procesamiento.

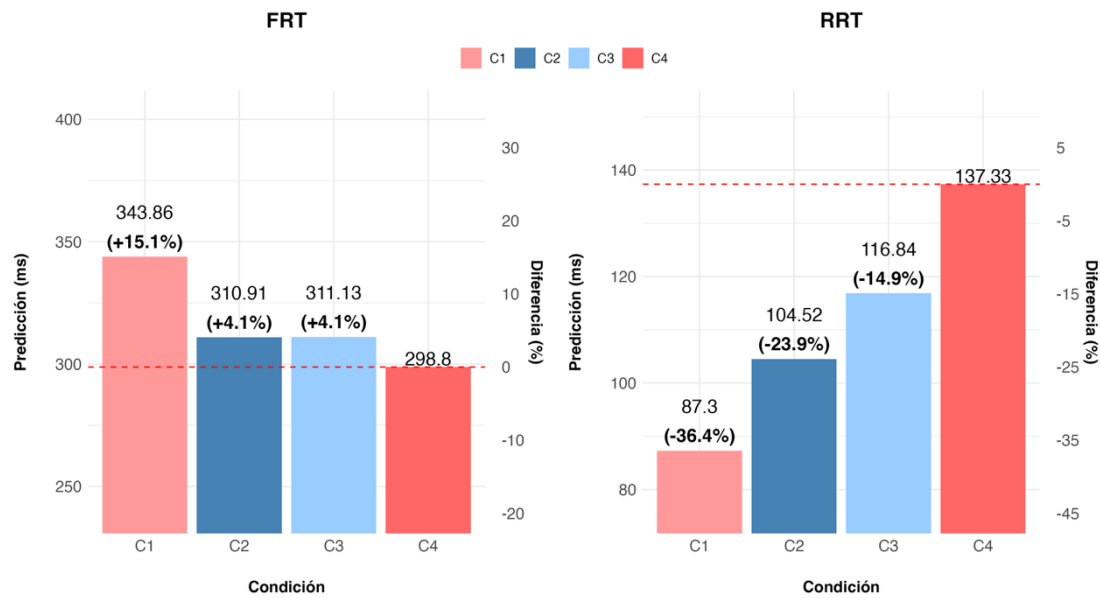


Gráfico 22. Comparativa de C1, C2, C3 y C4 según tiempos de lectura de FRT y RRT en la OT

En cuanto a la primera lectura, el cambio referencial (C4) requiere menor tiempo de procesamiento (298.80 ms) en comparación con las demás condiciones. Las diferencias entre la presencia parcial (C2) y la duplicación referencial (C3) son mínimas respecto al cambio referencial (C4), donde se refleja una carga cognitiva superior con un efecto pequeño del 4.1%. El efecto resulta considerablemente muy grande respecto a la ausencia total de sujeto pronominal, ya que esta condición (C1) demanda un mayor tiempo de procesamiento en la primera lectura (343.86 ms) y provoca un esfuerzo cognitivo mayor que el cambio referencial del 15.1%.

Por otro lado, el parámetro de relectura (RRT) presenta unas diferencias considerables en los tiempos de lectura entre las condiciones. El cambio referencial demanda un mayor tiempo de relectura (137.33 ms) y, por lo tanto, el procesamiento de estas estructuras supone una carga cognitiva superior que el resto de las condiciones. En otras palabras, la presencia parcial del sujeto (C2) constituye un esfuerzo mental inferior al 23.9%; la repetición del referente (C3) supone una sobrecarga cognitiva con un efecto grande del 14.9%; y, considerablemente, la ausencia total del sujeto (C1) tiene un coste menor al 36.4%.

Para poder estudiar con más detalle si los esfuerzos cognitivos se producen en la primera (OC1: *Él inventaba historias preciosas*) o la segunda (OC2: *Yo imaginaba personajes fantásticos*) parte de la oración, la Tabla 50 recoge los tiempos de lectura total según la condición analizada en comparación con el cambio referencial.

TRT	Comparación	Ítem 1 (control)	Ítem 2	Diferencia (ms)	Diferencia (%)
OC1	C4 vs C3	545.04	438.28	106.76	-19.58
	C4 vs C2	545.04	438.19	106.85	-19.60
	C4 vs C1	545.04	468.59	76.45	-14.02
OC2	C4 vs C3	424.83	423.83	1.00	-0.23
	C4 vs C2	424.83	413.87	10.96	-2.57
	C4 vs C1	424.83	408.79	16.04	-3.77

Tabla 50. Comparativa de C4, C1, C2 y C3 según el tiempo de lectura de TRT en la OC1 y OC2

Mientras que en la segunda cláusula los tiempos de lectura no presentan diferencias en cuanto a la carga cognitiva que supone su procesamiento, ya que todas las diferencias son mínimas, la primera cláusula muestra diferencias considerables. Por un

lado, en la condición de cambio de referente, la primera parte de oración es la que más tiempo requiere para ser procesada (545.04 ms) y refleja una carga cognitiva mayor que la presencia parcial (C2) y la duplicación referencial (C3) del 19%, lo que se refleja en un esfuerzo cognitivo elevado. Estas disparidades también se detectan en la comparación con la condición de ausencia total de sujeto, donde el cambio de referente incrementa en un 14% la carga cognitiva.

Con el objetivo de identificar si el esfuerzo cognitivo se produce en la primera construcción del supuesto (FRT) o si, por el contrario, la mayor carga cognitiva proviene de la reconsideración de lo leído (RRT), el Gráfico 23 representa los tiempos de lectura de las dos cláusulas (OC1 y OC2) en función de los parámetros de FRT y de RRT.

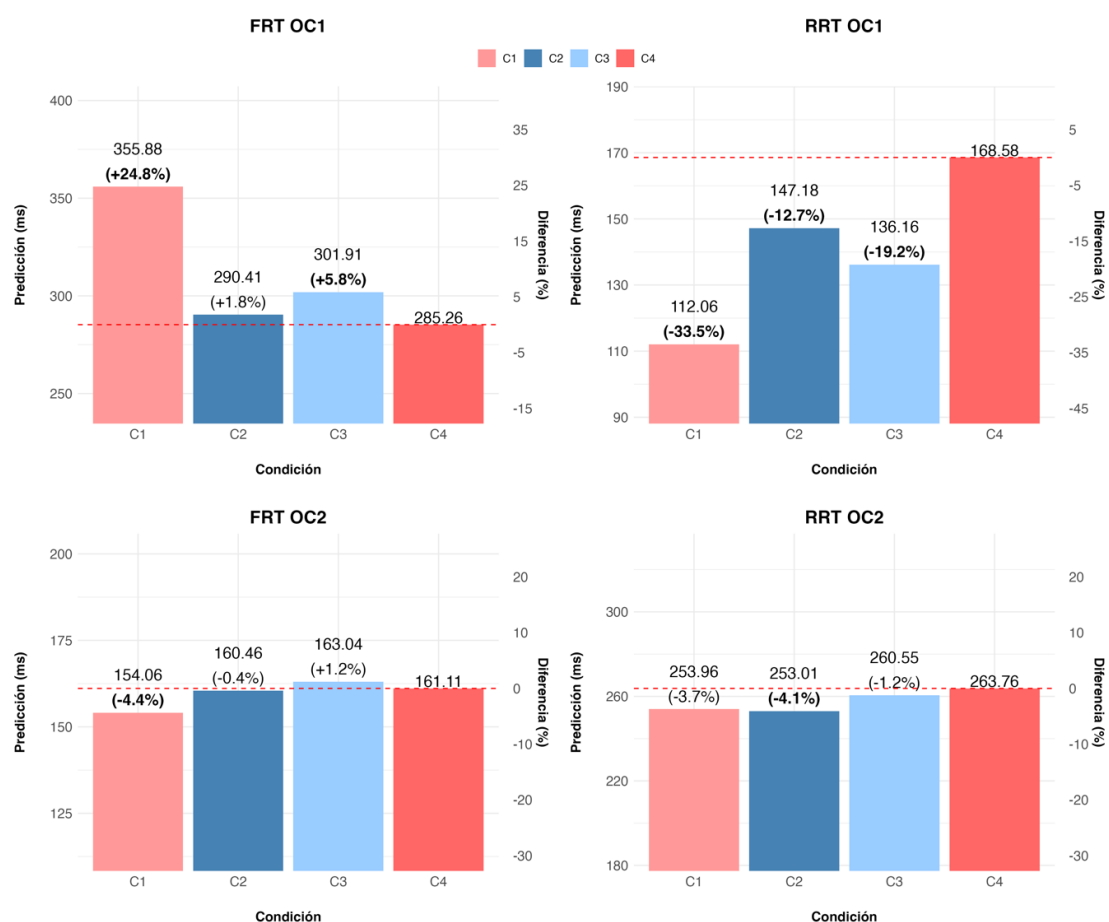


Gráfico 23. Comparativa de C4, C1, C2 y C3 tiempos de lectura de FRT y RRT en la OC1 y OC2

En cuanto al tiempo de la primera lectura (FRT), la parte inicial del estímulo (OC1) logra una lectura más rápida en la condición de cambio referencial (285.26 ms) que en el resto de las condiciones. Si bien la comparativa entre el cambio referencia (C4)

y la presencia parcial (C2) apenas presenta diferencias triviales, el procesamiento de la duplicación pronominal (C3) implica un esfuerzo cognitivo superior al cambio de referente (C4), que se refleja en un efecto mediano del 5.84%. Las diferencias entre la ausencia total de sujetos explícitos (C1) frente al cambio de referente (C4) son considerables, puesto que la condición con ausencia tiene un mayor coste de procesamiento (355.88 ms) en la primera oración, que se traduce en un efecto muy grande del 24.8% respecto a la condición con cambio de sujeto (285.26 ms). En esta comparativa, también detectamos diferencias de pequeño efecto (4.4%) en la segunda parte de la oración (OC2), donde la C4 necesita más tiempo para ser procesada que la C2. Las demás diferencias en cuanto al tiempo de la primera lectura en la segunda cláusula son mínimas y no logran generar efectos cognitivos dispares.

De acuerdo con el parámetro de RRT, la condición de cambio de referente (C4) presenta, durante la primera parte de la oración (OC1), un mayor tiempo de lectura destinado aclarar o confirmar lo interpretado tras la primera lectura. Al comparar el cambio referencial con otras condiciones, las diferencias entre ellas presentan efectos cognitivos considerables: la duplicación referencial (-19.2% para la C3) y la presencia parcial del sujeto (-12.7% para la C2). Cabe destacar las grandes disparidades en relación con el tiempo de relectura de la condición con ausencia total de sujeto (112.06 ms), que supone un 33.5% menos de costo cognitivo. La segunda cláusula (OC2) muestra una relectura similar en las cuatro condiciones, a excepción de la condición de presencia parcial (253.01 ms en la C2), que demanda un esfuerzo cognitivo menor que el cambio referencial (263.76 ms en la C2) con un efecto pequeño del 4%.

Con el objetivo de corroborar si la introducción de un cambio de referente implica diferencias en la percepción cognitiva del estímulo, confrontamos las áreas de interés de pequeña escala referentes a la segunda parte de la oración (OC2), donde se produce la explicitación pronominal del nuevo referente. Extendemos esta comparación, por un lado, entre el cambio referencial (C4) y la duplicación (C3), puesto que en ambas cláusulas se presenta el sujeto de forma expresa, aunque uno indique repetición del referente anterior y el otro sea un referente nuevo. Por otro lado, cotejamos los tiempos de lectura de la condición de presencia parcial (C2), entendida como continuidad referencial sin sujeto expreso, frente al cambio referencial (C4). Para ello, comparamos aquella parte de la

oración que es idéntica, excepto por el elemento que marca la diferencia entre condiciones.

El Gráfico 24 compara los tres parámetros de tiempo para el área de interés PV (*imaginaba personajes fantásticos*) en las condiciones C3 y C4. A pesar de que en el tiempo de total de lectura no se detectan diferencias entre las dos condiciones —466.31 ms para la cuarta condición y 461.58 ms para la tercera condición—, el tiempo de primera lectura (TRT) refleja que la duplicación referencial (180.02 ms para la C3) incrementa el coste de procesamiento un 9%, lo que supone una magnitud de efecto mediano respecto al cambio referencial (165.3 ms). En lo que respecta a las relecturas, la condición que introduce un nuevo referente requiere un tiempo de relectura mayor (301.05 ms en la C4) con un efecto mediano del 6%. Por lo tanto, aunque el tiempo total de lectura sea muy similar en ambas condiciones, la duplicación referencial tarda más en procesarse en la primera lectura y requiere una menor relectura y, por el contrario, el cambio referencial exige una mayor relectura para garantizar la comprensión de la oración.

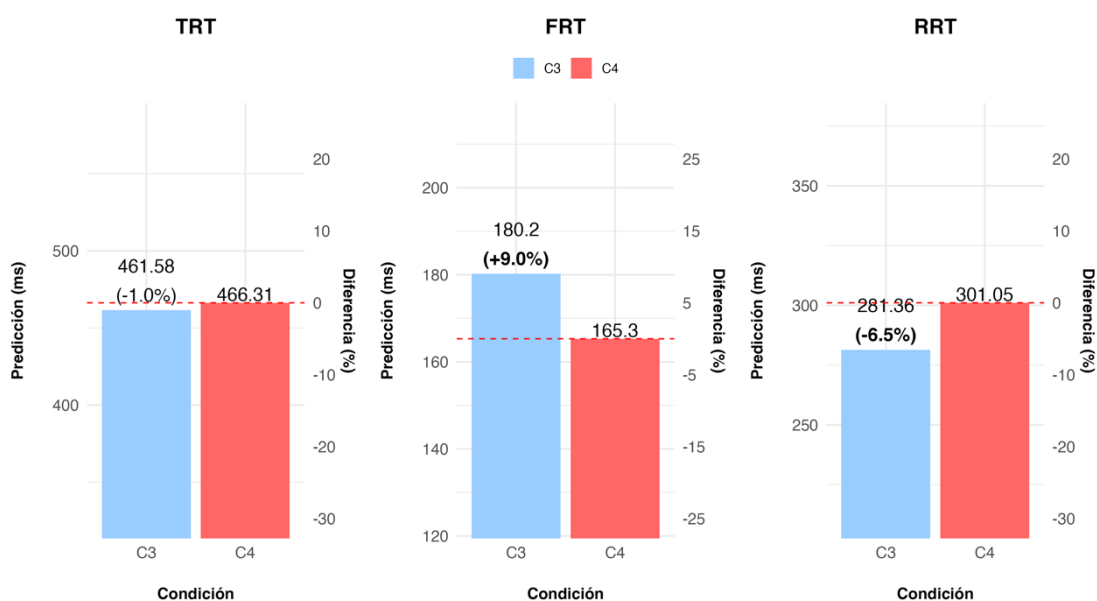


Gráfico 24. Comparativa de C3 y C4 según tiempos de lectura de TRT, FRT y RRT en la PV2

Por último, el Gráfico 25 presenta la comparativa de los tiempos obtenidos según la condición de presencia parcial en la OC2 (*Imaginaba personajes fantásticos*) y en el PV2 de la condición de cambio de referente (*imaginaba personajes fantásticos*), dado que estas áreas de interés contienen la misma información.

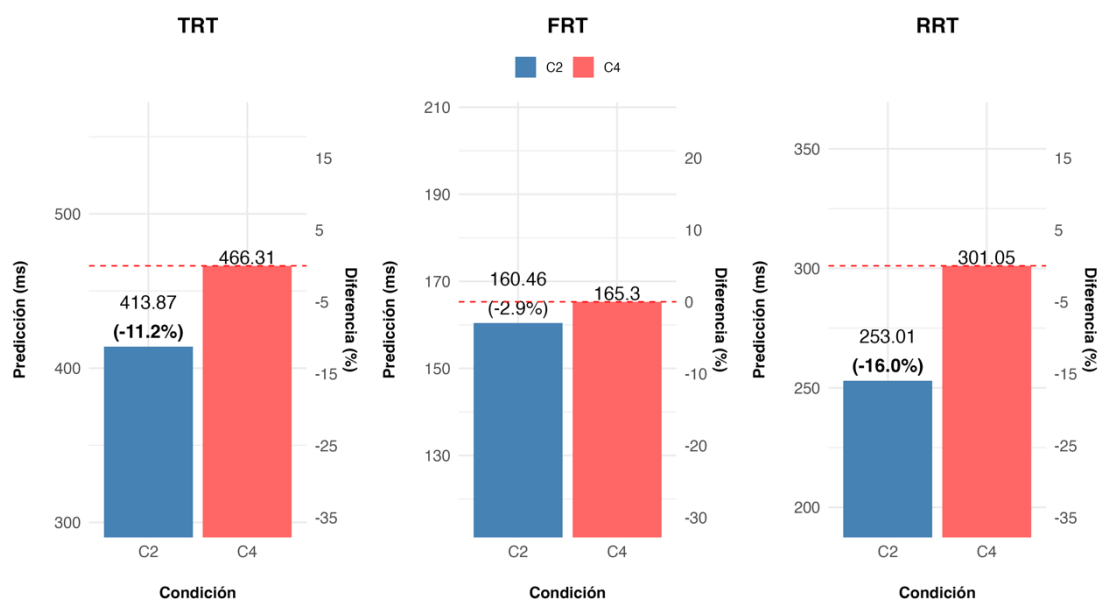


Gráfico 25. Comparativa de C2 y C4 según tiempos de lectura de TRT, FRT y RRT en la OC2 y PV2

En una primera exploración de los datos, identificamos que la condición de presencia parcial (C2) precisa de un menor coste de procesamiento en los tres parámetros de lectura. El tiempo total de lectura ya refleja que la introducción de un nuevo referente implica una carga cognitiva superior a la continuidad referencial con un efecto grande del 11%. La primera lectura (FRT) no presenta diferencias entre las condiciones, pero el cambio referencial (301.05 ms) incrementa su tiempo de relectura un 16% respecto a la presencia parcial (253.01 ms).

3.6. Discusión y conclusiones parciales

Este estudio de percepción cognitiva a través de la herramienta de *eye-tracking* pone de manifiesto la existencia de diferentes patrones de procesamiento según las condiciones estudiadas. Los factores lingüísticos implementados en los estímulos —la ausencia/presencia del sujeto, la ambigüedad morfológica y la correferencialidad— activan diferentes dinámicas de procesamiento en términos de esfuerzo cognitivo.

A partir del tiempo total invertido en el procesamiento de cada estructura, establecemos una comparativa entre las condiciones estudiadas en función de su grado de

complejidad cognitiva. El Gráfico 26 toma como referencia la condición 2, al presentar esta la percepción cognitiva más eficiente.

