

La lengua hablada en Santander

Corpus PRESEEA-Santander

UAH MONOGRAFÍAS
HUMANIDADES 102

La lengua hablada en Santander

Corpus PRESEEA-Santander

Inmaculada Martínez Martínez

Marta Gancedo Ruiz



Universidad
de Alcalá

EDITORIAL
UNIVERSIDAD DE ALCALÁ

Reservados todos los derechos, ni la totalidad ni parte de este libro puede reproducirse o transmitirse por procedimiento electrónico o mecánico, incluyendo fotocopias, grabación magnética o cualquier almacenamiento de información y sistemas de recuperación sin el permiso previo y por escrito del autor.



Esta publicación es parte del proyecto de I+D+i *Agenda 2050. El español del centro-norte de España. Procesos de variación y cambio espaciales y sociales* (PID2019-104982GB-C51), financiado por el Ministerio de Ciencia e Innovación-Agencia Estatal de Investigación (10.13039/501100011033)

Este libro se publica bajo una licencia Creative Commons Atribución-NoComercial-SinObraDerivada 4.0 Internacional (CC BY-NC-ND 4.0).

© De los textos: sus autores.
© De las imágenes: sus autores.
© De la ilustración de portada: sus autores.

© Editorial Universidad de Alcalá, 2024
Plaza de San Diego, s/n
28801 Alcalá de Henares
www.uah.es

I.S.B.N. electrónico (PDF): 978-84-10432-51-2

Composición: Solana e Hijos, A. G., S.A.U.
Impreso en España

PRESENTACIÓN

1. EL PROYECTO PRESEEA²

1.1. El PRESEEA en Santander

El Proyecto para el Estudio Sociolingüístico del Español de España y América (PRESEEA) nace en abril de 1993, en el *X Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina* (ALFAL). En la Comisión de Sociolingüística de esta asociación se decide iniciar un proyecto que aborde el estudio sociolingüístico de las ciudades iberoamericanas y españolas. Las intenciones generales del proyecto incluyen “llegar a ser algo tan valioso para el futuro conocimiento de la lengua española, como útil para las personas que se ocupan de ella” (Moreno Fernández 2021a:5).

El Proyecto PRESEEA se concreta en la creación de corpus sociolingüísticos sincrónicos de lengua española hablada, representativos del mundo hispánico en su variedad geográfica y social. En la actualidad, agrupa a cerca de 50 equipos de investigación sociolingüística coordinados en torno a una misma metodología –que se detallará más adelante (§ 2)– y a idénticos principios teóricos de carácter sociolingüístico: la concepción del dialecto como propiedad de una comunidad de habla, la variabilidad como rasgo caracterizador de la lengua, la cuantificación como método analítico y la representatividad de las muestras de habla (Moreno Fernández 2021a).

La adhesión al proyecto PRESEEA por parte del equipo de Santander tuvo lugar en noviembre de 2013 con la firma del acuerdo de colaboración que incluye los siguientes compromisos (Moreno Fernández 2021a: 28): seguir las pautas metodológicas generales propuestas para el proyecto; poner los materiales recogidos dentro del proyecto a disposición del resto de los equipos de la red; ofrecer información permanentemente actualizada sobre el desarrollo de la investigación del equipo;

² Esta publicación se inscribe dentro de las actividades del proyecto de I+D+i *Agenda 2050. El español del centro-norte de España: procesos de variación y cambio espaciales y sociales* (PID2019-104982GB-C51), financiado por MICIU/AEI/10.13039/501100011033. Nos gustaría agradecer el inmenso apoyo recibido de la coordinadora del proyecto, la profesora Ana María Cestero Mancera, de la Universidad de Alcalá.

mencionar en todas las publicaciones derivadas de la investigación la vinculación del equipo a PRESEEA.