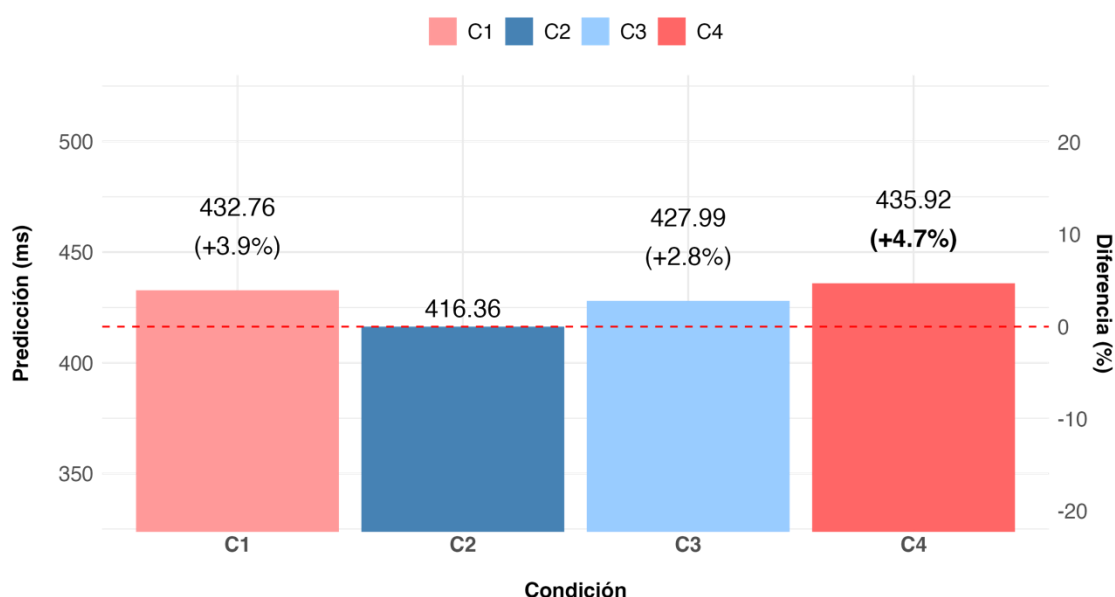


Gráfico 26. Comparativa de las cuatro condiciones según el tiempo total de lectura del conjunto de los enunciados (OT)

El estímulo con presencia parcial de sujeto constituye la condición que refleja un equilibrio óptimo entre el esfuerzo invertido en el procesamiento inicial de la oración y la necesidad de volver a leer, especialmente el primer enunciado, para asegurar la correcta comprensión. De este resultado se infiere que la presencia del sujeto pronominal en una de las cláusulas facilita la percepción cognitiva de las estructuras, de modo que la ambigüedad morfológica de los verbos quedaría neutralizada con la explicitación pronominal. Esto favorece la interpretación del segundo enunciado como continuidad referencial, cuyo antecedente se encontraría en la cláusula inmediatamente anterior.

Al contrastar esta percepción cognitiva con el uso real del sujeto pronominal observado en los corpus PRESEEA, podemos corroborar de manera empírica la propuesta de Silva-Corvalán sobre el papel desambiguador del contexto cuando la forma verbal es morfológicamente ambigua. Además, nuestros resultados de la ciudad de Málaga revelan que, ante situaciones de ambigüedad no contextual, el pronombre se omite en un 72% en el corpus de 1995 y un 78% en el corpus de 2015. Por tanto, se podría confirmar que las altas tasas de ausencia están justificadas cognitivamente, debido a que la omisión optimiza la comprensión global del enunciado.

Por el contrario, la duplicación referencial, aunque presenta un patrón de lectura análogo a la presencia parcial de sujeto, conlleva unos tiempos de lectura superiores en los tres parámetros analizados. Al tratarse de contextos ya definidos como no ambiguos por la explicitación del sujeto en la primera cláusula, la repetición del sujeto en la segunda parte de la oración genera una sobrecarga en el procesamiento del estímulo en general. Eso ralentiza especialmente la relectura necesaria para confirmar que esta doble explicitación pronominal no conduce a un nuevo referente, sino que reitera el referente anterior. Aunque estos casos no son demasiado numerosos en nuestros corpus, se observan ejemplos de duplicación referencial en las entrevistas analizadas, donde el pronombre se usa de manera explícita en un 28% en el corpus de 1995 y en un 22% en el corpus de 2015.

Asimismo, detectamos dos dinámicas de procesamiento distintas ante la resolución de la ambigüedad y el cambio referencial. Por un lado, las oraciones de ausencia total son las que más tiempo de lectura demandan para el procesamiento y la recreación de un primer supuesto. Ante la ausencia total de sujetos explícitos y formas verbales morfológicamente ambiguas, el individuo tarda más en procesar por primera vez la oración. Esto podría deberse probablemente a que esté decidiendo a quién atribuirle la acción de los dos posibles sujetos (1ª o 3ª persona del singular). No obstante, una vez resuelta la ambigüedad o, por el contrario, abandonada la selección del sujeto potencial, el tiempo dedicado a la relectura para garantizar la comprensión del enunciado total es significativamente menor que el resto de las condiciones. Por lo tanto, deducimos que la ausencia total de sujeto pronominal, acompañada de formas ambiguas, obliga al individuo a emplear un gran coste cognitivo en el procesamiento de dichas estructuras y, además, puede comprometer la comprensión del discurso.

Aunque tanto la ausencia total del sujeto como la introducción de un nuevo referente exigen más esfuerzo cognitivo por parte del lector para su comprensión, el cambio referencial presenta el patrón de procesamiento inverso al de la ambigüedad. En este tipo de estructuras, la primera recreación mental del estímulo leído se produce de manera más rápida que en el resto de los casos, pero el proceso posterior de asignar correctamente quién realiza cada acción ante dos referentes incrementa considerablemente la carga cognitiva. Si contrastamos el uso del sujeto en contextos de cambio referencial con verbos ambiguos morfológicamente, la presencia del sujeto ronda

el 50% en ambos corpus. De este modo, si se omitiese el sujeto en este tipo de enunciados, el receptor podría percibirlo como continuidad referencial si solo se omite uno de los sujetos, o como una oración potencialmente ambigua en caso de doble omisión.

La incorporación de este análisis complementario mediante *eye-tracking* ha proporcionado evidencia experimental de los patrones de procesamiento implicados en la lectura de determinadas estructuras. La integración del análisis de corpus PRESEEA-Málaga, con el fin de identificar las variables que afectan a la presencia del pronombre y determinar su frecuencia de uso, junto con los resultados experimentales ofrece una visión integradora del fenómeno lingüístico estudiado. Este estudio complementario refuerza las hipótesis planteadas y permite conocer con detalle cómo se procesan cognitivamente ciertas estructuras detectadas en el uso real. En síntesis, hemos corroborado las hipótesis planteadas inicialmente.

La condición con presencia parcial de sujeto (C2) resulta menos exigente en términos de esfuerzo cognitivo, de la cual deducimos que la omisión pronominal en la segunda cláusula facilitaría la consideración de la continuidad correferencial. La optimización en el procesamiento del enunciado completo que supone la presencia parcial del sujeto corrobora la *hipótesis de la ventaja*. Nuestros resultados coinciden con la Teoría planteada por Ariel (1990), quien señala una alta accesibilidad cuando la conexión entre el antecedente y la anáfora es estrecha y se refleja un menor costo cognitivo en su procesamiento.

En cambio, la duplicación referencial (C3), aunque es la segunda condición que requiere menos tiempo de procesamiento, ha demostrado que la explicitación del pronombre personal sujeto en sus dos cláusulas penaliza el procesamiento de la oración, ya que la repetición parece reflejar una sobrecarga cognitiva. Este resultado se alinea con el planteamiento de Gernsbacher (1989) y, por lo tanto, corrobora nuestra *hipótesis de penalización por repetición*.

La doble ausencia de sujeto explícito (C1) ocupa el tercer lugar en cuanto a tiempo de procesamiento requerido. Esta ausencia de sujeto parece generar cierto grado de ambigüedad y, en consecuencia, un esfuerzo cognitivo notablemente mayor para identificar el referente, de modo que se verifica la *hipótesis de ambigüedad*. Los resultados obtenidos coinciden con lo esperado porque el efecto de competición o el efecto de múltiples candidatos (Gernsbacher, 1989; Badecker & Straub, 2002) exige al

individuo más esfuerzo de procesamiento. No obstante, una vez que la ambigüedad ha sido resuelta, el participante demanda menos tiempo de relectura.

Por último, el cambio de referente (C4) implica el mayor esfuerzo cognitivo por parte del lector. Estos indicios siguen la línea de los obtenidos en el estudio de Khan (1999), ya que considera que el cambio referencial, al concebirse como eventos separados, puede retrasar y dificultar el proceso de comprensión. De este modo, se confirma la *hipótesis de cambio referencial*, que anticipaba un mayor coste de procesamiento debido a la introducción de un nuevo referente.

Las cuatro estructuras lingüísticas estudiadas han arrojado diferencias en sus tiempos de lectura y, por lo tanto, en el procesamiento cognitivo que hace el participante al leer los estímulos. Gracias al análisis de los diferentes parámetros temporales, hemos detectado que los tiempos de lectura entre el FRT y el RRT varían al comparar las condiciones que presentan ambigüedad referencial respecto al resto de las condiciones, especialmente, en los estímulos de cambio referencial. A pesar de las diferentes estrategias llevadas a cabo para la percepción cognitiva de los enunciados, el Gráfico 27 representa el tiempo total de lectura (TRT) en milisegundos de las cuatro condiciones estudiadas.

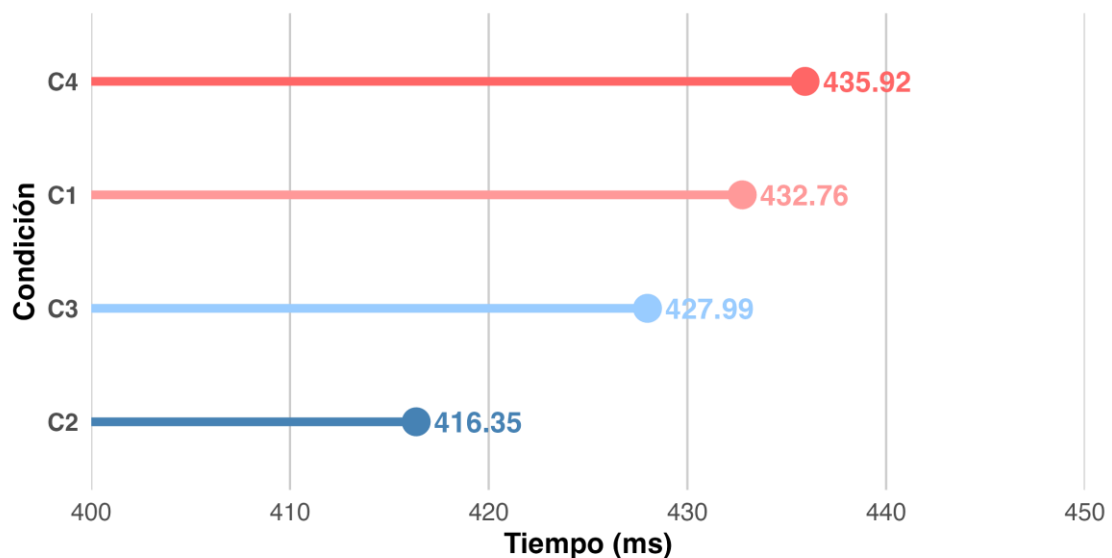


Gráfico 27. Ordenación de las cuatro condiciones según el tiempo total de lectura del conjunto de los enunciados (OT)

Estos datos confirman que la presencia del sujeto en una de las cláusulas es un factor clave cognitivamente para facilitar el proceso de comprensión, ya que el estímulo con presencia parcial del sujeto es el que requiere menor tiempo para la lectura total del enunciado (416.35 ms). A partir de ahí, las demás condiciones se alejan de manera progresiva. Por su parte, la duplicación referencial que introduce una pequeña sobre carga al tener que gestionar la información redundante (427.99 ms). Al eliminar por completo cualquier anclaje cognitivo en el discurso, los estímulos con ambigüedad total incrementan su tiempo de lectura (432.76 ms), ya que el lector tiene que activar mecanismos compensatorios para resolver dicha ambigüedad. Finalmente, los estímulos con cambio de referente son los que más tiempo requieren (435.92 ms), ya que la incorporación un nuevo referente obliga al lector a reconstruir el modelo mental del texto.

La diferencia de aproximadamente 20 ms entre la condición más eficiente y la más costosa cognitivamente revela el papel decisivo de la presencia o ausencia del sujeto pronominal en la comprensión del enunciado: cuando su presencia resulta superflua — solo aporta información redundante— o bien se explicita para señalar un cambio referencial, el lector tiende a incrementar su esfuerzo cognitivo para comprender el enunciado que acaba de leer; en cambio, el uso parcial del sujeto entendido como una continuidad referencial agiliza y facilita la comprensión textual.

CAPÍTULO 4. CONCLUSIONES

El estudio de la variación del sujeto pronominal en español abordado desde el enfoque innovador de la sociolingüística cognitiva (Caravedo, 2014; Moreno Fernández, 2012) ha aportado contribuciones notables por varios motivos. La observación de la variación sintáctica a partir de la integración dos ópticas complementarias, como son la producción y la percepción, nos ha permitido acercarnos a una visión más real del funcionamiento de este fenómeno lingüístico. La producción ha resultado esencial para determinar la frecuencia de aparición y detectar qué variables lingüísticas —y, en ocasiones, sociales— han motivado una mayor explicitación del sujeto. Por su parte, la percepción ha complementado los resultados obtenidos en los corpus al adentrarse en el procesamiento cognitivo observado en los individuos ante la lectura de las oraciones mediante la técnica experimental de *eye-tracking*.

En lo referente a la producción, la variación sintáctica que conlleva la ausencia y presencia del pronombre personal ha manifestado ser una realidad multidimensional, puesto que se ha demostrado la influencia de factores de diversa naturaleza: fonéticos, morfológicos, sintácticos, pragmáticos, discursivos y sociales. Superada queda ya la afirmación de que la variación sintáctica, y especialmente la ausencia o presencia del

pronombre, responde a una alternancia libre atribuida a cuestiones estilísticas más que gramaticales (Fernández Soriano, 1999). En este trabajo se ha podido dejar constancia de la existencia de ciertos factores recurrentes que motivan la presencia del sujeto, y también se ha evidenciado una similitud en cuanto a los condicionantes gramaticales identificados en las investigaciones anteriores.

En primer lugar, este trabajo ha aportado datos inéditos referidos al uso de la expresión del sujeto pronominal en el habla de Málaga y, de esta forma, se ha podido situar esta variable sintáctica en el mapa de la ciudad lingüística. Gracias a los corpus recogidos por el grupo VUM, hemos podido estudiar las voces de la ciudad en dos cortes temporales con un lapso de 20 años. Este estudio diacrónico ha manifestado continuidad en los usos lingüísticos y en las variables que condicionan la ausencia y la presencia del sujeto pronominal, pero también nos ha ofrecido indicios acerca de la posible evolución en el tiempo del empleo del sujeto pronominal.

Además, al realizarse esta investigación en el marco del proyecto PRESEEA, la utilización de la metodología consensuada por todos los equipos que forman parte del proyecto nos ha permitido establecer comparaciones con los datos obtenidos previamente en otras comunidades de habla hispana. A pesar de que se ha puesto de manifiesto la existencia de ciertas disparidades entre las áreas geográficas analizadas —como las diferentes tasas de presencia del sujeto pronominal—, también se ha constatado una notable coherencia diatópica en cuanto a los factores condicionantes y se han detectado reglas generales y parciales que justifican la expresión del sujeto pronominal.

Con el objetivo de adaptar y actualizar la propuesta metodológica a la ciudad de Málaga, hemos incorporado algunas variables como el referente del sujeto y la ambigüedad fonética y, además, se han revisado aquellas variables donde la incorporación del contexto a su análisis proporcionaba un enfoque más integrador, especialmente ante las posibles ambigüedades tanto fonéticas como morfológicas. Esta ampliación y reconfiguración orientada a explorar las variables dentro su contexto, en lugar de analizarlos de manera aislada, nos ha permitido observar de forma más real la dinámica natural entre los factores lingüísticos que se entrelazan y se condicionan mutuamente en el discurso.

El objetivo principal de este estudio de corpus se ha cumplido al determinar la frecuencia de aparición y al identificar los factores lingüístico-sociales que inciden en el uso del sujeto pronominal en la ciudad de Málaga. Asimismo, consideramos que los

resultados respaldan con garantías las hipótesis generales desarrolladas al inicio del trabajo, ya que algunas de ellas no solo han hallado apoyo en la ciudad de Málaga, sino que también han evidenciado su transversalidad al observarse en otras comunidades hispanohablantes.

La primera de nuestras hipótesis (*hipótesis de la gramática subyacente*) planteaba que, a pesar de las diferencias dialectales y temporales, existiría una gramática subyacente que unificaría los mecanismos lingüísticos que conducen a una mayor expresión del sujeto pronominal (Orozco et al., 2024). Esto ha podido verificarse a través de la identificación de un patrón lingüístico extensible a doce trabajos desarrollados en diferentes áreas dialectales, lo que ha revelado que la persona del sujeto es la variable más explicativa en todos los estudios; especialmente, la 1ª y 2ª persona del singular son las que condicionan la mayor aparición del pronombre. El factor correferencialidad, cuando se produce un cambio en el referente, requiere más explicitación del sujeto. Los verbos que acompañan al pronombre favorecen la expresión cuando están en tiempo presente e imperfecto, cuando el modo es el indicativo y la clase semántica del verbo es de procesos mentales o de cognición.

Los resultados han demostrado la mayor influencia de las variables lingüísticas sobre las variables sociales, que no han presentado una tendencia tan uniforme y no han conseguido ser tan explicativas como las lingüísticas, cumpliéndose así nuestra segunda hipótesis (*hipótesis del peso de las variables*) que atribuía una mayor importancia a los factores lingüísticos (Martín Butragueño, 2020a). No obstante, es probable que un estudio en profundidad de las variables sociales más allá de las empleadas para la preestratificación de los hablantes, como las relacionadas con la influencia del capital cultural del informante (Bourdieu, 1984, 1986), pueda revelar diferencias en el uso del pronombre personal sujeto que todavía no se estén considerando.

La combinación del análisis de corpus y la experimentación cognitiva han aportado una mayor solidez metodológica, ya que nos ha permitido complementar el uso de diferentes variables con el procesamiento mental que el individuo experimenta ante las estructuras presentes en su lengua (Recio Fernández et al., 2021). Estudiar cómo la variación de la ausencia y presencia del sujeto pronominal influye en el entendimiento de los enunciados ha proporcionado información esencial sobre la función que cumple el sujeto en ciertos contextos y, además, se han podido identificar diferencias en función del

esfuerzo cognitivo requerido según el estímulo presentado. Con ello, hemos cumplido con el propósito principal del bloque de percepción.

El análisis experimental de *eye-tracking* también ha revelado diferentes estrategias cognitivas, especialmente ante la resolución de la ambigüedad y la incorporación de un nuevo referente en los enunciados. Los estímulos que presentaban una ambigüedad referencial total (C1) eran los que requerían más esfuerzo cognitivo en la primera lectura, y contrastaba, a su vez, con el menor tiempo de procesamiento en relectura. En cambio, la introducción de un cambio referencial (C4) presentaba la tendencia contraria: la primera lectura era más rápida, pero sus relecturas indicaban un alto coste en el procesamiento de la información.

Nuestras hipótesis iniciales se han visto respaldadas, puesto que el pronombre personal sujeto ha funcionado como anclaje cognitivo antes la resolución de la ambigüedad con verbos morfológicamente idénticos. De este modo, los estímulos con un solo sujeto de forma explícita (C2) han resultado ser enunciados óptimos en cuanto al esfuerzo manifestado por parte del individuo y la recompensa, en este caso, el entendimiento de la oración. Por lo tanto, se ha verificado nuestra *hipótesis de la resolución pronominal* (Ariel, 2001), ya que el pronombre personal sujeto ha reducido la carga cognitiva respecto a las oraciones con ausencia total de sujeto.