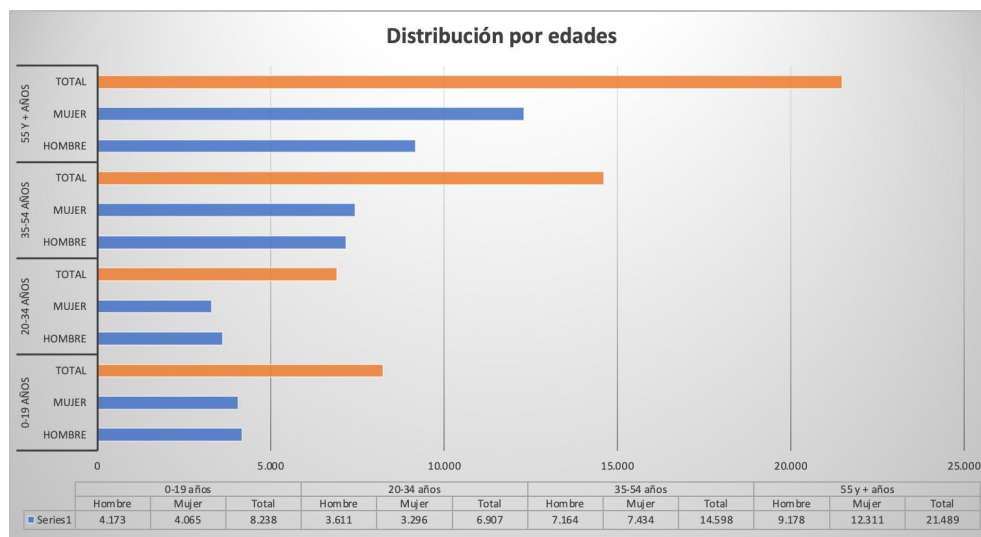
1.2. La ciudad de Santander: caracterización social y lingüística

Santander es una ciudad situada en el norte de España, capital de la comunidad autónoma uniprovincial de Cantabria. Con 173 638 habitantes (2024), es la urbe más poblada de la comunidad autónoma (56,1 % del total de habitantes de Cantabria). Caben destacarse como rasgos identitarios sus límites con la meseta y con el mar, pues ambos determinan en el siglo XIX la verdadera expansión urbana de Santander: es notable el auge del comercio de harinas procedentes de Castilla, así como el aumento del tráfico en el puerto de Santander hasta el punto de llamársele el “Liverpool de España” (Biblioteca Nacional de España, 2024).

La muestra para el estudio sociolingüístico de la ciudad de Santander se ha conformado siguiendo los requisitos mínimos contemplados en la metodología común del proyecto internacional. Entre ellos se destaca, en primer lugar, el hecho de que los núcleos de población en los que se recogen los datos deben corresponderse con núcleos de población asentada: Santander posee conciencia de comunidad de habla acreditada, en torno a la cual surge el hablante santanderino (Martínez y Ueda 2023a). Autores como Peña Arce (2021: 444) señalan que este español “presenta una serie de características que, sin ser privativas de este dominio, dibujan unos contornos lingüísticos que lo personalizan”.

Santander muestra el rostro de una ciudad dedicada al sector terciario, convertida en centro administrativo, comercial, financiero, turístico y cultural; es, por tanto, escasa la implantación industrial dentro de los límites municipales, ya que la mayoría de las empresas han ubicado sus factorías en los municipios vecinos. Es básicamente una ciudad del siglo XX, pues es entonces cuando se construye la mayor parte del tejido urbano actual (Ayuntamiento de Santander 2020: 14). Es, de este modo, una ciudad relativamente joven, en comparación con otras capitales españolas. El proceso de terciarización se ha llevado a cabo a base de cambios morfológicos y demográficos.

Las características de esta estructura demográfica son, principalmente, la escasa proporción de población correspondiente a grupos de edad jóvenes (0-19 años), unida al marcado peso de la población perteneciente al grupo de personas mayores de 55 años; ello provoca un índice de envejecimiento municipal levemente superior a la media regional y de forma clara superior a la media del país (Ayuntamiento de Santander 2020: 221). El relevo generacional, por tanto, corre un riesgo claro, tal y como se manifiesta con nitidez en el Gráfico 1.

GRÁFICO 1. LA POBLACIÓN DE SANTANDER POR EDADES

(Fuente: ICANE a partir de explotación de microdatos del Censo Anual de Población, INE)

La población santanderina se distribuye en 8 distritos censales, tal y como se aprecia en la Imagen 1, el último de los cuales, el 8, aparece integrado por los barrios de Monte, La Pereda, El Alisal y Peñacastillo. Con una función claramente residencial, constituye la última creación del tejido urbano en Santander y reúne el número más alto de habitantes (46356). En el extremo opuesto se sitúa el distrito Centro, el número 1 que, siendo el menos poblado, reúne tan solo a 9441 habitantes (Instituto Cántabro de Estadística, ICANE 2023). Este hecho nos acerca a una configuración urbanística en la que las entidades periurbanas integradas por el distrito 8 abarcan más de la mitad de la superficie del municipio. El movimiento de trasvase e instalación de la población es, pues, “centrífugo en ámbitos colindantes al municipio” (PGOU 2004: 22). La municipalidad actual de Santander presenta, pues, una realidad cambiante, del mismo modo a como lo hace la sociedad que en ella se desarrolla. Estos cambios atañen, así mismo, a la entrada de población extranjera, que supone tan solo el 10% del total de la población de Santander y cuyo perfil mayoritario es mujer, de entre 35 y 54 años. La población que ha migrado a la ciudad se concentra, de mayor a menor porcentaje, en los siguientes distritos: 6, Numancia-San Fernando (15,82 %); 1, Centro (13,73 %) y 5, Barrio Pesquero (13,65 %). El menor índice se da en el distrito 4, correspondiente a la zona de El Sardinero y Puertochico (6,09%), que alberga a los grupos sociales de economía más fuerte.

IMAGEN 1. DISTRIBUCIÓN DE LOS DISTRITOS EN EL MUNICIPIO DE SANTANDER

(Fuente: *El Diario Montañés*)

Si atendemos al nivel educativo, podemos comprobar en el Gráfico 2 cómo el mayor número de personas se agrupa en torno a la educación superior (59 049), que integra, entre otros, los títulos propios universitarios, de duración igual o superior a 2 años, los grados universitarios de 240 créditos ECTS, diplomados universitarios, títulos propios universitarios de experto o especialista, así como los licenciados, másteres, especialidades en Ciencias de la Salud por el sistema de residencia y similares. A esta población le sigue en número la que ha cursado la primera etapa de educación secundaria (38 715) y la segunda etapa de educación secundaria, bien con orientación general o bien con orientación profesional. El porcentaje más bajo se corresponde con la educación primaria e inferior (19 029) (Instituto Cántabro de Estadística 2023).

Podemos intuir, a tenor de los datos, que nos encontramos ante un núcleo urbano en el que la población dispone de un nivel educativo alto y, en el contexto regional, por encima de la media española: la población adulta (entre 25 y 64 años) en Cantabria presenta un porcentaje de finalización de estudios superiores a la media española (45,6 %, frente al 41,1 % de la media correspondiente a todo el territorio nacional). En esto, el núcleo urbano de Santander tendrá, sin duda, un peso porcentual importante³.

³ <https://www.educacionyfp.gob.es/gl/mc/redie-eurydice/sistemas-educativos/contexto/nivel-formacion-poblacion-adulta.html> . Red europea de Información sobre Educación (Eurydice España- rediE)

GRÁFICO 2. LA POBLACIÓN DE SANTANDER POR NIVELES EDUCATIVOS

(Fuente: Instituto Cántabro de Estadística a partir de Censo de Población 2023, INE)

Cabe destacarse, para concluir esta caracterización sociolingüística, los rasgos que determinan a la ciudad de Santander como un núcleo urbano poseedor de una fuerte conciencia de habla santanderina, tal y como se señaló al comienzo de este epígrafe. Es palpable, así mismo, cierta homogeneidad, debida a la ausencia de mezclas lingüísticas: su situación geográfica en el punto central de la provincia y mirando al mar ha provocado que no se produzcan trasvases lingüísticos significativos desde las comunidades limítrofes (Asturias y País Vasco) (Martínez y Ueda, 2023a).