A pesar del sustento cognitivo que supone la presencia del sujeto pronominal en estos estímulos, se detecta que el sujeto favorece o dificulta la lectura según su grado de aparición. La duplicación referencial (C3), en lugar de facilitar el entendimiento de la oración por acompañar a formas morfológicamente ambiguas, ha reflejado en sus tiempos de lectura una sobrecarga en comparación con la presencia parcial del sujeto. De esta forma, los resultados permiten apoyar la *hipótesis del grado de aparición*, ya que la explicitación del sujeto favorece la comprensión del enunciado, pero la repetición de un mismo referente o el cambio referencial han demostrado aumentar la carga cognitiva (Badecker & Straub, 2002; Gernsbacher, 1989; Grant et al., 2020).

El estudio de la ausencia y presencia del sujeto pronominal en español, abordado desde la perspectiva de la sociolingüística cognitiva, ha aportado una visión integradora al relacionar lo lingüístico y lo social con el procesamiento cognitivo. Nuestro trabajo ha puesto de manifiesto la necesidad de analizar los fenómenos lingüísticos desde la multidimensionalidad que supone considerar la producción —a partir del uso real—, y la

percepción —a través de los recursos cognoscitivos implicados en su procesamiento—. Además, los resultados obtenidos han puesto de relieve la necesidad de situar el análisis en su contexto comunicativo, desde el entorno lingüístico más inmediato del enunciado hasta el conocimiento que ya se encuentra integrado en la mente del individuo.

Es preciso mencionar también que este trabajo no está exento de limitaciones. La primera de ellas reside en la naturaleza de los corpus utilizados, ya que, al estar basados en entrevistas semidirigidas, las intervenciones están más controladas en el habla espontánea y el contexto comunicativo está limitado al entrevistador y al entrevistado, lo que ha podido restringir el uso de las personas gramaticales del plural. Resultaría conveniente en futuras investigaciones incorporar situaciones comunicativas más variadas y naturales, como conversaciones espontáneas y cotidianas que puedan aportar una visión más natural del fenómeno lingüístico estudiado.

En segundo lugar, pese al avance que supone el uso de técnicas como el *eye-tracking* para medir el procesamiento cognitivo, este método también presenta limitaciones intrínsecas, como la posible interferencia del equipamiento o la artificialidad del entorno experimental. En futuras investigaciones se podría considerar la inclusión de otras pruebas complementarias de carácter perceptivo, como cuestionarios *offline* que puedan enriquecer los resultados obtenidos mediante *eye-tracking*.

La lengua, al ser un fenómeno vivo, en constante cambio y evolución, no acaba nunca de ser observada por completo. Esta realidad invita a realizarnos nuevas preguntas, algunas surgidas a partir de las reflexiones finales de esta investigación y otras que llevan presentes desde los primeros planteamientos de este trabajo. Como consecuencia, en lugar de considerar estos primeros hallazgos como punto de llegada, los entendemos como una oportunidad para seguir desarrollando y expandiendo el alcance del análisis de esta variable lingüística.

4.1. Próximos objetivos

Todavía quedan muchos aspectos relacionados con este fenómeno de variación lingüística que pueden ser explorados en futuros trabajos; por ello, a continuación,

destacamos algunas líneas de investigación que, bajo nuestro criterio, resultarían especialmente relevantes.

En primer lugar, consideramos imprescindible seguir ampliando la dimensión diacrónica de la observación lingüística de la ciudad de Málaga. Con el objetivo de obtener un entendimiento más profundo de los cambios de uso pronominal en el español hablado en la urbe malagueña, sería pertinente retrotraernos a los materiales del corpus Arqueoling (Ávila Muñoz et al., 2024) que recogen las voces de la ciudad desde los años 40 hasta los 90. Este análisis contrastivo nos permitiría documentar los cambios históricos, por ejemplo, en cuanto a las preferencias de uso respecto a los pronombres de 2ª persona, concretamente la variación en el uso de los sujetos pronominales *tú* y *usted*.

Además, gracias a la recopilación de entrevistas más actuales, trabajar con los datos contenidos en el corpus MAYHO (Mujer Ayer y Hoy) (Cruces Rodríguez, en prensa) ofrecería la oportunidad de investigar si las transformaciones sociales que han experimentado las mujeres en las últimas décadas tienen consecuencias en sus usos lingüísticos. En este corpus de nueva planta podríamos confirmar si los indicios detectados sobre una mayor explicitación del sujeto pronominal por parte de las mujeres representan el inicio de un cambio lingüístico. Incluso, si esta tendencia hacia una mayor explicitación pronominal pudiera llegar a constituir uno de los rasgos característicos de la variedad intermedia identificada previamente en el contexto sociolingüístico malagueño.

Por último, sugerimos un estudio acústico-perceptivo detallado sobre la relación de la ambigüedad fonética y la expresión del sujeto pronominal. Aunque la elisión de /-s/ en coda silábica de la forma verbal es un fenómeno fonético ampliamente extendido y estudiado en la ciudad de Málaga (Vida Castro, 2004), no se han realizado previamente investigaciones que consideren la influencia de la posible pérdida del segmento /-s/ en otros elementos presentes en el contexto comunicativo, como el sujeto pronominal. Al abordarlo desde la dualidad de la producción y percepción, podríamos obtener la imagen completa de la integración de factores fonéticos y sintácticos desde el uso y el entendimiento. Además del análisis acústico previo, nuestra intención es evaluar el reconocimiento perceptivo de los jueces sobre las formas verbales de 2ª y 3ª persona del singular con el propósito de comprobar si la presencia del sujeto pronominal puede desempeñar un papel desambiguador.

Además de las posibles ampliaciones mencionadas en cuanto a determinados aspectos de la variación lingüística del sujeto pronominal, somos conscientes de que todavía la dimensión social puede ser indagada con mayor profundidad en relación con esta variación sintáctica. Aunque se han abordado las variables sociales empleadas en la preestratificación de los corpus —grupo etario, sexo y nivel de instrucción de los informantes—, los hallazgos aquí presentados no agotan las múltiples posibilidades que podrían ser exploradas con el objetivo de ahondar en la perspectiva más social de la variación.

A partir de las múltiples variables de naturaleza social recogidas en los cuestionarios sociológicos se desprenden incontables líneas de investigación que nos permitirían explorar la influencia del perfil sociológico de los individuos. Por ejemplo, la incorporación de otros indicadores como la competencia lectora del informante, el interés por la cultura, la formación no necesariamente académica, la voluntad que impulsa al individuo a seguir aprendiendo o el entorno cultural que lo rodea pueden ser elementos relevantes para reconstruir el capital cultural del individuo (Bourdieu, 1984, 1986). Este planteamiento nos proporcionaría una visión más realista y precisa que el nivel de instrucción formal y, de este modo, se podría evaluar con mayor exactitud si la competencia gramatical y el bagaje cultural de los hablantes podría tener alguna repercusión en la variación de la ausencia o presencia del sujeto pronominal.

Asimismo, en trabajos posteriores puede ser clave hacer estudios de mayor calado poniendo en relación variables de corte social con otras de naturaleza lingüística. Esta combinación cumpliría con la perspectiva integradora de lo lingüístico y lo social que promueve la sociolingüística cognitiva. De este modo, podríamos observar si existen diferentes tendencias en la variación sintáctica que estén determinadas por las características sociales de los individuos, así como la percepción que estos tienen sobre la variación ausencia y presencia del sujeto pronominal, lo cual nos permitiría seguir reconstruyendo una gramática sociocognitiva.

Esta Tesis Doctoral, al estudiar el fenómeno de la variación en la expresión del sujeto pronominal desde la óptica de la sociolingüística cognitiva, ha puesto de manifiesto la importancia de considerar los usos lingüísticos que producen los hablantes dentro de su contexto comunicativo y también los procesos cognitivos que se desencadenan en la mente de los individuos ante la recepción del material lingüístico. Esta propuesta teórica y

metodológica que integra ambas perspectivas ha confirmado la necesidad de abordar fenómenos lingüísticos complejos desde una visión multidimensional e integradora que sea capaz de abarcar las diferentes aristas presentes en la naturaleza de la lengua. Por ello, el estudio de la lengua, además de observar el modo de comunicarnos, implica entender cómo construimos nuestra realidad social y cognitiva.

BIBLIOGRAFÍA

- Almor, A., & Nair, V. A. (2007). The form of referential expressions in discourse. *Language and Linguistics Compass*, 1(1-2), 84-99. <https://doi.org/10.1111/j.1749-818X.2007.00009.x>
- Alonso-Ovalle, L., Fernández-Solera, S., & Frazier, L. (2002). Null vs. Overt pronouns and the topic-focus articulation in Spanish. *Journal of Italian Linguistics*, 2(14), 151-169.
- Ariel, M. (1990). *Accessing Noun Phrase antecedents*. Routledge.
- Ariel, M. (2001). Accessibility theory: An overview. En T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 29-87). John Benjamins Publishing Company.
- Ávila Muñoz, A. M. (1998). *Elaboración, anotación y análisis del corpus oral del proyecto V.U.M. Léxico de frecuencia del español hablado en la ciudad de Málaga* [Tesis Doctoral, Universidad de Málaga]. <https://produccioncientifica.ugr.es/documentos/63930bb97a05941066fa8ba7>
- Ávila Muñoz, A. M. (2017). Entre limonetas y malaguitas: Estudio del léxico divergente

de proximidad. Nuevos datos del Proyecto CONVERLEX-Málaga. *Hispania*, 100(4), 538-553.

Ávila Muñoz, A. M., Lasarte, M., & Villena Ponsoda, J. A. (2008). *El español hablado en Málaga*. Sarriá.

Ávila Muñoz, A., Molina García, Á., Villena Ponsoda, J., & Sionakidis, G. (2024). *Las voces de Málaga. Reconstrucción histórica y archivo sonoro de la Ciudad Lingüística*. Comares.

Ávila-Jiménez, B. I. (1996). *Subject pronoun expression in Puerto Rican Spanish: A sociolinguistic, morphological, and discourse analysis* [Tesis Doctoral, Cornell University].

<https://www.proquest.com/openview/3500b742d0611367c8de2432994d5ed5/1?pq-origsite=gscholar&cbl=18750&diss=y>

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 748-769. <https://doi.org/10.1037/0278-7393.28.4.748>

Barrenechea, A. M., & Alonso, A. (1973). Los pronombres personales sujetos en el español hablado en Buenos Aires. *Studia iberica: Festschrift für Hans Flasche, 1973*, págs. 75-91, 75-91. <https://dialnet.unirioja.es/servlet/articulo?codigo=5496752>

Bayley, R., & Pease-Alvarez, L. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change*, 9(3), 349-371. <https://doi.org/10.1017/S0954394500001964>

Beckner, C., Ellis, N. C., Blythe, R., Holland, J., Bybee, J., Ke, J., Christiansen, M. H., Larsen-Freeman, D., Croft, W., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59(1), 1-26. <https://doi.org/10.1111/j.1467-9922.2009.00533.x>

Bentivoglio, P. (1980). *Why canto and not yo canto? The problem of first-person subject pronoun in spoken Venezuelan Spanish*. [Tesis de Máster]. Universidad de California.

Bentivoglio, P. (1987). *Los sujetos pronominales de primera persona en el habla de*

Caracas. Universidad Central de Venezuela.

Bentivoglio, P., Ortiz, L., & Silva-Corvalán, C. (2011). *La variable «expresión del sujeto pronominal» Guía de codificación PRESEEA*.

Blakemore, D. (2002). *Relevance and Linguistics Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge University Press.
<http://dx.doi.org/10.1017/CBO9780511486456>

Blanco Canales, A. (1998). *Estudio sociolingüístico de la ciudad de Alcalá de Henares* [Tesis Doctoral, Universidad de Alcalá].
<https://portalcientifico.uah.es/documentos/6142af3f27af2147d14441ff>

Blas Arroyo, J. L. (2005). *Sociolingüística del español: Desarrollos y perspectivas en el estudio de la lengua española en contexto social*. Cátedra.

Booth, R. W., & Weger, U. W. (2013). The function of regressions in reading: Backward eye movements allow rereading. *Memory & Cognition*, 41(1), 82-97.
<https://doi.org/10.3758/s13421-012-0244-y>

Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge University Press.

Bourdieu, P. (1984). Capital et marche linguistique. *Linguistische Berichte*, 90, 3-24.

Bourdieu, P. (1986). The forms of capital. En J. G. Richardson (Ed.), *Handbook of Theory and Research for the Sociology Education* (pp. 241-258). Greenwood Press.

Bybee, J. (2001). *Frequency and the emergence of linguistic structure*. John Benjamins Publishing Company.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Cameron, R. (1993). Ambiguous agreement, functional compensation, and nonspecific tú in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change*, 5(3), 305-334. <https://doi.org/10.1017/S0954394500001526>

Caravedo, R. (2014). *Percepción y variación lingüística: Enfoque sociocognitivo*. Iberoamericana.

Caravedo, R. (2018). Variación y cambio desde una perspectiva sociocognitiva. *Actas del X Congreso Internacional de Historia de la Lengua Española: Zaragoza, 7-11 de septiembre de 2015*, 1, 67-96.

- Casielles Suárez, E. (1996). ¿Es la interpretación arbitraria realmente arbitraria? *Revista Española de Lingüística*, 26(2), 359-378.
- Castellano Alemán, Á. (2001). Aspectos sociolingüísticos de la variación gramatical. En *Aspectos sociolingüísticos de la variación gramatical*. Servicio de Publicaciones y Producción Documental de la Universidad de Las Palmas de Gran Canaria.
- Chamorro, G. (2018). Offline interpretation of subject pronouns by native speakers of Spanish. *Glossa: A Journal of General Linguistics*, 3(1), 1-16. <https://doi.org/10.5334/gjgl.256>
- Chamorro, G., Sorace, A., & Sturt, P. (2016). What is the source of L1 attrition? The effect of recent L1 re-exposure on Spanish speakers under L1 attrition. *Bilingualism: Language and Cognition*, 19(3), 520-532. <https://doi.org/10.1017/S1366728915000152>
- Claes, J. (2011). ¿Constituyen las Antillas y el Caribe continental una sola zona dialectal?: Datos de la variable expresión del sujeto pronominal en San Juan de Puerto Rico y Barranquilla, Colombia. *Spanish in Context*, 8(2), 191-212. <https://doi.org/10.1075/sic.8.2.01cla>
- Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.
- Cristerna Román, D. (2020). *Implicit causality, pronominal form and anaphora resolution in Spanish* [Tesis de Máster, Universidad de Alberta]. <https://doi.org/10.7939/r3-ssag-4e36>
- Cruces Rodríguez, M. (en prensa). *Corpus Mujer Ayer y Hoy (MAYHO). Transformaciones sociales y lingüísticas en la ciudad de Málaga en el siglo XXI*.
- Cruz, A., & Loureda Lamas, Ó. (2019). Processing patterns of focusing in Spanish. En Ó. Loureda Lamas, I. Recio Fernández, L. Nadal, & A. Cruz (Eds.), *Empirical studies of the construction of discourse* (pp. 195-228). John Benjamins Publishing Company.
- Devís Márquez, P. P. (2012). Presencia y ausencia del sujeto con verbos en forma personal en español. *Zeitschrift für romanische Philologie*, 128(2), 286-318. <https://doi.org/10.1515/zrp-2012-0025>

- Duchowski, A. T. (2007). *Eye tracking methodology. Theory and practice* (2.^a ed.). Springer.
- Eckert, P. (1989). *Jocks and burnouts: Social categories and identity in the high school*. Teachers College Press.
- Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41(1), 87-100. <https://doi.org/10.1146/annurev-anthro-092611-145828>
- El Esbozo. (1973). *Esbozo de una nueva gramática de la lengua española*. Espasa-Calpe.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143-188. <https://doi.org/10.1017/S0272263102002024>
- Enríquez, E. V. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*. Editorial CSIC.
- Erker, D., & Guy, G. 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 Ramírez, S. (1987). *Gramática española: El pronombre: Vol. 3.2* (2.^a ed.). Arco Libros.
- Fernández Soriano, O. (1999). El pronombre personal. Formas y distribuciones. Pronombres átonos y tónicos. En V. Demonte Barreto & I. Bosque (Eds.), *Gramática descriptiva de la lengua española* (Vol. 1, pp. 1209-1274). Espasa Calpe España.
- Filiaci, F., Sorace, A., & Carreiras, M. (2014). Anaphoric biases of null and overt subjects in Italian and Spanish: A cross-linguistic comparison. *Language, Cognition and Neuroscience*, 29(7), 825-843. <https://doi.org/10.1080/01690965.2013.801502>
- Flores-Ferrán, N. (2004). Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact? *Language Variation and Change*, 16(1), 49-73. <https://doi.org/10.1017/S0954394504161048>
- Flores-Ferrán, N. (2007). A bend in the road: Subject personal pronoun expression in

Spanish after 30 years of sociolinguistic research. *Language and Linguistics Compass*, 1(6), 624-652. <https://doi.org/10.1111/j.1749-818X.2007.00031.x>

García, E. C. (1985). Shifting variation. *Lingua*, 67, 189-224.

García Salido, M. (2013). *La expresión pronominal de sujeto y objetos en español: Estudio con datos conversacionales*. Servicio de Publicaciones e Intercambio Científico. Universidad de Santiago de Compostela.

García-Miguel, J. M., Costas, L., & Martínez, S. (2003). Diátesis verbales y esquemas construccionales: Verbos, clases semánticas y esquemas sintáctico-semánticos en el proyecto ADESSE. *Actas del VI Congreso Internacional de Lingüística Hispánica*. VI Congreso Internacional de Lingüística Hispánica., Leipzig. <https://adesse.uvigo.es/>

Geeraerts, D. (2005). Lectal variation and empirical data in Cognitive Linguistics. En F. J. Ruiz de Mendoza & M. S. Peña (Eds.), *Cognitive linguistics. Internal dynamics and interdisciplinary interaction* (pp. 225-244.). Mouton/De Gruyter.

Gelormini Lezama, C., & Almor, A. (2011). Repeated names, overt pronouns, and null pronouns in Spanish. *Language and cognitive processes*, 26(3), 437-454. <https://doi.org/10.1080/01690965.2010.495234>

Gelormini-Lezama, C., & Almor, A. (2014). Singular and plural pronominal reference in Spanish. *Journal of Psycholinguistic Research*, 43, 299-313. <https://doi.org/10.1007/s10936-013-9254-6>

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, 32(2), 99-156. [https://doi.org/10.1016/0010-0277\(89\)90001-2](https://doi.org/10.1016/0010-0277(89)90001-2)

Gili Gaya, S. (1960). *Curso superior de sintaxis española* (7.^a ed.). Publicaciones y Ediciones Spes.

Giora, R. (1997). Discourse coherence and theory of relevance: Stumbling blocks in search of a unified theory. *Journal of Pragmatics*, 1(27), 17-34.

Givón, T. (1976). Topic, pronoun and grammatical agreement. En N. L. Charles (Ed.), *Subject and topic* (pp. 151-188). Academic Press.

Givón, T. (1984). *Syntax. A functional-typological introduction*. (Vol. 1). John Benjamins Publishing Company.

- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17(3), 311-347. https://doi.org/10.1207/s15516709cog1703_1
- Grant, M., Sloggett, S., & Dillon, B. (2020). Processing ambiguities in attachment and pronominal reference. *Glossa: A Journal of General Linguistics*, 5(1), 1-30. <https://doi.org/10.5334/gjgl.852>
- Grice, H. P. (1975). Logic and conversation. En P. Cole & J. L. Morgan (Eds.), *Syntax and Semantic. Speech Acts* (pp. 41-58). Academic Press.
- Guerrero González, S. (2017). La recolección de narraciones orales y su estudio en correlación con factores sociales: El valor de la entrevista sociolingüística. *Philologica Canariensia*, 23, 79-89. <https://doi.org/10.20420/PhilCan.2017.150>
- Guerrero González, S., & Manjón-Cabeza Cruz, A. (2022). *Guía PRESEEA de estudio de la expresión del sujeto pronominal*. <https://doi.org/10.37536/PRESEEA.2022.guia5>
- Guillén Jiménez, L. D. (2023). *Counter-argumentation and anaphoric substitution*. Peter Lang. <https://www.peterlang.com/document/1405828>
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2), 274. <https://doi.org/10.2307/416535>
- Hochberg, J. G. (1986). Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language*, 62(3), 609-621. <https://doi.org/10.1353/lan.1986.0041>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65-70.
- Holmquist, K., Nyström, M., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye-tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Hurtado, L. M. (2005). Condicionamientos sintáctico-semánticos de la expresión del sujeto en el español colombiano. *Hispania*, 88(2), 335-348. <https://doi.org/10.2307/20140954>
- Hyönä, J. (1995). An eye movement analysis of topic-shift effect during repeated reading.

Journal of Experimental Psychology. Learning, Memory, and Cognition, 21(5), 1365-1373. <https://doi.org/10.1037/0278-7393.21.5.a>

Jaeggli, O. A. (1986). Arbitrary plural pronominals. *Natural Language and Linguistic Theory*, 4(1), 43-76.

Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 3(1), 359-383. <https://doi.org/10.1111/j.1749-818X.2008.00108.x>

Johnson, D. E. (2016). *Rbrul versión 3.0* [Software]. <https://danielezrajohnson.shinyapps.io/Rbrul/>

Just, M. A., & Carpenter, P. A. (1980a). A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review*.

Just, M. A., & Carpenter, P. A. (1980b). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354. <https://doi.org/10.1037/0033-295X.87.4.329>

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, 37(1), 1-32. <https://doi.org/10.1017/S0272263114000187>

Keating, G. D., Jegerski, J., & Vanpatten, B. (2016). Online processing of subject pronouns in monolingual and heritage bilingual speakers of Mexican Spanish. *Bilingualism: Language and Cognition*, 19(1), 36-49. <https://doi.org/10.1017/S1366728914000418>

Khan, G. (1999). *A grammar of neo-aramaic: The dialect of the jews of Arbel*. Brill.

Kristiansen, G. (2001). *Social and linguistic stereotyping: A cognitive approach to accents*. 9, 129-145.

Kristiansen, G. (2003). *Referencia exofórica y estereotipos lingüísticos: Una aproximación sociocognitiva a la variación alofónica libre en el lenguaje natural* [Tesis Doctoral, Universidad Complutense de Madrid]. <http://webs.ucm.es/BUCM/tesis/fll/ucm-t27033.pdf>

Kristiansen, G. (2008). A sociocognitive approach to lectal variation. En F. J. Ruiz de Mendoza & M. S. Peña (Eds.), *Cognitive linguistics. Internal dynamics and*

interdisciplinary interaction (pp. 45-88). Mouton/De Gruyter.

Kristiansen, G., & Dirven, R. (2008). *Cognitive sociolinguistics. Language variation, cultural models, social systems*. Mouton/De Gruyter.

Labov, W. (1963). The Social Motivation of a Sound Change. *WORD*, 19(3), 273-309. <https://doi.org/10.1080/00437956.1963.11659799>

Labov, W. (1972). *Sociolinguistic patterns* (Vol. 4). University of Pennsylvania Press.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford University Press.

Lapidus Shin, N., & Otheguy, R. (2005). Overt nonspecific ellos in Spanish in New York. *Spanish in Context*, 2(2), 157-174. <https://doi.org/10.1075/sic.2.2.03lap>

Lasarte Cervantes, M. de la C., Sánchez Sáez, J. M., Ávila Muñoz, A. M., & Villena Ponsoda, J. A. (2009). *El español hablado en Málaga*. Sarriá.

Lastra, Y., & Martín Butragueño, P. (2015). Subject pronoun expression in oral Mexican Spanish. En *Subject pronoun expression in Spanish. A Cross-Dialectal Perspective* (pp. 39-57). Georgetown University Press.

Lavandera, B. R. (1978). Where does the sociolinguistic variable stop? *Language in Society*, 7(2), 171-182. <https://doi.org/10.1017/S0047404500005510>

Levshina, N. (2015). *How to do linguistics with R. Data exploration and statistical analysis*. John Benjamins Publishing Company.

Li, C., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. University of California Press.

López Morales, H. (2009). *El estudios de la variación lingüística* (M. V. Camacho Taboada, J. J. Rodríguez Toro, & J. Santana Marrero, Eds.; pp. 9-34). Vervuert Verlagsgesellschaft. <https://doi.org/10.31819/9783964566126-002>

Loureda Lamas, Ó., Recio Fernández, I., & Rudka, M. (2021). *Comunicación, partículas discursivas y pragmática experimental*. Arco/Libros.

Loureda, Ó., Recio Fernández, I., Cruz, A., & Nadal, L. (2019). Introduction. En Ó. Loureda, I. Recio Fernández, L. Nadal, & A. Cruz (Eds.), *Pragmatics & Beyond New Series* (Vol. 305, pp. 1-13). John Benjamins Publishing Company.

<https://doi.org/10.1075/pbns.305.int>

- Luján, M. (1999). Expresión y omisión del pronombre personal. En V. Demonte Barreto & I. Bosque (Eds.), *Gramática descriptiva de la lengua española. Sintaxis básica de las clases de palabras* (Vol. 1, pp. 1275-1316). Espasa Calpe España. <https://dialnet.unirioja.es/servlet/articulo?codigo=2152399>
- MacDonald, M. C., & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language*, 29(4), 469-492. [https://doi.org/10.1016/0749-596X\(90\)90067-A](https://doi.org/10.1016/0749-596X(90)90067-A)
- Manjón-Cabeza Cruz, A. (2022). Dos factores lingüísticos en pugna en la explicación de la tasa pronominal: Ambigüedad y tiempo verbal. *Lingüística en la Red*, XIX, 1-12. <https://doi.org/10.37536/linred.2022.XIX.1871>
- Manjón-Cabeza Cruz, A., & Granados Romero, I. (2022). Similitudes y diferencias en la expresión del sujeto pronominal en una misma comunidad de habla: Málaga frente a Granada. *Spanish in Context*, 19(2), 217-243. <https://doi.org/10.1075/sic.19034.man>
- Manjón-Cabeza Cruz, A., Pose Furest, F., & Sánchez García, F. J. (2016). Factores determinantes en la expresión del sujeto pronominal en el corpus PRESEEA de Granada. *Boletín de Filología*, 51(2), 181-207. <https://doi.org/10.4067/S0718-93032016000200007>
- Martín Butragueño, P. (1994). Hacia una tipología de la variación gramatical en sociolingüística del español. *Nueva Revista de Filología Hispánica (NRFH)*, 42(1), 29-75. <https://doi.org/10.24201/nrfh.v42i1.1822>
- Martín Butragueño, P. (2020a). 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. <https://doi.org/10.1075/sic.00060.mar>
- Martín Butragueño, P. (2020b). La expresión del sujeto pronominal en la Ciudad de México: Explorando la variación lingüística con efectos estadísticos fijos y con efectos mixtos. En A. E. Pérez Barajas & A. Hernández (Eds.), *Propuestas metodológicas para el trabajo y la investigación lingüística. Aplicaciones teóricas y descriptivas* (pp. 639-684). Universidad de Colima.

- Martín Morales, M. (2018). Factores lingüísticos y sociales en el proceso de debilitamiento de /-n/ final de palabra y el uso de formas vulgares en la ciudad de Málaga. *ELUA*, 32, 237-262. <https://doi.org/10.14198/ELUA2018.32.11>
- Martínez Lara, J. A. (2013). Estudio de la expresión variable del sujeto de 3ª persona en el habla de Caracas. Un análisis cuantitativo. *Núcleo*, 30, 69-96.
- Martínez Lara, J. A. (2015). Estudio diacrónico del sujeto variable de 3ª persona en el habla de Caracas. *Boletín de Lingüística*, XVII(43-44), 48-62.
- Martínez Lara, J. A. (2016). *Estudio sociolingüístico del sujeto variable de 3ª persona singular en los dialectos de Caracas, Medellín y Montevideo*. [Tesis de Máster]. Universidad Central de Venezuela.
- Martínez Lara, J. A. (2022). Sujeto variable de 3ª persona singular en el habla de Montevideo. *Onomázein Revista de lingüística filología y traducción*, 57, 215-240. <https://doi.org/10.7764/onomazein.57.10>
- Martínez Lara, J. A., Guerrero, S., & González Vergara, C. (2021). (Yo) opino: La variación del sujeto pronominal en el habla de Santiago de Chile. En B. M. A. Rogers & M. Figueroa Candia (Eds.), *Lingüística del castellano chileno: Estudios sobre variación, innovación, contacto e identidad* (pp. 283-308). Vernon Press.
- McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, 28(10), 1107-1118. [https://doi.org/10.1016/0042-6989\(88\)90137-X](https://doi.org/10.1016/0042-6989(88)90137-X)
- Milroy, J. (1992). *Linguistic variation and change: On the historical sociolinguistics of English*. Blackwell. <http://archive.org/details/linguisticvariat0000milr>
- Moeschler, J. (1989). Pragmatic connectives, argumentative coherence and relevance. *Argumentation*, 3(3), 321-339. <https://doi.org/10.1007/BF00128944>
- Morales, A. (1997). La hipótesis funcional y la aparición de sujeto no nominal: El español de Puerto Rico. *Hispania*, 80(1), 153. <https://doi.org/10.2307/345995>
- Moreno Fernández, F. (2005). Corpus para el estudio del español en su variación geográfica y social. El corpus PRESEEA. *Oralia: Análisis del Discurso Oral*, 8, 123-139. <https://doi.org/10.25115/oralia.v8i1.8333>
- Moreno Fernández, F. (2009). *Principios de sociolingüística y sociología del lenguaje* (4.ª

ed.). Ariel.

Moreno Fernández, F. (2012). *Sociolingüística cognitiva: Propositiones, escolios y debates*. Iberoamericana Vervuert.

Nadal, L., Recio Fernández, I., & Loureda Lamas, Ó. (2016). El significado procedimental y las partículas discursivas del español: Una aproximación experimental. *Revista Signos. Estudios de Lingüística*, 49, 52-77.

Nadal, L., Recio Fernández, I., Rudka, M., & Loureda Lamas, Ó. (2017). Processing additivity in Spanish: Incluso vs. Además. En A.-M. De Cesare & C. Andorno (Eds.), *Focus on additivity: Adverbial modifiers in Romance, Germanic and Slavic languages* (pp. 137-154). John Benjamins Publishing Company.

Newman, I., & Benz, C. R. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Educational Leadership Faculty Publication.

NGLE. (2025). *Nueva gramática de la lengua española*. <https://www.rae.es/gramática/>

O'Regan, J. K. (1990). Eye movements and reading. *Reviews of Oculomotor Research*, 4, 395-453.

O'Regan, J. K., & Jacobs, A. M. (1992). Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1), 185-197. <https://doi.org/10.1037/0096-1523.18.1.185>

O'Regan, J. K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, 23(8), 765-768. [https://doi.org/10.1016/0042-6989\(83\)90198-0](https://doi.org/10.1016/0042-6989(83)90198-0)

Orozco, R., & Guy, G. R. (2008). El uso variable de los pronombres sujetos: ¿qué pasa en la costa Caribe colombiana? En M. Westmoreland & J. A. Thomas (Eds.), *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics* (pp. 70-80). Cascadilla Proceedings Project.

Orozco, R., & Hurtado, L. M. (2020). A Variationist Study of Subject Pronoun Expression in Medellín, Colombia. *Languages*, 6(1), 5. <https://doi.org/10.3390/languages6010005>

Orozco, R., & Hurtado, L. M. (2021). Variable subject pronoun expression revisited: This

is what the Paisas do. *Proceedings of the Linguistic Society of America*, 6(1), 713.
<https://doi.org/10.3765/plsa.v6i1.5006>

Orozco, R., Marmolejo, M., & Grajales, R. (2024). Estudio sociolingüístico de la expresión pronominal en Cali. *Lingüística*, 40, 1-28. <https://doi.org/10.5935/2079-312X.20240006>

Orozco, R., & Orozco, J. (2022). The effect of the verb on pronominal expression: A reanalysis. *Proceedings of the Linguistic Society of America*, 7(1), 1-15.
<https://doi.org/10.3765/plsa.v7i1.5286>

Paredes Silva, V. L. (1993). Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change*, 5(1), 35-49.
<https://doi.org/10.1017/S0954394500001381>

Pérez Brabandere, V. (2010). *Los sujetos pronominales de primera persona en el español de Caracas 2004-2010* [Tesis de máster]. Universidad Central de Venezuela.

Pérez Córdoba, A. L., & Gomes Camacho, R. (2019). A expressão do pronome pessoal sujeito no espanhol falado no Caribe colombiano. *Estudos Linguísticos*, 48(1), 404-424. <https://doi.org/10.21165/el.v48i1.2148>

Pizarro Pedraza, A. (2016). Variación semántica y significado social: Hacia una sociolingüística cognitiva de la Tercera Ola. *Dicenda. Estudios de lengua y literatura españolas*, 34, 311-338. <https://doi.org/10.5209/DICE.53562>

Pollatsek, A., & Rayner, K. (1990). Eye movements and lexical access in reading. En D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading*. Routledge.

Prieto Muñoz, S. Y. P. (2011). Los pronombres personales sujetos: Muestra general de tres ciudades. *Cuadernos de Lingüística Hispánica*, 17, 53-66.

Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, 110(2), 81-94. <https://doi.org/10.1016/j.bandl.2008.12.001>

RAE. (2024, enero 19). *Ortografía de la lengua española*. <https://www.rae.es/ortografia>

RAE, & ASELE. (2009). *Nueva gramática de la lengua española*. Espasa-Calpe.

- Ramajo Ballester, Á. (2022). *Autolingua* [Software].
<https://github.com/ramajoballester/autolingua>
- Ranson, D. L. (1991). Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change*, 3(2), 133-152.
<https://doi.org/10.1017/S0954394500000491>
- Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, 8(1), 21-30. <https://doi.org/10.1068/p080021>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 2(62), 1457-1506.
<https://doi.org/10.1080/17470210902816461>
- Recio Fernández, I. (2019). *The impact of procedural meaning on second language processing: A study on connectives* [Tesis Doctoral, Universität Heidelberg].
https://archiv.ub.uni-heidelberg.de/volltextserver/28641/1/RecioFernandez_Dissertation.pdf
- Recio Fernández, I., Loureda, Ó., & Sanders, T. J. M. (2021). Constructing discourse: An experimental approach. *Revista Signos*, 54(107), 1004-1025.
<https://doi.org/10.4067/S0718-09342021000301004>
- Recio Fernández, I., Nadal, L., & Loureda Lamas, Ó. (2018). On argumentative relations in Spanish: Experimental evidence on the grammaticalization of cause-consequence discourse markers. En S. Pons Bordería & Ó. Loureda Lamas (Eds.), *Beyond grammaticalization and discourse markers* (Vol. 18, pp. 384-409). Brill.
- Repede, D. (2019). Condicionantes sociolingüísticos de la expresión del sujeto pronominal en el sociolecto alto del corpus PRESEEA-Sevilla. *Revista de Investigación Lingüística*, 22, 397-423. <https://doi.org/10.6018/ril.390031>
- Romaine, S. (1981). The status of variable rules in sociolinguistic theory. *Journal of Linguistics*, 17(1), 93-119.
- Roselló Verdeguer, J. (2021). Factores determinantes en la expresión del sujeto

pronominal en el español hablado de Valencia. *Linred: Lingüística en la Red*, 18(18), 1-24. <https://doi.org/10.37536/LINRED.2021.XVIII.2>

Rosengren, P. (1974). *Presencia y ausencia de los pronombres personales sujetos en español moderno*. Almqvist & Wiksell.

Salvá, V. (1988). *Gramática de la lengua castellana según ahora se habla*. Arco Libros.

Sankoff, D. (1988). Sociolinguistics and syntactic variation. En F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey*: (Vol. 4, pp. 140-161). Cambridge University Press.

Scrivner, O., & Dí-az Campos, M. (2016). Language Variation Suite: A theoretical and methodological contribution for linguistic data analysis. *Proceedings of the Linguistic Society of America*, 1, 29:1-15. <https://doi.org/10.3765/plsa.v1i0.3734>

Serrano, M. J. (2007). Historia que ya es historia: Evolución y actualidad del concepto y la metodología de la variación sintáctica. *Boletín de Lingüística*. <https://www.redalyc.org/articulo.oa?id=34702806>

Silva-Corvalán, C. (1982). Subject expression and placement in Mexican-American Spanish. En J. Amastae & L. Elías-Olivares (Eds.), *Spanish in the United States: Sociolinguistic aspects* (pp. 93-120). Cambridge University Press.

Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español*. Georgetown University.

Silva-Corvalán, C. (2003). Otra mirada a la expresión del sujeto como variable sintáctica. En *Lengua, variación y contexto: Estudios dedicados a Humberto López Morales* (Vol. 2, pp. 849-860). Arco/Libros.

Silva-Corvalán, C., & Enrique Arias, A. (2017). *Sociolingüística y pragmática del español* (2.^a ed.). Georgetown University.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2.^a ed.). Blackwell.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. En M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 327-342). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198568971.013.0019>

- Tagliamonte, S. A. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Wiley-Blackwell.
- Travis, C. E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, 19(2). <https://doi.org/10.1017/S0954394507070081>
- Travis, C. E., & Torres Cacoullos, R. (2012). What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics*, 23(4), 711-748. <https://doi.org/10.1515/cog-2012-0022>
- van Hoek, K. (2007). Pronominal anaphora. En D. Geeraerts & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics* (pp. 890-915). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199738632.013.0034>
- Vida Castro, M. (2004). *Estudio sociofonológico del español hablado en la ciudad de Málaga*. Servicio de Publicaciones. <https://dialnet.unirioja.es/servlet/libro?codigo=322461>
- Vida Castro, M. (2007). *El español hablado en Málaga*. Sarriá.
- Vida Castro, M. (en prensa). *Corpus oral complementario del español hablado en Málaga*. Publicación del Grupo Vernáculo Urbano Malagueño.
- Vida Castro, M., & Villena Ponsoda, J. A. (2012). La influencia del prestigio social en la reversión de los cambios fonológicos. Constricciones universales sobre la variación en el español ibérico meridional. Un caso de nivelación dialectal. En *Estudios sobre el español de Málaga. Pronunciación, vocabulario y sintaxis*. (pp. 67-127). Sarriá.
- Villena Ponsoda, J. A. (2014). Estatus, red e individuo. Fundamentos del análisis escalonado de la variación lingüística. Elisión de /d/ en el español de Málaga. *Actas del X Congreso de Lingüística General*, 953-969.
- Villena Ponsoda, J. A., Ávila Muñoz, A. M., & Von Essen, M. C. (2017). *Efecto de la estratificación, la red social y las variables de pequeña escala en la variación léxica. Proyecto de investigación sobre la convergencia del léxico dialectal en la ciudad de Málaga (CONVERLEX)*. Venice University Press. <https://riuma.uma.es/xmlui/handle/10630/31309>
- Villena Ponsoda, J. A., & Vida Castro, M. (2017a). Horizontal and vertical convergence

and divergence of dialects in Southern Spain: Between local and standard varieties. En I. Buchstaller & B. Siebenhaar (Eds.), *Language Variation—European Perspectives VI: Selected papers from the Eighth International Conference on Language Variation in Europe (ICLaVE 8), Leipzig, May 2015* (pp. 125-140). John Benjamins Publishing Company. <https://doi.org/10.1075/silv.19.08vil>

Villena Ponsoda, J. A., & Vida Castro, M. (2017b). Variación, identidad y coherencia en el español meridional. Sobre la indexicalidad de las variables convergentes del español de Málaga. *Linred: Lingüística en la Red*, 15. <https://ebuah.uah.es/dspace/handle/10017/34167>

Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition*, 111(1), 132-137. <https://doi.org/10.1016/j.cognition.2008.12.011>

Wilson, D., & Sperber, D. (2004). La teoría de la relevancia. *Revista de Investigación Lingüística*, 7, 233-283.

Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge University Press.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2.^a ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>

ANEXO I. GUÍA DE CODIFICACIÓN EXPRESIÓN DEL SUJETO PRONOMINAL EN CORPUS PRESEEA-MÁLAGA

Datos del informante y de la entrevista

- Código del informante
- Minuto de grabación más próximo señalado en la transliteración

Información sobre la cláusula de análisis

- Contexto lingüístico en el que aparece el ejemplo analizado
- Forma verbal analizada
- Infinitivo de la forma verbal analizada

Variables de control

- Contextos obligatorios: presencia obligatoria, ausencia obligatoria, variación
- Observaciones

Variable dependiente

1. Sujeto pronominal

- No expresado
- Expresado

Variables independientes lingüísticas

2. Persona del sujeto

- 1ª persona del singular (*yo*)
- 2ª persona del singular (*tu*)
- 3ª persona del singular (*él/ella*)
- 1ª persona del plural (*nosotros*)
- 2ª persona del plural (*vosotros/ustedes* con valor informal)
- 3ª persona del plural (*ellos/ellas*)
- 2ª persona del singular (*usted*)
- 2ª persona del plural (*ustedes* con valor formal)

3. Referente del sujeto

- Humano
- Animado (*no humano*)
- Inanimado (*cosas*)
- Neutros (*oración, ello, eso*)

4. Ambigüedad fonética potencial

- Sí (*Tú canta(s)/ella canta/ellos canta(n)*)
- No (*Nosotros cantamos/ Vosotros cantáis*)

5. Especificidad del sujeto

- No específico
- Específico

6. Modo

- Indicativo
- Subjuntivo
- Imperativo

7. Tiempo verbal

- Presente
- Pretérito
- Imperfecto
- Futuro morfológico
- Futuro perifrástico (*ir a + inf*)
- Condicional
- Pretérito perfecto compuesto
- Pretérito pluscuamperfecto
- Pretérito anterior
- Futuro compuesto
- Condicional compuesto

8. Progresividad

- No progresivo

- Progresivo

9. Perfectividad

- Tiempo simple
- Tiempo compuesto

10. Ambigüedad morfológica

- Forma verbal no ambigua
- Forma verbal morfológicamente ambigua

11. Clase semántica verbos según ADESSE

- Verbos de procesos mentales
- Verbos de procesos relacionales
- Verbos de procesos materiales
- Verbos de procesos verbales
- Verbos de procesos existenciales
- Verbos de modulación

12. Subtipos de verbos según ADESSE

(1-4 Verbos de procesos mentales)

- Sensación
- Percepción
- Cognición
- Elección

(5-6 Verbos de procesos relacionales)

- Atribución
- Posesión

(7-10 Verbos de procesos materiales)

- Espacio
- Cambio
- Otros hechos
- Comportamiento

(11-13 Verbos de proceso verbal)

- Comunicación
- Valoración
- Emisión de sonido

(14-16 Verbos de procesos existenciales)

- Existencia
- Fase-Tiempo
- Vida

(17-20 Verbos de modulación)

- Causación
- Disposición
- Aceptación
- Verbo de apoyo

13. Tipo de cláusula

- Afirmativa
- Negativa
- Interrogativa con partícula interrogativa
- Preguntas sin partícula interrogativa
- Orden/imperativa afirmativa
- Orden/imperativo negativo
- Exclamativa con partícula

14. Correferencialidad

- Correferencial con sujeto previo
- Correferencial con cualquier argumento que no sea sujeto
- No correferencial
- No aplicable

15. Turno de habla

- Mismo turno de habla
- Cambio de turno

16. Perífrasis verbales

- Forma no perifrástica
- Modales
- Aspectuales

17. Construcción pronominal

- Verbo no pronominal
- Verbo pronominal

18. Tipo de discurso

- Diálogo rápido
- Descripción
- Narración
- Argumentación

Variables independientes sociales

19. Edad

- Grupo etario 1 (20 a 34 años)
- Grupo etario 2 (35 a 55 años)
- Grupo etario 3 (<55 años)

20. Sexo

- Hombre
- Mujer

21. Nivel de estudios

- Nivel bajo
- Nivel medio
- Nivel alto

22. Corpus

- Corpus 1995
- Corpus 2015

ANEXO II. LISTADO DE LOS EFECTOS ALEATORIOS

Efectos aleatorios: verbos con un peso positivo para el corpus 1995

Variable	<i>log-odds</i>	Casos	P/P+A	Peso centrado
Creer	0.796	40	0.675	0.689
Salir	0.462	35	0.343	0.613
Nacer	0.410	10	0.500	0.601
Acordarse	0.378	49	0.429	0.593
Ir	0.312	151	0.258	0.577
Pensar	0.284	16	0.438	0.570
Vivir	0.282	36	0.306	0.570
Ser	0.270	182	0.209	0.567
Trabajar	0.262	36	0.278	0.565
Proponerse	0.236	2	1.000	0.558
Comprar	0.234	7	0.429	0.558
Hojear	0.233	2	1.000	0.558
Estudiar	0.228	26	0.231	0.556
Estar	0.221	213	0.225	0.555
Descubrir	0.217	3	1.000	0.554
Preparar	0.211	4	0.500	0.552
Hacer	0.208	142	0.190	0.551
Enterarse	0.198	5	0.600	0.549
Notar	0.194	4	0.750	0.548
Tener que trabajar	0.189	4	0.750	0.547
Adaptarse	0.177	2	1.000	0.544
Pagar	0.165	17	0.118	0.541
Llamar	0.164	12	0.167	0.541
Tener	0.161	303	0.251	0.540
Hacerse	0.160	7	0.429	0.540
Subir	0.160	4	0.500	0.540
Llamarse	0.159	6	0.333	0.539
Dar	0.156	55	0.164	0.539
Olvidar	0.154	1	1.000	0.538
Valer	0.152	2	1.000	0.538
Guiar	0.151	1	1.000	0.537

Irse a traer	0.151	1	1.000	0.537
Irse retirando	0.150	1	1.000	0.537
Poder decir	0.150	4	0.500	0.537
Querer	0.149	63	0.254	0.537
Depender	0.146	1	1.000	0.536
Recoger	0.146	1	1.000	0.536
Invitar	0.145	1	1.000	0.536
Cortar	0.144	2	0.500	0.536
Ir a liar	0.143	1	1.000	0.535
Tomar	0.143	2	0.500	0.535
Liarse	0.142	2	0.500	0.535
Recordar	0.142	12	0.417	0.535
Opinar	0.139	1	1.000	0.534
Marcharse	0.138	1	1.000	0.534
Montar	0.137	1	1.000	0.534
Ir a venir	0.133	1	1.000	0.533
Poder ofrecer	0.133	1	1.000	0.533
Irse alejando	0.132	1	1.000	0.533
Levantar	0.131	1	1.000	0.532
Seguir dibujando	0.129	1	1.000	0.532
Deber de dejar	0.127	1	1.000	0.531
Conocer	0.125	41	0.317	0.531
Cumplimentar	0.125	1	1.000	0.531
Escoger	0.125	3	0.667	0.531
Tener que molestar	0.125	1	1.000	0.531
Echar de menos	0.123	5	0.400	0.530
Llevar	0.121	31	0.226	0.530
Cometer	0.120	1	1.000	0.530
Darse cuenta	0.115	2	0.500	0.528
Defender	0.114	1	1.000	0.528
Escribir	0.114	2	0.500	0.528
Tener que callar	0.114	1	1.000	0.528
Caer	0.113	1	1.000	0.528
Escuchar	0.113	6	0.333	0.528
Tener que ordenar	0.111	1	1.000	0.527
Mudarse	0.110	5	0.200	0.527

Aguantar	0.108	2	0.500	0.527
Comprarse	0.107	1	1.000	0.526
Fumar	0.107	4	0.250	0.526
Colaborar	0.106	2	0.500	0.526
Relacionarse	0.106	2	0.500	0.526
Tener que acudir	0.106	1	1.000	0.526
Ir evolucionando	0.104	1	1.000	0.526
Tratar	0.103	4	0.250	0.525
Seguir trabajando	0.102	1	1.000	0.525
Ir recogiendo	0.099	3	0.333	0.524
Pelear	0.099	7	0.143	0.524
Plantearse	0.099	2	0.500	0.524
Elegir	0.095	1	1.000	0.523
Copiar	0.094	1	1.000	0.523
Conservar	0.093	2	0.500	0.523
Dejarse engañar	0.091	1	1.000	0.522
Preocuparse	0.091	3	0.333	0.522
Criar	0.086	1	1.000	0.521
Hartarse	0.086	1	1.000	0.521
Necesitar	0.085	4	0.250	0.521
Aferrarse	0.084	1	1.000	0.521
Mariposear	0.084	1	1.000	0.521
Sacar	0.084	14	0.214	0.521
Dejarse	0.083	2	0.500	0.520
Identificarse	0.082	2	0.500	0.520
Intervenir	0.082	1	1.000	0.520
Poder ampararse	0.082	1	1.000	0.520
Ir a representar	0.080	1	1.000	0.520
Criarse	0.078	1	1.000	0.519
Injertar	0.078	1	1.000	0.519
Faltar	0.070	3	0.333	0.517
Acostarse	0.069	1	1.000	0.517
Poder perjudicar	0.067	1	1.000	0.516
Considerarse	0.065	1	1.000	0.516
Deber estudiar	0.064	1	1.000	0.516
Fijarse	0.064	6	0.167	0.516
Ir a dar	0.064	2	0.500	0.516

Coger	0.060	25	0.200	0.515
Inventar	0.060	3	0.333	0.515
Ir dando	0.060	3	0.333	0.515
Bajar	0.059	3	0.333	0.514
Ayudar	0.058	13	0.231	0.514
Vulnerar	0.058	1	1.000	0.514
Tener que decir	0.054	2	0.500	0.513
Correr	0.053	7	0.286	0.513
Matar	0.053	3	0.333	0.513
Venirse	0.053	25	0.160	0.513
Comer	0.051	6	0.333	0.512
Meter	0.048	12	0.167	0.512
Rezar	0.048	2	0.500	0.512
Cerrar	0.046	2	0.500	0.511
Poder tener	0.046	4	0.250	0.511
Tener que comer	0.044	5	0.200	0.511
Cuidar	0.043	3	0.333	0.510
Tirar	0.034	3	0.333	0.508
Repetir	0.026	6	0.333	0.506
Jugar	0.025	4	0.250	0.506
Traer	0.020	8	0.125	0.505
Encontrarse	0.016	4	0.250	0.504
Tener que salir	0.013	5	0.200	0.503
Oír	0.012	6	0.167	0.503
Traerse	0.009	3	0.333	0.502
Quedarse	0.007	25	0.120	0.501
Imaginar	0.005	14	0.286	0.501
Admirar	0.000	2	0.500	0.500

Efectos aleatorios: informantes con un peso positivo para el corpus 1995

Variable	<i>log-odds</i>	Casos	P/P+A	Peso centrado
MA_M21_008	1.098	43	0.372	0.745
MA_H13_715	1.057	48	0.354	0.737
MA_H22_731	1.050	84	0.464	0.736
MA_M21_069	0.917	28	0.321	0.709
MA_H32_705	0.807	105	0.295	0.686

MA_M21_906	0.785	32	0.312	0.681
MA_M32_722	0.700	65	0.308	0.662
MA_M22_709	0.639	63	0.302	0.649
MA_H32_214	0.629	40	0.350	0.646
MA_M12_729	0.614	80	0.300	0.643
MA_M11_023	0.485	15	0.467	0.613
MA_M32_717	0.438	52	0.269	0.602
MA_H12_707	0.396	79	0.342	0.591
MA_M31_002	0.343	14	0.214	0.579
MA_M13_046	0.340	53	0.340	0.578
MA_H31_108	0.272	16	0.250	0.561
MA_M33_723	0.265	72	0.250	0.559
MA_H22_730	0.244	48	0.271	0.554
MA_M31_004	0.235	25	0.160	0.552
MA_M31_077	0.217	76	0.211	0.548
MA_H11_114	0.216	60	0.167	0.547
MA_M11_075	0.206	23	0.174	0.545
MA_H21_734	0.180	51	0.294	0.538
MA_H33_725	0.175	39	0.282	0.537
MA_M33_713	0.123	75	0.200	0.524
MA_M33_706	0.111	54	0.259	0.521
MA_H21_905	0.107	68	0.206	0.520
MA_H11_916	0.093	31	0.226	0.517
MA_H22_012	0.085	65	0.246	0.515
MA_M22_013	0.082	5	0.400	0.514
MA_M12_017	0.076	56	0.250	0.512
MA_H23_061	0.037	13	0.231	0.503
MA_H22_014	0.034	57	0.228	0.502

Efectos aleatorios: verbos con un peso positivo para el corpus 2015

Variable	<i>log-odds</i>	Casos	P/P+A	Peso centrado
Creer	0.840	100	0.740	0.698
Vivir	0.546	26	0.615	0.633
Pensar	0.451	28	0.714	0.611
Estar	0.408	145	0.310	0.600
Fijarse	0.299	9	0.444	0.574

Montar	0.288	4	0.750	0.571
Apuntarse	0.283	4	0.750	0.570
Entrar	0.245	24	0.333	0.561
Poder	0.223	47	0.213	0.555
Pagar	0.218	17	0.235	0.554
Recordar	0.211	13	0.615	0.552
Seguir	0.21	7	0.429	0.552
Ponerse	0.193	9	0.333	0.548
Visitar	0.173	2	1	0.543
Hablar	0.17	26	0.231	0.542
Meterse	0.17	4	0.5	0.542
Poder meter	0.164	3	0.667	0.541
Sacar	0.151	12	0.25	0.537
Admirar	0.15	2	1	0.537
Sembrar	0.15	1	1	0.537
Permitir	0.145	2	0.5	0.536
Pasearse	0.144	2	1	0.536
Salvar	0.144	1	1	0.536
Ir a decir	0.143	2	1	0.535
Beber	0.142	3	0.667	0.535
Huir	0.138	1	1	0.534
Enviudar	0.136	1	1	0.534
Encargarse	0.134	1	1	0.533
Ir comprar	0.134	1	1	0.533
Ir a endiñar	0.133	1	1	0.533
Ir	0.13	161	0.23	0.532
Vender	0.13	7	0.286	0.532
Desear	0.129	1	1	0.532
Sentirse	0.128	3	0.667	0.532
Reservar	0.127	1	1	0.531
Subirse	0.126	1	1	0.531
Rendirse	0.125	1	1	0.531
Abrir	0.124	2	0.5	0.531
Querer ganar	0.124	1	1	0.531
Percibir	0.123	1	1	0.53
Perder	0.123	1	1	0.53
Matarse	0.12	1	1	0.53

Retocar	0.12	1	1	0.53
Ir a juntar	0.119	1	1	0.529
Entrevistar	0.118	1	1	0.529
Comerciar	0.117	2	0.5	0.529
Seguir viendo	0.116	1	1	0.529
Estudiar	0.115	9	0.333	0.528
Preferir	0.115	4	0.75	0.528
Tener que llevar	0.115	1	1	0.528
Asomarse	0.114	1	1	0.528
Ver	0.114	3	0.667	0.528
Entretenerse	0.11	1	1	0.527
Marcharse	0.11	2	0.5	0.527
Reunirse	0.11	2	0.5	0.527
Sumar	0.11	1	1	0.527
Acordar	0.109	1	1	0.527
Tener que acostarse	0.109	1	1	0.527
Matricularse	0.108	1	1	0.527
Calcular	0.107	1	1	0.526
Calentar	0.104	1	1	0.526
Salir	0.104	19	0.368	0.526
Proceder	0.103	1	1	0.525
Elegir	0.102	2	0.5	0.525
Seguir trabajando	0.1	1	1	0.525
Acompañar	0.098	1	1	0.524
Pasarse	0.095	1	1	0.523
Fallecer	0.094	1	1	0.523
Intentar	0.092	5	0.2	0.523
Mirar	0.092	9	0.222	0.523
Deber	0.09	4	0.25	0.522
Estrenar	0.086	2	0.5	0.521
Poder llegar	0.086	1	1	0.521
Llamar	0.081	4	0.25	0.52
Ir a arbitrar	0.08	1	1	0.52
Incluir	0.078	1	1	0.519
Poder haber	0.076	1	1	0.519
Conocer	0.075	41	0.317	0.518

Ser	0.075	147	0.259	0.518
Acceder	0.074	1	1	0.518
Enterarse	0.074	3	0.667	0.518
Aparcar	0.073	2	0.5	0.518
Conducir	0.073	1	1	0.518
Seguir estando	0.073	1	1	0.518
Tratar	0.068	5	0.4	0.517
Compararse	0.063	1	1	0.515
Poder decir	0.063	2	0.5	0.515
Irse	0.062	50	0.3	0.515
Pasar	0.062	14	0.286	0.515
Correr	0.061	1	1	0.515
Crear	0.057	3	0.333	0.514
Apretar	0.054	1	1	0.513
Chupar	0.054	1	1	0.513
Probar	0.054	1	1	0.513
Conocerse	0.053	3	0.333	0.513
Decir	0.053	177	0.186	0.513
Destacar	0.046	1	1	0.511
Llevar trabajando	0.044	1	1	0.511
Tirarse	0.044	4	0.5	0.511
Presentarse	0.042	2	0.5	0.51
Sacarse	0.042	3	0.333	0.51
Opinar	0.041	2	0.5	0.51
Morir	0.04	2	0.5	0.51
Tener que pagar	0.04	2	0.5	0.51
Preguntar	0.035	4	0.25	0.508
Tirar	0.034	6	0.167	0.508
Comer	0.032	8	0.375	0.508
Llevarse	0.032	4	0.25	0.508
Quedar	0.031	2	0.5	0.507
Llamarse	0.028	10	0.2	0.507
Imaginar	0.02	9	0.333	0.505
Oír	0.013	6	0.333	0.503
Pegarse	0.012	2	0.5	0.503
Llevar	0.011	32	0.188	0.502
Ver	0.011	100	0.37	0.502

Casarse	0.01	2	0.5	0.502
Comprar	0.01	20	0.25	0.502
Usar	0.009	3	0.333	0.502
Poder comprar	0.001	3	0.333	0.5

Efectos aleatorios: informantes con un peso positivo para el corpus 2015

Variable	log-odds	Casos	P/P+A	Peso centrado
MA_M32_008	0.939	74	0.324	.715
MA_M31_025	0.755	51	0.471	.677
MA_M33_048	0.724	51	0.471	.670
MA_M31_015	0.612	63	0.365	.644
MA_H11_052	0.575	37	0.243	0.636
MA_H12_007	0.555	53	0.453	0.631
MA_M21_042	0.454	61	0.377	0.608
MA_M23_002	0.414	65	0.308	0.598
MA_M22_034	0.408	56	0.339	0.596
MA_H12_055	0.399	32	0.312	0.594
MA_H13_021	0.399	27	0.296	0.594
MA_H21_018	0.396	74	0.284	0.594
MA_H33_056	0.390	45	0.400	0.592
MA_M11_038	0.348	59	0.339	0.582
MA_M13_036	0.336	58	0.310	0.579
MA_M23_019	0.319	39	0.333	0.575
MA_H21_049	0.261	51	0.275	0.561
MA_H22_030	0.249	66	0.273	0.558
MA_M11_053	0.216	67	0.313	0.550
MA_M13_027	0.161	58	0.397	0.536
MA_M21_004	0.145	49	0.327	0.532
MA_M22_031	0.143	48	0.250	0.531
MA_M33_039	0.131	61	0.246	0.528
MA_H32_037	0.096	22	0.273	0.520
MA_M21_041	0.089	58	0.293	0.518
MA_H31_024	0.074	54	0.241	0.514
MA_H32_009	0.071	31	0.355	0.513
MA_H31_051	0.018	40	0.150	0.500

ANEXO III. HISTORIAS DISEÑADAS PARA EL EXPERIMENTO CON EYE-TRACKING

Historia 1

Contexto: Carlos y yo éramos compañeros de clase y pasábamos mucho tiempo juntos.

Relleno 1: Como niños, teníamos una gran imaginación.

Condiciones

- | | | |
|---|---------------------|--|
| 1 | Ambiguo total | Inventaba historias preciosas. Imaginaba personajes fantásticos. El relato era indescriptible. |
| 2 | Ambiguo parcial | Él inventaba historias preciosas. Imaginaba personajes fantásticos. El relato era indescriptible. |
| 3 | Mismo referente | Él inventaba historias preciosas. Él imaginaba personajes fantásticos. El relato era indescriptible. |
| 4 | Cambio de referente | Él inventaba historias preciosas. Yo imaginaba personajes fantásticos. El relato era indescriptible. |

Relleno 2: A día de hoy, Carlos ya ha publicado tres novelas de ficción.

Historia 2

Contexto: Cristina y yo asistíamos a clases de pintura por las tardes y explorábamos diferentes técnicas.

Relleno 1: Compartíamos la misma afición por capturar momentos únicos.

Condiciones

- | | | |
|---|-----------------|---|
| 1 | Ambiguo total | Creaba obras asombrosas. Dibujaba paisajes otoñales. La pintura era una pasión. |
| 2 | Ambiguo parcial | Ella creaba obras asombrosas. Dibujaba paisajes otoñales. La pintura era una pasión. |
| 3 | Mismo referente | Ella creaba obras asombrosas. Ella dibujaba paisajes otoñales. La pintura era una pasión. |
-

4	Cambio de referente	Ella creaba obras asombrosas. Yo dibujaba paisajes otoñales. La pintura era una pasión.
----------	---------------------	---

Relleno 2: Recordar esas tardes me llena de alegría.

Historia 3

Contexto: Marta y yo nos conocimos en la universidad y compartíamos la fascinación por el mundo de la moda. (FOTO)

Relleno 1: Solíamos asistir a eventos para descubrir nuevas tendencias.

Condiciones

1	Ambiguo total	Seleccionaba vestidos atrevidos. Elegía combinaciones con estilo propio. La moda era una forma de expresión.
2	Ambiguo parcial	Ella seleccionaba vestidos atrevidos. Elegía combinaciones con estilo propio. La moda era una forma de expresión.
3	Mismo referente	Ella seleccionaba vestidos atrevidos. Ella elegía combinaciones con estilo propio. La moda era una forma de expresión.
4	Cambio de referente	Ella seleccionaba vestidos atrevidos. Yo elegía combinaciones con estilo propio. La moda era una forma de expresión.

Relleno 2: Años más tarde, Marta montó su propio atelier.

Historia 4

Contexto: Lucía y yo participábamos en voluntariados y nos gustaba conocer proyectos sociales

Relleno 1: Sentíamos que de esta forma contribuíamos al bienestar comunitario.

Condiciones

1	Ambiguo total	Colaboraba en proyectos educativos. Participaba en actividades con jóvenes. La labor era muy gratificante.
----------	---------------	--

- | | | |
|---|---------------------|--|
| 2 | Ambiguo parcial | Ella colaboraba en proyectos educativos. Participaba en actividades con jóvenes. La labor era muy gratificante. |
| 3 | Mismo referente | Ella colaboraba en proyectos educativos. Ella participaba en actividades con jóvenes. La labor era muy gratificante. |
| 4 | Cambio de referente | Ella colaboraba en proyectos educativos. Yo participaba en actividades con jóvenes. La labor era muy gratificante. |

Relleno 2: Cada gesto solidario, ha dejado una marca imborrable en nuestra percepción del mundo

Historia 5

Contexto: Julio y yo éramos compañeros de piso y pasábamos los domingos cocinando. (FOTO)

Relleno 1: En verano preparábamos aperitivos para invitar a nuestros amigos.

Condiciones

- | | | |
|---|---------------------|---|
| 1 | Ambiguo total | Cocinaba unas tartas sabrosas. Tenía buena mano para los platos. La comida era una debilidad. |
| 2 | Ambiguo parcial | Él cocinaba unas tartas sabrosas. Tenía buena mano para los platos. La comida era una debilidad. |
| 3 | Mismo referente | Él cocinaba unas tartas sabrosas. Él tenía buena mano para los platos. La comida era una debilidad. |
| 4 | Cambio de referente | Él cocinaba unas tartas sabrosas. Yo tenía buena mano para los platos. La comida era una debilidad. |

Relleno 2: La cocina se convirtió en una excusa para reunirnos más a menudo.

Historia 6

Contexto: Álvaro y yo adorábamos viajar y solíamos aventurarnos en destinos exóticos.

Relleno 1: Explorábamos nuevas culturas desde muchas perspectivas.

Condiciones

- | | | |
|---|---------------------|---|
| 1 | Ambiguo total | Hablaba con gente del lugar. Descubría diferentes costumbres. Los viajes dejaban huella. |
| 2 | Ambiguo parcial | Él hablaba con gente del lugar. Descubría diferentes costumbres. Los viajes dejaban huella. |
| 3 | Mismo referente | Él probaba la gastronomía local. Él visitaba rincones remotos. Los viajes dejaban huella. |
| 4 | Cambio de referente | Él probaba la gastronomía local. Yo visitaba rincones remotos. Los viajes dejaban huella. |

Relleno 2: Tenemos miles de historias divertidas que contar a nuestros amigos.

Historia 7

Contexto: Jana y yo escuchábamos música en el coche y recorríamos los diferentes estilos musicales.

Relleno 1: Decidimos apuntarnos a clases de guitarra en verano.

Condiciones

- | | | |
|---|---------------------|---|
| 1 | Ambiguo total | Componía melodías propias. Exploraba ritmos novedosos. Las canciones eran inspiradoras. |
| 2 | Ambiguo parcial | Ella componía melodías propias. Exploraba ritmos novedosos. Las canciones eran inspiradoras. |
| 3 | Mismo referente | Ella componía melodías propias. Ella exploraba ritmos novedosos. Las canciones eran inspiradoras. |
| 4 | Cambio de referente | Ella componía melodías propias. Yo exploraba ritmos novedosos. Las canciones eran inspiradoras. |

Relleno 2: Aquella afición nos llevó a componer un disco.

Historia 8

Contexto: Antonio y yo éramos aficionados a la jardinería y dedicábamos muchas horas a cuidar de nuestro jardín.

Relleno 1: Sembrábamos algunas verduras en nuestro huerto.

Condiciones

- | | | |
|---|---------------------|--|
| 1 | Ambiguo total | Plantaba lechugas rizadas. Recogía tomates maduros. Todo era muy sano. |
| 2 | Ambiguo parcial | Ella plantaba lechugas rizadas. Recogía tomates maduros. Todo era muy sano. |
| 3 | Mismo referente | Ella plantaba lechugas rizadas. Ella recogía tomates maduros. Todo era muy sano. |
| 4 | Cambio de referente | Ella plantaba lechugas rizadas. Yo recogía tomates maduros. Todo era muy sano. |

Relleno 2: La cosecha de nuestro huerto se convirtió en una recompensa por tanto esfuerzo.

Distractor 1

Contexto: Laura y yo somos hermanas y en vacaciones ayudábamos en un refugio de animales.

Relleno 1: Nos encargábamos de los cachorros abandonados.

Estímulo Ella salía a pasear a algunos perros. Mientras rellenaba los comederos. Aquella labor nos reconfortaba.

Relleno 2: Al final de verano, adoptamos un perro del refugio para seguir cuidándolo en casa.

Distractor 2

Contexto: Blanca y yo bailábamos ballet y queríamos acceder al conservatorio.

Relleno 1: Desde muy pequeñas, nuestro sueño era ser bailarinas profesionales.

Estímulo Ella se movía al ritmo de la música. Yo era menos hábil para seguir los pasos. La danza nos encantaba.

Relleno 2: Nuestros caminos se separaron cuando ella logró entrar al conservatorio.

Distractor 3

Contexto: Javier y yo surfeábamos cuando íbamos de vacaciones al norte.

Relleno 1: Nos adentrábamos en playas cercanas con nuestras tablas.

Estímulo Desafiaba olas gigantes sin miedo. No tenía mucha destreza para mantener el equilibrio.

Relleno 2: Un día, tuvo un accidente y ya jamás volvió a surfear.

ANEXO IV. MODELO ESTADÍSTICO COMPLETO EYE-TRACKING

Total Reading Time

AOI_Condición	Estimación	Error estándar	N.º letras por palabra obs.	N.º letras por palabra fijada	TRT predicho	Error estándar TRT predicho
OT_C1	420.72	22.93	7.64	6.71	432.76	22.81
OT_C2	-16.40	22.68	6.92	6.71	416.36	21.98
OT_C3	-4.77	22.90	6.41	6.71	427.99	22.12
OT_C4	3.16	22.85	6.33	6.71	435.92	22.04
OC1_C1	36.83	23.26	7.35	6.71	469.59	22.77
OC1_C2	5.43	22.75	6.56	6.71	438.19	22.00
OC1_C3	5.52	22.88	6.45	6.71	438.28	22.12
OC1_C4	21.28	22.79	6.48	6.71	454.04	22.03
OC2_C1	-23.97	23.30	7.52	6.71	408.79	22.84
OC2_C2	-18.89	22.67	7.48	6.71	413.87	22.13
OC2_C3	-8.93	22.87	6.47	6.71	423.83	22.14
OC2_C4	-7.93	22.84	6.35	6.71	424.83	22.07
V1_C1	120.18	23.26	7.89	6.71	552.94	22.87
V1_C2	200.86	22.62	7.99	6.71	633.62	22.18
V1_C3	125.10	22.73	7.87	6.71	557.86	22.28
V1_C4	155.45	22.63	7.81	6.71	588.21	22.17
V2_C1	78.88	23.40	7.89	6.71	511.64	23.02
V2_C2	50.42	22.77	7.87	6.71	483.18	22.31
V2_C3	111.91	22.85	7.78	6.71	544.67	22.40
V2_C4	77.91	22.75	7.82	6.71	510.67	22.30
S1_C2	-134.07	27.66	2.98	6.71	298.69	26.98
S1_C3	-132.42	27.87	2.92	6.71	300.34	27.23
S1_C4	-135.35	27.62	2.98	6.71	297.41	26.96
S2_C3	-89.16	27.87	2.92	6.71	343.60	27.23
S2_C4	-97.28	31.88	2.00	6.71	335.48	31.48
P1_C1	9.37	23.25	7.24	6.71	442.13	22.73
P1_C2	1.60	22.62	7.48	6.71	434.36	22.08
P1_C3	32.95	22.74	7.30	6.71	465.71	22.19

P1_C4	55.29	22.66	7.33	6.71	488.06	22.13
P2_C1	-58.04	23.51	7.56	6.71	374.72	23.06
P2_C2	-40.11	22.93	7.50	6.71	392.65	22.39
P2_C3	6.93	23.01	7.54	6.71	439.69	22.51
P2_C4	25.01	22.93	7.60	6.71	457.77	22.44
PV1_C2	59.05	22.62	7.48	6.71	491.81	22.06
PV1_C3	59.57	22.73	7.32	6.71	492.33	22.17
PV1_C4	81.18	22.64	7.34	6.71	513.94	22.08
PV2_C3	28.82	22.80	7.50	6.71	461.58	22.30
PV2_C4	33.55	22.71	7.55	6.71	466.31	22.22

First Time Reading

AOI_Condición	Estimación	Error estándar	N.º letras por palabra obs.	N.º letras por palabra fijada	FRT predicho	Error estándar FRT predicho
OT_C1	339.31	13.52	7.64	6.71	343.86	13.44
OT_C2	-32.95	14.56	6.92	6.71	310.91	12.86
OT_C3	-32.73	14.69	6.41	6.71	311.13	12.95
OT_C4	-45.07	14.65	6.33	6.71	298.80	12.90
OC1_C1	12.02	14.94	7.35	6.71	355.88	13.41
OC1_C2	-53.45	14.60	6.56	6.71	290.41	12.87
OC1_C3	-41.95	14.68	6.45	6.71	301.91	12.96
OC1_C4	-58.60	14.62	6.48	6.71	285.26	12.89
OC2_C1	-189.80	14.95	7.52	6.71	154.06	13.46
OC2_C2	-183.40	14.55	7.48	6.71	160.46	12.96
OC2_C3	-180.82	14.68	6.47	6.71	163.04	12.97
OC2_C4	-182.75	14.65	6.35	6.71	161.11	12.92
V1_C1	-44.65	14.94	7.89	6.71	299.21	13.47
V1_C2	29.91	14.52	7.99	6.71	373.77	12.99
V1_C3	-3.13	14.59	7.87	6.71	340.73	13.06
V1_C4	-2.62	14.53	7.81	6.71	341.24	12.98
V2_C1	-81.95	15.00	7.89	6.71	261.91	13.58
V2_C2	-96.21	14.59	7.87	6.71	247.65	13.08
V2_C3	-75.67	14.65	7.78	6.71	268.19	13.14
V2_C4	-97.74	14.59	7.82	6.71	246.13	13.07

S1_C2	-179.94	17.64	2.98	6.71	163.93	16.16
S1_C3	-178.86	17.77	2.92	6.71	165.00	16.32
S1_C4	-180.66	17.62	2.98	6.71	163.21	16.14
S2_C3	-137.26	17.77	2.92	6.71	206.61	16.32
S2_C4	-152.11	20.11	2.00	6.71	191.75	18.89
P1_C1	-230.01	14.93	7.24	6.71	113.85	13.38
P1_C2	-186.22	14.52	7.48	6.71	157.64	12.93
P1_C3	-192.74	14.60	7.30	6.71	151.12	13.00
P1_C4	-178.00	14.55	7.33	6.71	165.86	12.95
P2_C1	-182.50	15.05	7.56	6.71	161.36	13.59
P2_C2	-193.70	14.66	7.50	6.71	150.16	13.10
P2_C3	-162.56	14.72	7.54	6.71	181.30	13.19
P2_C4	-159.98	14.67	7.60	6.71	183.88	13.14
PV1_C2	-49.74	14.52	7.48	6.71	294.12	12.91
PV1_C3	-55.45	14.59	7.32	6.71	288.41	12.98
PV1_C4	-81.27	14.54	7.34	6.71	262.60	12.92
PV2_C3	-163.66	14.63	7.50	6.71	180.20	13.07
PV2_C4	-178.56	14.57	7.55	6.71	165.30	13.02

Re-reading Time

Condición	Estimación	Error estándar	N.º letras WD_obs	N.º letras WD_fix	RRT predicho	Error estándar RRT predicho
OT_C1	79.98	19.77	7.64	6.71	87.30	19.65
OT_C2	17.22	22.24	6.92	6.71	104.52	18.72
OT_C3	29.54	22.45	6.41	6.71	116.84	18.87
OT_C4	50.03	22.40	6.33	6.71	137.33	18.79
OC1_C1	24.76	22.82	7.35	6.71	112.06	19.61
OC1_C2	59.88	22.31	6.56	6.71	147.18	18.74
OC1_C3	48.86	22.43	6.45	6.71	136.16	18.88
OC1_C4	81.28	22.34	6.48	6.71	168.58	18.78
OC2_C1	166.67	22.85	7.52	6.71	253.96	19.68
OC2_C2	165.71	22.23	7.48	6.71	253.01	18.88
OC2_C3	173.26	22.43	6.47	6.71	260.55	18.90
OC2_C4	176.46	22.39	6.35	6.71	263.76	18.82
V1_C1	164.77	22.81	7.89	6.71	252.07	19.71

V1_C2	171.03	22.18	7.99	6.71	258.33	18.94
V1_C3	128.52	22.29	7.87	6.71	215.82	19.05
V1_C4	158.48	22.20	7.81	6.71	245.77	18.92
V2_C1	162.62	22.95	7.89	6.71	249.92	19.88
V2_C2	148.74	22.33	7.87	6.71	236.04	19.09
V2_C3	189.70	22.41	7.78	6.71	277.00	19.18
V2_C4	177.67	22.31	7.82	6.71	264.96	19.07
S1_C2	44.56	26.99	2.98	6.71	131.86	24.09
S1_C3	45.05	27.21	2.92	6.71	132.35	24.36
S1_C4	44.28	26.96	2.98	6.71	131.58	24.06
S2_C3	46.71	27.21	2.92	6.71	134.00	24.36
S2_C4	49.03	31.11	2.00	6.71	136.33	28.82
P1_C1	239.69	22.80	7.24	6.71	326.99	19.56
P1_C2	188.10	22.18	7.48	6.71	275.40	18.83
P1_C3	226.30	22.30	7.30	6.71	313.60	18.95
P1_C4	233.80	22.22	7.33	6.71	321.10	18.88
P2_C1	127.08	23.06	7.56	6.71	214.37	19.93
P2_C2	156.66	22.48	7.50	6.71	243.95	19.17
P2_C3	172.62	22.57	7.54	6.71	259.92	19.31
P2_C4	188.20	22.49	7.60	6.71	275.50	19.23
PV1_C2	108.68	22.18	7.48	6.71	195.97	18.81
PV1_C3	115.28	22.29	7.32	6.71	202.58	18.93
PV1_C4	162.77	22.20	7.34	6.71	250.07	18.83
PV2_C3	194.06	22.36	7.50	6.71	281.36	19.07
PV2_C4	213.75	22.27	7.55	6.71	301.05	18.98

ANEXO IV. ABSTRACT

As an essential component of human communication, observing language involves the careful study of two closely related processes: production, associated with a speaker who is in charge of constructing a message; and perception, linked to the audience, who must then decode the message using their analytical and evaluative capacities. Although the study of linguistic production—which production rules are activated in certain contexts and with what frequency—is fundamental, the analysis of perception, understood as one of the cognitive mechanisms that mediate between the internal and external worlds via the senses, should not be neglected either. Attempting to study language, or more specifically the processes of linguistic variation, without attending to its implicit social character or the cognitive processes triggered in the human mind would make for a partial and biased analysis of a broad and complex phenomenon.