2. METODOLOGÍA

2.1. Criterios de selección de informantes

Como parte integrante del PRESEEA, el equipo de Santander se ha planteado como objetivo amplio reunir materiales que permitan estudiar el español actual de una zona determinada, el norte de España, más específicamente, Cantabria. Para concretar la muestra de estudio, en primer lugar, se han seguido las directrices apuntadas en el documento Metodología del Proyecto para el estudio sociolingüístico del español de España y América (PRESEEA) (Moreno Fernández, 2021a), de modo que el conjunto del material oral compilado procede de informantes que se integran en un núcleo asentado con una comunidad de habla a la cual se sienten adscritos.

Asimismo, partiendo de que el proyecto se incardina en la línea de la sociolingüística variacionista y, por lo tanto, se enfoca en el estudio de comunidades pertenecientes a núcleos urbanos, el corpus se ha limitado a recoger materiales de hablantes de la ciudad de Santander. Si bien la comunidad de habla estudiada puede trascender los límites geográficos de este núcleo urbano, dadas las características descritas arri-

ba (§ 1.2), el corpus PRESEEA-Santander está formado por hablantes santanderinos nacidos en la capital o residentes en ella desde hace más de 20 años.

Una vez delimitada la pertenencia de los informantes seleccionados a la ciudad de Santander, el segundo paso ha sido procurar la representatividad de la muestra de modo que esta sea válida para su estudio y comparación sociolingüísticos. Para tal fin, se han seguido las indicaciones del proyecto PRESEEA, en cuya base se encuentra la consideración de García Ferrando (1994) sobre la necesidad de tener en cuenta que la configuración de la muestra debe depender no tanto del tamaño de la comunidad de habla, sino de su complejidad social. De esta manera, siguiendo el espíritu general de todos los equipos del proyecto PRESEEA, se ha trabajado por cuotas con afijación uniforme. Este sistema supone preestablecer unas variables sociales determinadas y dividir en función de ellas el universo relativo de informantes en subpoblaciones, estratos o cuotas. Como sostiene Moreno Fernández (1996, 2021a) se prefiere este sistema de muestreo y no la muestra aleatoria o probabilística, ya que las cuotas permiten la comparación estadística entre los estratos internos, tanto de la propia muestra como de otra diferente; igualmente, las cuotas con afijación interna “garantizan la representación de hablantes de los perfiles sociales considerados como fundamentales” (Moreno Fernández, 2021a: 13).

Como en todos los corpus que se recogen al abrigo del proyecto PRESEEA, y como una decisión avalada por la bibliografía previa de estudios sobre sociolingüística, en PRESEEA-SANTANDER se ha llevado a cabo la preestratificación atendiendo a tres variables sociales (Moreno Fernández, 1996, 2021a; Cestero Mancera, 2012; entre otros;): sexo, edad y nivel de instrucción.

En cuanto a la variable social sexo, la muestra ha quedado preestratificada en dos grupos: hombres y mujeres. Con respecto a la posibilidad de contar con hablantes adscritos a géneros no binarios, pese a no contar con ningún informante de estas características, se ha tomado la decisión de que, llegado el caso, se propondrá la autoadscripción del hablante a una categoría.

En cuanto a la variable social edad, la muestra se ha preestratificado en tres grupos etarios:

- Grupo 1: hablantes de edades comprendidas entre los 20 y los 34 años.
- Grupo 2: hablantes de edades comprendidas entre los 35 y los 54 años.
- Grupo 3: hablantes de edades superiores a los 55 años.