In this research, the study of the social nature of language is not limited to its association with variables that allow us to stratify the sample under analysis, nor is it treated as a mere accessory to linguistic facts, but rather incorporated through an holistic perspective, in line with the proposals of Moreno Fernández (2012) and Caravedo (2014). This approach is based on the premise of a complete interweaving between language and society: there is no natural language without a social dimension, nor any autonomous linguistic structure without social involvement. In this way, the traditional disjunction that considers the structurally linguistic as internal and the social as external is overcome, so that both aspects are seen as complementary and inseparable if a comprehensive and cohesive approach to the linguistic phenomenon is to be adopted.

This doctoral thesis is grounded in the theoretical foundations of cognitive sociolinguistics, which is interested in the variable production of linguistic manifestations and, in turn, in the perception that speakers have of variability (Moreno Fernández, 2012, p. 89). Thus, our study focuses on grammatical variation in Spanish, specifically on the alternation of the presence and absence of the subject personal pronoun from the perspective of cognitive sociolinguistics. This approach leads us to structure our work in two main sections: the uses of the pronominal subject; and the cognitive perception of the absence/presence variation of the personal pronoun subject.

The first part of the work is dedicated to production. It deals with the use of subject personal pronouns in the Spanish spoken in the city of Malaga. The data were gathered from two PRESEEA corpora, collected in 1995 and 2015, respectively. Multiple factors are analysed, both linguistic (phonetic, morphological, lexical, pragmatic, discursive, and semantic) and social, in order to explain the manifestation of the pronominal subject (Chapter II). Following the analysis of production, the second phase of the work, oriented towards perception, comprises a study using eye-tracking techniques. To provide experimental evidence, the cognitive processing pattern is observed in relation to variations in the presence or absence of the subject personal pronoun (Chapter III).

1. Variation of the pronominal subject: theoretical framework and antecedents

In contrast to languages such as English, French, or German, in some Romance languages, such as Spanish and Italian, it is possible in certain utterances to indicate the subject exclusively through verbal inflection. For this reason, tacit subjects have pronominal properties and agree in person and number with the verb, and in gender and number with other elements of the utterance (Fernández Soriano, 1999; RAE & ASELE, 2009; etc.).

Traditional grammars have typically considered the omission of the pronoun as the default option. They most commonly explain its presence with one of three reasons: redundancy, emphasis, or disambiguation. In contrast to these arguments, Fernández Soriano (1999, p. 1227) considered the pronominal subject to be a free variable, whose presence is more a question of style than of grammar. However, this assertion has been categorically denied by Bentivoglio (1987, p. 66) and Silva Corvalán (2003). In this, the subsequent development of variationist studies has made it possible to clarify that -excluding contexts where alternation is impossible- the expression of the subject is probabilistically correlated with a complex network of pragmatic, semantic, syntactic, and morphological factors, such as co-referentiality, topicality, and ambiguity of the verbal form, although these factors do not condition variation in a categorical way.

1.1. Variationist sociolinguistics

Language is, by its nature, an intrinsically variable system and reflects this variability in its use. When speakers interact, they select different linguistic resources to express different meanings, or use alternative forms to refer to the same concept (Moreno Fernández, 2009, p. 21). This variation may be motivated by linguistic as well as social factors, but the social connection inherent in any linguistic study is indisputable, as without this social dimension language could not fulfil its main function.

The discrepancies between variation at the phonetic-phonological and grammatical levels are obvious. Silva-Corvalán (2001, p. 129-132) stated that phonological variation cannot be equated with syntactic variation, since there is less of the latter than the former. Moreover, syntactic variation is more difficult to measure and quantify, because the contexts in which it occurs are less frequent and more difficult to delimit.

Although it is undeniable that observing, analysing, and establishing the degree of identity between grammatical variants is an extremely complex task, it is precisely this complexity that is proposed as an object of study in the field of syntactic variation. Silva-Corvalán (2001, p. 129-131) observed that the difficulties arising from the analysis of meaning-bearing elements have raised challenges in connecting sociolinguistics with semantics and discourse pragmatics more closely. Thanks to more recent approaches, such as that proposed by cognitive sociolinguistics (Moreno Fernández, 2012), the boundaries of study are widening towards a more holistic view, the foundations of which are the cognitive resources involved in discourse, contextualised language use, and social perception. This approach offers a useful framework for dealing with the complexity of variation, especially syntactic variation, by showing that semantic, pragmatic, and cognitive factors converge with those of a social and stylistic order in the observation of the complex web of correlations present in language use.

Thus, languages, considered as complex adaptive systems (Beckner et al., 2009), are characterised by the intervention of multiple agents, the cumulative use of interactions, the concurrence of factors in the speaker's behaviour, as well as the emergence of patterns of interaction, experience, and cognition. Communicative cooperation between individuals makes language an emergent phenomenon with two dimensions: an increasingly complex social dimension shared by interlocutors and, in turn, a more individualised dimension

based on factors of a cognitive, psychomotor, perceptual, and experiential nature, where the preferred variants of language use gradually become embedded in the individual mind. Consequently, if the researcher's intention is to observe language in all its complexity and dynamism, it is necessary to employ theoretical constructs that allow for a multidimensional and comprehensive conception of the linguistic, social, and cognitive components.

In light of the above, cognitive sociolinguistics is conceived of as a metatheoretical model that seeks to analyse language relationally in its natural, cultural, social, and situational environment. Its aim is to approach both the linguistic and the social as realities from the perspective of human cognition. Cognitive sociolinguistics therefore focuses on the study of the cognitive resources involved in language processing and its contextualised use. This perspective places particular emphasis on speakers' knowledge of language in its social context, considering aspects such as communicative environments, interaction processes, linguistic variation and change, as well as how these are perceived by speakers.

From this point of view, the study of language variation is not conceived as an objective and autonomous phenomenon, but as a two-dimensional process: variation is externalised in production and acquires its intrinsic meaning in perception. In the words of Caravedo (2018), limiting the analysis of variation exclusively to production, without taking perception into consideration, would be a mistake. To make up for this limitation, cognitive sociolinguistics addresses this multidimensionality by taking an interest in production, as manifested in language use, but also in the perception that speakers themselves have of variation (Moreno Fernández, 2012, p. 89). In this framework, it is essential to specify what we mean by production and perception.

Production consists of the output, i.e., the concrete linguistic manifestations which the speaker emits in conversational interaction, and which give rise to the speech acts to which all other sociolinguistic uses can be ascribed (Moreno Fernández, 2012, p. 77). Sociolinguistic variation deals mainly with the organised, objective, and systematic study of speakers' production (Caravedo, 2014, p. 36), the analysis of which may be based, for example, on the examination of oral corpora.

The study of production is fundamental, as cognitive sociolinguistics requires systematic observation when analysing language on the basis of its use and, therefore, quantitative and qualitative analyses are essential to understand how it is produced and

how general rules of prediction are established in use (Moreno Fernández, 2012, p. 166-171).

Perception, as defined by Caravedo (2018, p. 70-71), is one of the main components in the development of cognition: it acts as a basic mechanism for approaching the knowledge of a language and is essential in cognitive activity. Perception brings the senses into play to recognise and know the external world, i.e., it is a bridge that connects the inside with the outside. Perception is linked to the senses—visual, auditory, tactile, and olfactory—and through these different channels, especially auditory and visual, essential external information is received. Perception is therefore understood as the reception of linguistic material produced by the speaker, and its subsequent decoding and processing by the listener (input).

1.2. Background

For almost fifty years, researchers have shown great interest in the phenomenon of pronominal subject expression, resulting in a substantial body of literature. Although the use of the subject personal pronoun has been the focus of this study, examining how speakers perceive particular syntactic structures can also contribute to a better understanding of this linguistic phenomenon.

In the context of production, studies on the use of the pronominal subject have been approached from different perspectives and, on occasion, have been based on diverse corpora, often without shared methodological principles, which has made it difficult to compare the results of different works. In this sense, the studies could be divided into two stages, since the appearance of the Project for the Sociolinguistic Study of Spanish in Spain and America (PRESEEA) can be used as a differentiating milestone. The creation of PRESEEA has made it possible to overcome many of the previous problems, such as the lack of comparable corpora, the absence of a unified methodology, and the lack of representativeness in the samples.

Although research on the expression of the pronominal subject in Spanish has traditionally focused on its use, we also find studies which, from the perspective of perception, analyse how listeners or readers process and resolve possible ambiguities in the discourse. These studies focus on the influence of the presence or absence of different

elements, such as the personal pronoun, on the interpretation of the referent. These ambiguous contexts typically occur when there is no clear referent or when there are several possible candidates for the subject or object of a sentence, so that the receiver must activate disambiguation mechanisms to correctly identify to whom the speaker is referring.

1.3. Objectives and hypothesis

In order to carry out the study of pronominal subject variation in the city of Malaga from the approach of cognitive sociolinguistics, this work is divided into two main sections, production and perception.

Accordingly, the following initial hypotheses are formulated, aimed at clarifying the causes that lead to the presence or absence of the subject personal pronoun in the Spanish spoken in Malaga.

1. Hypothesis of underlying grammaticalisation: the linguistic-social determinants identified as significant in previous research on similar speech communities (Manjón-Cabeza Cruz et al., 2016; Repede, 2019; Roselló Verdeguer, 2021) will also affect the speech of the city of Málaga. This would corroborate the existence of an underlying grammar common to all varieties of Spanish, even if superficially there are differences in the rates at which pronouns are omitted (Orozco et al., 2024).
2. Hypothesis of the impact of linguistic-social variables: linguistic variables will have a greater weight than social variables in conditioning the variation of the pronominal subject (Martín-Butragueño, 2020).
3. Hypothesis of internal language restriction: the appearance or omission of the subject will be conditioned by characteristics inherent to the nature of the subject, such as specificity, or animacy.

Based on these questions and hypotheses, which will guide our research, we establish the following objectives:

- a. To identify the conditioning factors in usage. This first objective is mainly concerned with analysing the use of the pronominal subject in the Spanish spoken in the city of Malaga. Data from the two oral corpora of PRESEEA-Malaga are used for this objective. Specifically, the aim is to determine the relative frequency of the

presence and absence of the pronominal subject, and the factors -both linguistic and social- that condition its use.

In addition to this primary objective, we define two secondary objectives in order to establish diachronic and synchronic comparisons with reference to the city of Malaga.

- b. To critically assess the internal evolution of subject pronoun usage in Malaga (1995 and 2015). This objective aims to detect possible changes in the expression of the subject over two decades in the city of Malaga. In addition to comparing the frequency of occurrence of the pronominal subject in Malaga speech in each corpus, significant variations are identified according to linguistic and social conditioning factors.
- c. Contrasting the results for Malaga with other cities in the PRESEEA project. The fact that the data were collected with a consistent method means that, despite the differences between the Spanish-speaking areas studied, we are able to compare the results, with the aim of sketching a potential general pattern that justifies the use of the pronominal subject.

In order to approach the study of the pronominal subject from the prism of cognitive sociolinguistics, the following section builds on the usage data obtained from the corpus analysis (production) to consider the cognitive perception of variation in the presence or absence of the pronominal subject. These questions have been formulated to guide the analysis and to develop the hypotheses proposed below with greater specificity:

- 4. Hypothesis of pronominal resolution: the presence of the personal pronoun will facilitate the resolution of ambiguity, as it constitutes a cognitive anchor that reduces the interpretative load (Ariel, 2001).
- 5. Hypothesis of degree of occurrence: although the subject may be a useful element in understanding utterances, different cognitive processing tendencies are expected depending on the degree of presence or absence. These differences may favour (or, alternately, hinder) certain readings, either because of recursivity issues or because of the introduction of new referents (Badecker & Straub, 2002; Gernsbacher, 1989; Grant et al., 2020).

Therefore, after establishing the expected questions and hypotheses in this research, the main objective of this section is the following:

d. To determine the role of the pronominal subject in the cognitive processing of utterances using the eye-tracking tool. This objective aims to identify how variation between the explicitness or omission of the subject personal pronoun affects sentence understanding. In addition, the aim is to establish which structures involve higher or lower cognitive effort.

The general objective derives from two specific objectives which are linked to the linguistic variables analysed:

e. To observe the cognitive strategies involved in the resolution of ambiguity. From the different parameters of reading time, the aim is to identify disparities in processing patterns when comparing sentences that may be ambiguous with those that are unambiguous.

f. To compare the effects of referential continuity versus referent change. This objective is aimed at analysing how the introduction of a new referent affects the cognitive load during the first reading and re-reading.

In short, the cognitive perception approach complements the duality expected from cognitive sociolinguistics by integrating linguistic production with the processes of comprehension and mental re-creation of variation.

2. The production of the subject personal pronoun in the PRESEEA-Málaga corpora

This chapter is focused on the study of the use of the pronominal subject from the methodological approach of sociolinguistic research. Using the materials collected in the oral corpora of specific speech communities, we investigate which conditioning factors of a linguistic (phonetic, morphological, lexical, pragmatic, discursive, and semantic) or social nature motivate the omission and presence of the pronominal subject. To establish the relationships between these variables, it is necessary to use quantitative methods to discern how the frequency of concurrence of different linguistic-social factors explains the variation.

On the basis of the common methodological framework proposed by PRESEEA, the necessary sample is established for each population according to the social macro-categories of age, gender and level of education. In addition, the linguistic phenomena considered a priority have a coding guide in which the exclusion of certain contexts of analysis and the selection of common variables are advised. In this way, this initial methodological scheme makes it possible for each speech community to adhere to the general principles established by PRESEEA, but also incorporate variables that reflect its specific sociolinguistic particularities.

This Doctoral Thesis analyses the oral speech of the city of Málaga based on two corpora that collect the voices of the city: The Spanish spoken in Málaga, which we will refer to as PRESEEA-Málaga 1995, since the collection of the materials dates from that date (Vida Castro, 2007; Ávila Muñoz et al., 2008; Lasarte Cervantes et al., 2009), and the complementary oral corpus of the Spanish spoken in the city of Málaga, which we will refer to as PRESEEA-Málaga 2015 (Vida Castro, in press).

The sample volume of the 1995 corpus is evenly distributed with four informants per cell (N=72). In the 2015 corpus, three informants are assigned per cell (N=54). In total, the sample we address in this paper consists of a total of 126 informants, whose distribution is presented in Table 1.

The materials in this study are coded according to the Coding Guide for the variable expression of the pronominal subject by Bentivoglio et al. (2011), which has recently been updated in the PRESEEA Guide for the study of the expression of the pronominal subject by Guerrero González and Manjón-Cabeza Cruz (2022). In our study, we analyse variables that come from both methodological proposals and, in addition, we incorporate factors that are not previously contemplated in either of the two guides. Thus, we use as a dependent variable the presence or absence of the personal pronoun subject along with as 19 independent variables of a linguistic nature and three of a social nature.

2.1. Results

At first sight, the omission of the pronoun is observed with a far greater regularity, making up between 75-80% of the total number of occurrences. Moreover, it is significant

that the 2015 corpus, despite presenting a lower number of total cases, exhibits a slightly higher rate of explicit pronoun usage, at 25%, compared to around 20% in the 1995 corpus.

The variable which exerts the greatest influence on the variation of the pronominal subject is the person of the subject, and it obtains the lowest p-values in both corpora. Also, both synchronies coincide in showing higher weights for the 2nd (0.85 in the 1995 corpus and 0.81 in the 2015 corpus) and 1st person singular (0.70 in 1995 and 0.73 in 2015). The second most important variable in the regression analyses is the pragmatic-discursive variable of co-referentiality. The factors promoting a higher presence of the pronominal subject are non-correferentiality (0.64 in 1995, 0.57 in 2015), not applicable (0.52 in 1995 and 0.51 in 2015), and in the 2015 corpus only correferentiality with object (0.57). The third most important variable is associated with a characteristic of the subject of the sentence: specificity. Despite the fact that the specific referents favour the explicitness of the personal pronoun subject in both corpora with a weight of 0.65 in the 1995 corpus and 0.71 in the 2015 corpus.

The verbal mode turns out to be the fourth variable in the regression model in both corpora, despite the fact that the factors have different importance: in the 1995 corpus the indicative (0.72) contributes significantly to the expression of the personal pronoun as does, to a considerably lesser extent, the subjunctive (0.58). By contrast, in the 2015 corpus, the subjunctive (0.68) influences the presence of the pronoun in a very similar way to the indicative (0.67).

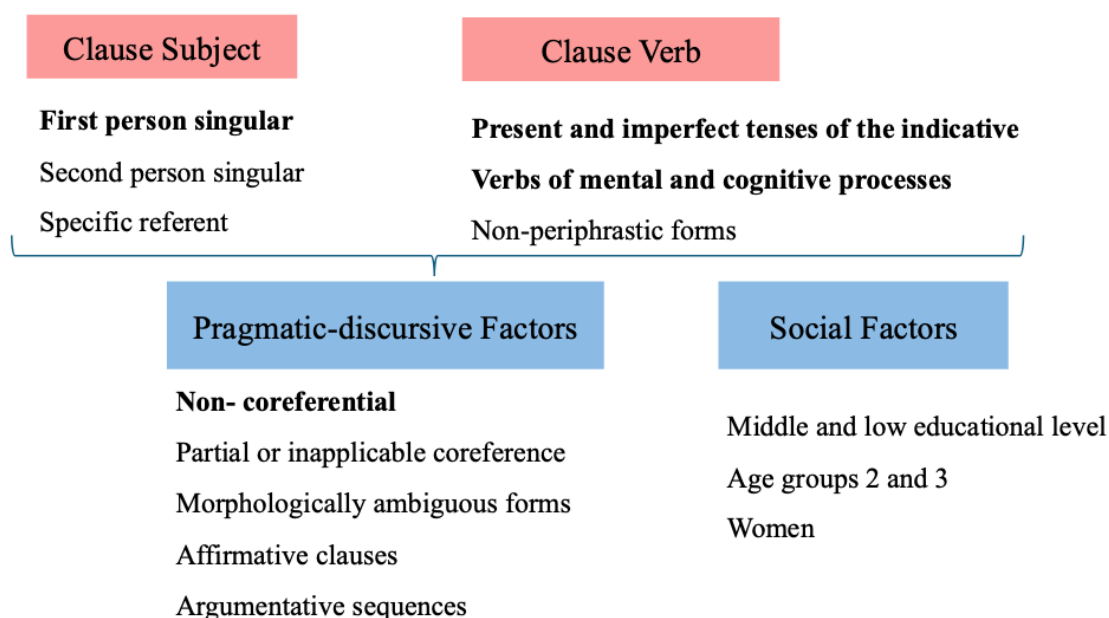
Finally, both synchronies coincide in considering clause type as a significant variable in their regression model, but not with the same level of importance, since in the 1995 corpus it occupies the sixth position below morphological ambiguity. In the two corpora, affirmative clauses obtain the highest weight, at 0.62 in the 1995 corpus and 0.58 in the 2015 corpus, and, in addition, negative clauses also favour the presence in the 1995 corpus (0.61).

In addition to the similarities presented in the two cross-sectional datasets, there are also differences. On the one hand, the 1995 corpus includes the following variables in its regression model: contextualised morphological ambiguity, whose ambiguous forms (0.57) and contextually unambiguous forms (0.54) favour the presence of the pronoun; non-periphrastic forms also correlate with the explicitness of the pronominal subject (0.60); and, finally, the type of discourse in argumentative sequences achieves a weight of

0.56. It should be noted that in the 2015 corpus the semantic classification of the verb according to the ADESSE typology is significant, and especially the existential (0.65), mental (0.54), relational (0.53) and material (0.52) process verbs are the ones that promote a higher occurrence of the pronominal subject.