En lo relativo a la tercera variable social de preestratificación, el nivel educativo o nivel de instrucción, se han diferenciado tres cuotas:

- Grupo 1. Grado de Instrucción 1: hablantes analfabetos, sin estudios, o que solo hayan completado la educación primaria.
- Grupo 2. Grado de Instrucción 2: hablantes que solo hayan completado la educación secundaria o enseñanzas técnicas medias. Se contemplan en este grupo

informantes que hayan estado escolarizados, aproximadamente, hasta los 18 años.

- Grupo 3. Grado de Instrucción 3: hablantes que hayan cursado, al menos, un año de educación superior (ya sean estudios universitarios o ciclos superiores de formación profesional).

En cuanto al tamaño completo de la muestra, dado que Santander cuenta con un número de habitantes inferior a 500 000, se ha optado por incluir tres informantes por cuota. De este modo, se cuenta con una muestra total de 54 informantes (lo que permite una representatividad de 1 / 9250).

Así pues, la muestra de Santander se compone de entrevistas realizadas a 54 encuestados, distribuidas, como puede verse en la Tabla 1, por cuotas fijas de acuerdo con las variables sociales preestablecidas de sexo, edad y nivel de instrucción.

TABLA 1. DISTRIBUCIÓN POR CUOTAS DE LOS INFORMANTES

	Generación 1		Generación 2		Generación 3	
	H	M	H	M	H	M
Nivel de instrucción 1	3	3	3	3	3	3
Nivel de instrucción 2	3	3	3	3	3	3
Nivel de instrucción 3	3	3	3	3	3	3

Además, cabe señalar que, dado que cada informante cumplimenta un cuestionario con información adicional, la muestra, en un proceso posterior, como sugieren Cestero Mancera, Molina Martos y Paredes García (2019), podría ser postestratificada atendiendo a variables de diversa índole como ingresos económicos, profesión o modo de vida. Se concluye, pues, que, al combinar la información ofrecida por estas variables con los datos recogidos durante las conversaciones, se haría posible, además, trabajar con la variable postestratificada del nivel sociocultural.

2.2. Metodología para la recogida de materiales

2.2.1. Tipo de cuestionario: la entrevista semidirigida

En el proyecto sociolingüístico de PRESEEA-SANTANDER, el objetivo general delimitado ha sido la recogida de materiales que proporcionaran información suficiente sobre el habla de los distintos estratos sociales de la ciudad. Para ello, ha sido fundamental seleccionar un cuestionario de recogida de datos adecuado para el estudio sociolingüístico del español de Santander. De acuerdo con este objetivo, la herramienta empleada para la obtención de los materiales lingüísticos ha sido la en-

trevista semidirigida empleada de manera coordinada por los diferentes equipos de trabajo que integran el proyecto. Como bien han señalado numerosos sociolingüistas, este instrumento de recogida de información, iniciado por Labov (1983, 1996), permite obtener gran cantidad de información variada y estratificada (gracias a las cuotas estratificadas) en poco tiempo y con un despliegue razonable de medios (solo es necesario un dispositivo de grabación) (López Morales, 1994; Recalde y Vázquez Rozas, 2009).

En concreto, la entrevista sociolingüística semidirigida, método empleado para la compilación del conjunto de corpus PRESEEA, es un intercambio comunicativo entre dos interlocutores (investigador e informante), *a priori* desconocidos, planificado por el entrevistador, aunque con cierto margen de laxitud. En todo momento, el informante es conocedor de que sus respuestas están siendo grabadas por un magnetófono u otro dispositivo de grabación, aunque ignora que el objetivo de la investigación del entrevistador sea lingüístico, para evitar cualquier adulteración de los datos; se trata de minimizar, así, la conocida como paradoja del observador laboviana (Labov, 1972, 1984).

Así pues, la finalidad de este tipo de acto es obtener información del encuestado, momento en que el entrevistador debe velar por la producción natural de material lingüístico por parte del entrevistado, lo que responde al fin fundamental del proceso de recogida de datos planteado: generar material suficiente para describir el idiolecto del informante (Calderón y Alvarado, 2011) y, por extensión, de un representante del estrato social estudiado de la ciudad analizada.

Así pues, si bien el reparto de roles epistémicos en este tipo de entrevista es asimétrico (en tanto que el entrevistador ostenta el rol de poder), el papel fundamental del entrevistador-investigador queda limitado a procurar que se produzcan los segmentos discursivos necesarios para el análisis de las muestras; en este sentido, su discurso ha de ser cuantitativamente inferior al del informante. De manera especial, ha de procurar la comodidad del entrevistado para que sus respuestas sean lo más naturales posibles y representativas de su dialecto vernáculo.

Además, como ya apuntaron Cestero Mancera, Molina Martos y Paredes García (2019), la entrevista semidirigida ofrece a la comunidad investigadora la posibilidad de analizar todos los niveles lingüísticos: el fonético-fonológico, el léxico-semántico, el morfosintáctico y el pragmático-discursivo.

En cuanto al estudio del plano fonético-fonológico, la entrevista semidirigida, como cualquier otro tipo de encuesta que pueda ser reproducida en un magnetófono y posteriormente transcrita, ofrece datos suficientes para desarrollar investigaciones en este nivel. Esto se debe a que, como señala López Morales (2004) y recuerdan Cestero Mancera, Molina Martos y Paredes García (2019: 19), el plano fonético-fonológico, al “tratarse de un número limitado y relativamente escaso de unidades que pertenecen a inventarios cerrados [...] presentan un índice de frecuencia muy alto en el habla espontánea” y, por lo tanto, su presencia en los registros que se materializan en una entrevista semidirigida serán suficientes para analizar los segmentos

fonéticos y fonológicos que sean de interés para caracterizar las variedades sociolingüísticas, en este caso, del español.

En lo referente al estudio del nivel léxico semántico, si bien la aplicación de herramientas organizadas en torno a campos semánticos puede parecer el método más adecuado, la entrevista semidirigida también ofrece la oportunidad de obtener información léxica, como dialectalismos, préstamos de otras lenguas, unidades neológicas o cultismos.

En tercer lugar, como ya lo han advertido Cestero Mancera, Molina Martos y Paredes García (2019), el estudio sociolingüístico del nivel morfosintáctico presenta mayores dificultades que los anteriores, pues aparecen obstáculos a la hora de obtener el rendimiento adecuado de las técnicas morfosintácticas más utilizadas. Además, como señalan estos investigadores, si bien para la fonética, la fonología e incluso la semántica hay cierto consenso en el ámbito hispánico sobre qué segmentos tienen una correlación lingüística y social, no existe tal acuerdo en el ámbito morfosintáctico. En el caso del proyecto PRESEEA, los equipos de la Comunidad de Madrid, pioneros y diseñadores del guion de la encuesta semidirigida, apoyándose en los trabajos de Silva Corvalán (1994) y Martín Butragueño (1994), desarrollaron este recurso de modo que hiciera posible la aparición, entre otras cuestiones, de los aspectos morfosintácticos que se enumeran a continuación (Cestero Mancera, Molina Martos y Paredes García, 2019: 19-20):