One of the objectives proposed in this work concerns the possibility of establishing a general pattern regarding the uses of the personal pronoun that can be extrapolated to different Spanish-speaking areas. To this end, we proceed to compare the results obtained in the city of Malaga with previous studies carried out in other Spanish-speaking communities. Thanks to the studies on the expression of the pronominal subject within the scope of PRESEEA, we have a large set of data from different Spanish-speaking areas with a partially homogeneous methodology, and, in this way, the conclusions reached should be robust and transversal.

Not all the factors are of equal importance, and not all have an impact in all of the cities analysed, so we distinguish between general rules shared by the great majority of the works discussed here (presented in bold) and partial rules, which require further empirical evidence to confirm their validity in all Spanish-speaking areas.



Graph 1. Pattern of pronominal subject use.

Thus, the factors that seem to have the greatest influence on the presence of the subject personal pronoun are the person of the subject, co-referentiality, tense, mood, and

the semantic class of the verb. The explicitness of the pronoun tends to be more frequent when:

- The subject is 1st person singular, and in some cases, also 2nd person singular.
- In contexts where there is no co-referentiality the subject pronoun tends to appear more frequently as it does, albeit to a lesser extent, in situations where co-referentiality does not apply or is of a partial type.
- The verb tenses are present or imperfect, as well as when the verb is in the indicative mood.
- The verbs belong to the semantic class denoting mental or cognitive processes.

Although these comparisons should be interpreted with caution – given that the methodological procedures are not completely identical and there are variations in the statistical models that depend on the consideration (or not) of random effects, among other details – the pattern of subject use reveals transversal tendencies in the different Spanish-speaking areas. Thus, it can be seen that, although explicitness rates vary according to geographical area, there are nevertheless underlying linguistic mechanisms that determine the appearance of the pronominal subject in Spanish.

3. The perception of morphological ambiguity and co-referentiality.

Experimental analysis through eye-tracking

The methodological combination of corpus analysis and experimentation provides research with complementary evidence, making the data more robust and reliable when dealing with a phenomenon from different perspectives (Recio Fernández et al., 2021, p. 1007). In experimentation, discourse is approached as an activity, from which data on reactions (e.g., cognitive effort measured in milliseconds) to certain stimuli (text fragments or certain grammatical structures) are extracted. The function of this work is to clarify possible correlations between specific features of utterances operating at the discourse level and the processing patterns of language users (Loureda et al., 2019, p. 2-3).

In this way, our objective is to evaluate, in particular, the function of the pronominal subject in the processing of certain syntactic structures. Specifically, we are interested in determining whether, in situations of morphological and/or contextual ambiguity, the

explicit pronoun acts as a disambiguator to facilitate interpretation or whether, on the contrary, pronominal redundancy increases the costs of linguistic processing.

This experiment aims to observe how pronominal subject variation influences the cognitive processing of stimuli, specifically the reading metrics selected as dependent variables: Total Reading Time, First Reading Time and Re-reading Time. As for the independent variables, we considered variation in the presence or absence of the subject personal pronoun, the ambiguity of certain syntactic structures with morphologically ambiguous verbs, and co-referentiality, understood as referential continuity or change.

In these terms, our experimental design is composed of four conditions (C1, C2, C3 and C4), where each critical item is adapted within different stories (story 1 to story 8). For each list, we assign a minimum of 20 participants (a total of 80 participants), who read each condition twice in different stories. The initial sample collected consisted of 124 participants, but 25 participants were excluded based on the outlier exclusion criteria or after detecting errors in the recording. The analysis was therefore performed on a total of 99 participants.

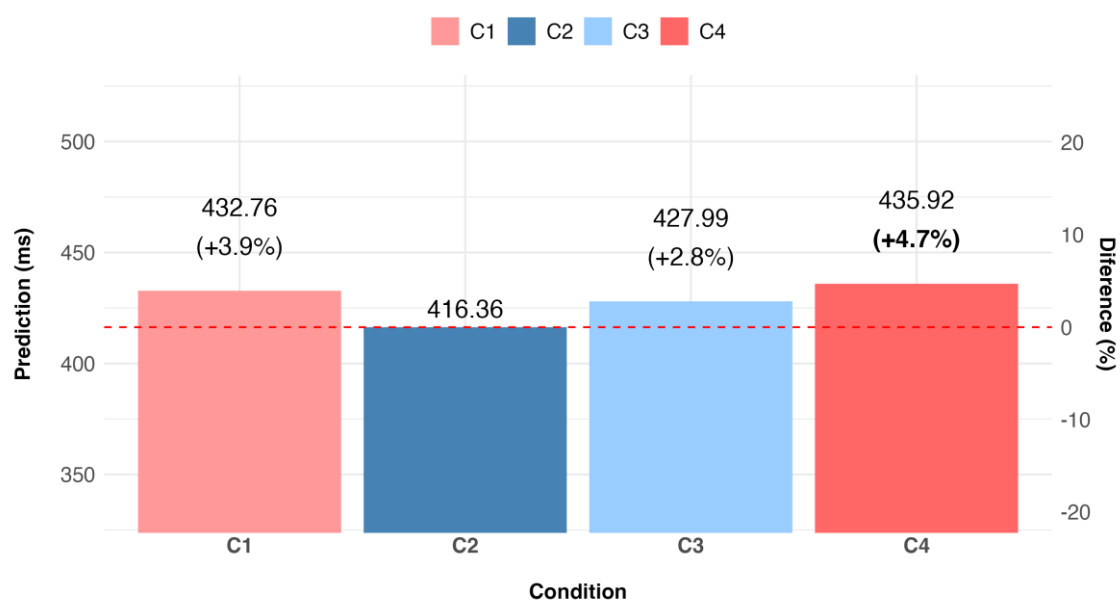
3.1. Results

This study of cognitive perception using the eye-tracking tool reveals the existence of different processing patterns depending on the conditions studied. The linguistic factors implemented in the stimuli – the presence or absence of the subject, morphological ambiguity and co-referentiality – activate different processing dynamics in terms of cognitive effort. Based on the total time invested in processing each structure, we establish a comparison between the conditions studied in terms of their degree of cognitive complexity.

The incorporation of this complementary analysis by means of eye-tracking has provided experimental evidence of the processing patterns involved in the reading of certain structures. The integration of the PRESEEA-Málaga corpus analysis, in order to identify the variables that affect the presence of the pronoun and determine its frequency of use, together with the experimental results offers a comprehensive view of the linguistic phenomenon studied. This complementary study reinforces the hypotheses put forward

and allows us to know in detail how certain structures detected in real use are cognitively processed.

As we can see in the Graph 2, the condition with partial presence of the subject (C2) is the least demanding in terms of cognitive effort. From this we deduce that omission of the pronoun in the second clause facilitates the consideration of co-referential continuity. The optimisation in the processing of the complete utterance that the partial presence of the subject implies corroborates the hypothesis of the advantage. Our results coincide with the theory put forward by Ariel (1990), who highlighted the high level of accessibility when the connection between the antecedent and the anaphora is close, reflected in a lower cognitive cost in its processing.



Graph 2. Comparison of the four conditions based on the Total Reading Time (TRT) of the entire utterances (OT)

In contrast, referential doubling (C3), although it is the second condition that requires less processing time, has shown that the explicitness of the subject personal pronoun in its two clauses penalises sentence processing, since repetition seems to reflect cognitive overload. This result is in line with Gernsbacher's (1989) approach and therefore corroborates our repetition penalty hypothesis.

The double absence of explicit subject (C1) ranks third in terms of processing time required. The absence of the subject seems to generate a certain degree of ambiguity and, consequently, a significantly higher cognitive effort is required to identify the referent,

verifying the ambiguity hypothesis. The empirical results are in line with expectations because the competition effect or the multiple-candidate effect (Gernsbacher, 1989; Badecker & Straub, 2002) imposes more processing effort on the individual. However, once the ambiguity has been resolved, the participant requires less time for rereading.

Finally, the change of referent (C4) involves the greatest cognitive effort on the part of the reader. These findings are in line with those obtained by Khan (1999), who considered that referential change, when conceived of as separate events, can delay and hinder the comprehension process. Thus, the referential change hypothesis, which anticipated a higher processing cost due to the introduction of a new referent, is confirmed.

The four linguistic structures studied showed differences in their reading times and, therefore, in the cognitive processing that the participant does when reading the stimuli. Thanks to the analysis of the different temporal parameters, we have detected that the reading times between the FRT and the RRT vary when comparing the conditions that present referential ambiguity with respect to the rest of the conditions, especially in the referential change stimuli. Despite the different strategies used for the cognitive perception of the utterances, Figure 27 represents the total reading time (TRT) in milliseconds for the four conditions studied.

These data confirm that the presence of the subject in one of the clauses is a key cognitive factor in facilitating comprehension, given that the stimulus with partial presence of the subject requires the least total reading time (416.35 ms). From this point onwards, the other conditions progressively diverge. The difference of approximately 20ms between the most efficient condition and the most cognitively demanding one reveals the decisive role of the presence or absence of the pronominal subject in comprehension: when its presence is superfluous - only providing redundant information- or is made explicit to signal a referential shift, the reader tends to increase his cognitive effort to understand the utterance he has just read; on the other hand, the partial use of the subject, understood as referential continuity, speeds up and facilitates textual comprehension.

ANEXO V. CONCLUSIONS

This study of pronominal subject variation in Spanish from the innovative approach of cognitive sociolinguistics has made notable contributions for several reasons. The observation of syntactic variation via the integration of two complementary lenses (production and perception) has allowed us to approach a more realistic view of how this linguistic phenomenon functions. Considering the matter from the perspective of production was essential in determining how frequently subject pronouns occur and in detecting which linguistic – and sometimes social – variables lead to subject explicitness. For its part, our analysis of perception complemented the results obtained from the corpora by allowing us to explore individuals' cognitive processing when reading sentences by means of the experimental eye-tracking technique.

Regarding production, the syntactic variation involved in the absence and presence of the personal pronoun has been shown to be multidimensional in nature, as the influence of various kinds of factors – phonetic, morphological, syntactic, pragmatic, discursive, and social – has been demonstrated. The claim that syntactic variation, more specifically the absence or presence of the pronoun, is free alternation, attributed to stylistic rather than grammatical issues, has been laid to rest. This study has documented the existence of certain recurring factors motivating the presence of the subject and has also evidenced similarities in terms of the grammatical conditioning factors identified by previous research.

Firstly, this work has provided original data on the use of the expression of the pronominal subject in the speech of Málaga and, in this way, it has been possible to situate this syntactic variable on the map of the linguistic city. Thanks to the corpus collected by the VUM group, we have been able to study the voices of the city in two sets of cross-sectional data separated by a period of 20 years. This diachronic study has shown continuity in the linguistic uses and in the variables that condition the absence and presence of the pronominal subject, but it has also provided us with indications of the possible evolution over time of the use of the pronominal subject.

Moreover, as this research was carried out within the framework of the PRESEEA project, the use of an agreed upon methodology shared by all project teams enabled us to

establish comparisons with data obtained from other Spanish-speaking communities. Although we observed certain differences –such as the different rates of presence of the pronominal subject– across the geographical areas analysed, we also found a notable diatopic coherence in terms of conditioning factors as well as having detected general and partial rules that justify the expression of the pronominal subject.

With the aim of adapting and updating the methodological proposal to the city of Malaga, we have incorporated some variables such as the referent of the subject and phonetic ambiguity and, in addition, we have revised those variables where the incorporation of the context to their analysis provided a more integrative approach, especially in the face of possible ambiguities, both phonetic and morphological. This extension and reconfiguration aimed at exploring the variables within their context, rather than analysing them in isolation, has allowed us to observe in a more realistic way the natural dynamics between the linguistic factors that intertwine and mutually condition each other in discourse.

The main objective of this corpus study has been achieved through the determination of the frequency with which the pronominal subject occurs and the identification of the linguistic-social factors which influence its use in the city of Malaga. Likewise, we consider that the general hypotheses developed at the beginning of the study have been corroborated with guarantees, since some of them have not only been verified in the city of Malaga, but have also shown their transversality when observed in other Spanish-speaking communities.

The first of our hypotheses (the underlying-grammar hypothesis) posited that, in spite of dialectal and temporal differences, an underlying grammar, unifying the linguistic mechanisms that lead to greater expression of the pronominal subject, would exist. This has been confirmed by identifying a linguistic pattern extendable to twelve studies conducted in different dialectal areas. This revealed that the person of the subject is the variable with most explanatory power across all the investigations—particularly, first and second person singular subjects condition the highest occurrence of the pronoun. The coreferentiality factor—when a shift in referent occurs—demands greater explicitness of the subject. Verbs accompanying the pronoun favour its expression when they are in the present or imperfect tense, when the mood is indicative, and when the verb's semantic class involves mental or cognitive processes.

The results have demonstrated that linguistic variables wield greater influence than social variables, with the latter exhibiting a less uniform tendency and offering less explanatory power. Thus, our second hypothesis (the weight-of-variables hypothesis), which attributed greater importance to linguistic factors, has been verified.

However, an in-depth study of social variables beyond those used for the pre-stratification of speakers—such as those related to the influence of the informant’s cultural capital—may reveal differences in the use of the subject personal pronoun that have not yet been taken into account.

The combination of corpus analysis and cognitive experimentation has provided greater methodological robustness, since it allowed us to complement the examination of different variables with an exploration of the mental processing that individuals carry out when faced with structures present in their language. Studying how the variation between absence and presence of the pronominal subject influences utterance comprehension has yielded essential information about the function the subject fulfils in certain contexts. Moreover, it has enabled us to identify differences depending on the cognitive effort required by each type of stimulus. In this way, we achieved the primary aim of the perception section.

The experimental eye-tracking analysis also revealed different cognitive strategies, especially in the resolution of ambiguity and the introduction of a new referent in the utterances. Stimuli presenting total referential ambiguity (C1) required the greatest cognitive effort on first reading yet exhibited the shortest re-reading times. By contrast, stimuli involving a referential shift (C4) showed the opposite tendency: faster initial readings but high processing costs during re-reads.

Our initial hypotheses have been corroborated, since the subject personal pronoun functioned as a cognitive anchor when resolving ambiguity in verb forms that are morphologically identical. Thus, stimuli containing a single explicit subject (C2) proved to be optimal utterances in terms of the effort expended by participants and the “reward”—namely, sentence comprehension. Therefore, our pronominal-resolution hypothesis has been validated, as the subject personal pronoun reduced cognitive load compared to sentences with a completely omitted subject.

Despite the cognitive support afforded by the presence of the pronominal subject in these stimuli, it is evident that the subject can either facilitate or impede reading depending on its degree of occurrence. Referential duplication (C3), rather than aiding sentence comprehension when paired with morphologically ambiguous forms, produced reading-time overloads compared to stimuli with only a partial presence of the subject. In this way, the degree-of-occurrence hypothesis was confirmed: while explicit subject marking promotes understanding, repeating the same referent or introducing a referential shift increases cognitive load.

The study of the absence and presence of the pronominal subject in Spanish—approached from the perspective of cognitive sociolinguistics—has yielded an holistic view by linking linguistic and social dimensions with cognitive processing. Our work has underscored the necessity of analysing linguistic phenomena from a multidimensional standpoint that encompasses both production—grounded in authentic usage—and perception—through the cognitive resources engaged during processing. Moreover, the results highlight the importance of situating analysis within its communicative context, from the immediate linguistic environment of the utterance to the background knowledge already present in the individual's mind.

It is also important to acknowledge the study's limitations. The first arises from the nature of the corpora employed: since they are based on semi-structured interviews, participants' interventions in spontaneous speech are to some extent controlled and the communicative context is confined to interviewer and interviewee, potentially restricting the use of plural grammatical persons. Future research would benefit from incorporating more varied and natural communicative settings—such as everyday, spontaneous conversations—to provide a more authentic perspective on the phenomenon under study.

Secondly, despite the advances afforded by techniques such as eye-tracking for measuring cognitive processing, this method also possess intrinsic limitations—such as potential interference from the equipment or the artificiality of the experimental setting. Future research could consider incorporating other complementary perceptual measures, such as offline questionnaires, to enrich the results obtained via eye-tracking.

Language, being a living phenomenon in constant change and evolution, can never be fully observed. This reality invites us to pose new questions—some of these have arisen from the final reflections of this research, while others have been present since its earliest

stages. Consequently, rather than viewing these initial findings as a final destination, we see them as an opportunity to continue developing and expanding the scope of analysis of this linguistic variable.

4.1. Future Objectives

Many aspects of this linguistic-variation phenomenon remain to be explored in future work. Therefore, we highlight below some lines of research that we consider especially relevant.

First, we consider it essential to further expand the diachronic dimension of linguistic observation in the city of Málaga. In order to achieve a deeper understanding of changes in pronominal usage in the Spanish spoken in Málaga over time, it would be appropriate to draw on the materials of the *Arqueoling* corpus, which record the voices of the city from the 1940s through the 1990s. A contrastive analysis of this nature would allow us to document historical shifts—for example, in the preference for second-person pronouns, specifically *tú* versus *usted*.

Moreover, thanks to the collection of more recent interviews, working with the data contained in the MAYHO corpus (*Mujer Ayer y Hoy* [“Women: Yesterday and Today”]) would offer the opportunity to investigate whether the social transformations experienced by women in recent decades have had consequences on their language usage. Using this newly assembled corpus, we could confirm whether the indications of increased explicitness of the pronominal subject by women represent the onset of a linguistic change—or even whether this trend toward greater pronominal explicitness might come to constitute one of the characteristic features of the intermediate variety previously identified within the sociolinguistic context of Málaga.

Exploring the influence of individuals’ sociological profiles beyond the pre-stratification variables is particularly promising, as it would allow for a deeper integration of the linguistic and the social. For example, incorporating other indicators—such as the informant’s reading proficiency, cultural interests, non-academic training, motivation to continue learning, or surrounding cultural environment—could form the relevant basis for a reconstruction of an individual’s cultural capital. This approach would yield a more realistic and precise picture than relying solely on formal education level and, in turn,

would allow us to more accurately assess whether speakers' grammatical competence and cultural background have any impact on the variation in presence or absence of the pronominal subject.

Finally, we suggest a detailed acoustic-perceptual study of the relationship between phonetic ambiguity and the expression of the pronominal subject. Although the elision of /-s/ in syllable-coda of verb forms is a widely documented and studied phonetic phenomenon in Málaga, no prior research has considered the influence that the potential loss of the /-s/ segment may have on other elements in the communicative context, such as the pronominal subject. By addressing this from the dual perspective of production and perception, we could obtain a complete picture of the integration of phonetic and syntactic factors during both usage and comprehension. In addition to the acoustic analysis, our intention is to evaluate judges' perceptual recognition of 2nd- and 3rd-person singular verb forms in order to determine whether the presence of the pronominal subject may play a disambiguating role.

This doctoral thesis, by examining the variation in pronominal-subject expression from the perspective of cognitive sociolinguistics, has underscored the importance of considering both the linguistic choices speakers make within their communicative contexts and the cognitive processes triggered in their minds upon receiving linguistic material. This theoretical and methodological proposal—integrating both perspectives—has confirmed the necessity of approaching complex linguistic phenomena from a multidimensional, comprehensive viewpoint capable of encompassing the various facets inherent to the nature of language. Hence, to study language is not only to observe how we speak but also to understand how we construct our social and cognitive reality.