1. Expresión y posición de sujeto, nominal y pronominal;
2. Orden de los argumentos;
3. Las formas de tratamiento, con especial atención a los usos de tú/usted y vosotros/ustedes;
4. Valores y usos del sistema verbal;
 - 4.1. Indicativo-subjuntivo/condicionales;
 - 4.2. Presente/pretérito con valor de pasado;
 - 4.3. Presente con valor de futuro;
 - 4.4. Indicativo/subjuntivo en oraciones subordinadas introducidas por las partículas *para que*, *cuando...*;
5. Verbos semiauxiliares de modalidad epistémica, dinámica y deóntica: *haber de/que*, *tener que*, *deber (de)*, *poder*, *ser capaz de*;
6. Variación en el uso de construcciones de tipo “reflexivo”;
7. Usos de los verbos *ser* y *estar*;
8. Usos de los verbos *haber* y *estar*;
9. Estilo directo e indirecto;
10. Deícticos pleonásticos;
11. Leísmo, laísmo, loísmo;
12. Duplicación de clíticos;
13. Posición de clíticos;
14. Posición de OD y OI;

15. Concordancia del verbo con usos impersonales;
16. Impersonalidad; cómo se marca (*se, tú*);
17. Uso de los verbos *haber* y *hacer* (“hay”/“hace niebla”...);
18. Perífrasis.

Por último, la entrevista semidirigida también se muestra muy productiva para analizar cuestiones relacionadas con el nivel pragmático-discursivo. Así pues, los rasgos situacionales (solidaridad reducida entre hablantes y distancia comunicativa), así como la finalidad discursiva, dominada por el interés del entrevistador para llegar a acuerdos y facilitar respuestas naturales, propician situaciones en las que afloran las estrategias discursivas orientadas a gestionar la dimensión social de la comunicación. Dichas estrategias se relacionan, tanto con cuestiones de cortesía, estrategias pragmáticas como la atenuación, la intensificación o la ironía, como con aspectos vinculados a la gestión de turnos o la expresión de la epistemicidad, entre otros. Además, por lo general, este tipo de herramientas permite recoger discursos en un registro formal o semiformal (dominante al inicio de la entrevista) combinados con fragmentos más cercanos al registro conversacional o coloquial.

Así pues, la entrevista semidirigida se muestra como una herramienta de recogida de materiales lingüísticos válida para estudiar los cuatro niveles de la lengua. No obstante, será pericia del entrevistador guiar la conversación para obtener la manifestación de los segmentos discursivos que el análisis requiere.

2.2.2. *El diseño de la entrevista*

Para lograr la homogeneidad de los materiales lingüísticos recogidos por los diferentes equipos del proyecto PRESEEA, y poder así llevar a cabo estudios coordinados y contrastivos (Moreno Fernández 2021c), todos los corpus incluidos, también el de Santander, han sido confeccionados con base en un guion. El orden y desarrollo de dicho guion podrán ser adaptados por el entrevistador a medida que se va desarrollando la conversación. Por último, el propio guion estructura la entrevista teniendo en cuenta los siguientes módulos temáticos:

- 1 Saludos
- 2 El tiempo
- 3 Lugar donde vive
- 4 Familia y amistad
- 5 Costumbres
- 6 Peligro de muerte
- 7 Anécdotas importantes en la vida
- 8 Deseo de mejora económica
- 9 Final.

La inclusión de esta diversidad de asuntos pretende favorecer la producción de un número importante de fenómenos lingüísticos de naturaleza variable en los cuatro niveles previamente mencionados.

El esquema que han seguido los entrevistadores del equipo PRESEEA-Santander sigue las indicaciones del equipo de Alcalá de Henares y Madrid, cuyas orientaciones pueden encontrarse, entre otros, en Cestero Mancera, Molina Martos y Paredes García (2019), Moreno Fernández (1996, 2021a). A continuación, reproducimos el esquema facilitado a los entrevistadores:

1. SALUDOS

Presentación.

¿Cómo quiere que le trate, de tú o de usted?

La verdad es que es un problema esto del tratamiento, nunca sabes cómo tratar a las personas ¿verdad? Por ejemplo, ¿usted /tú cómo trata(s) a /sus tus amigos, de tú o de usted? ¿Y si son personas mayores? ¿Si se trata de alguien joven al que no conoce(s)? ¿Y si es una persona mayor, hombre o mujer, al que por ejemplo le pregunta(s) por una calle? ¿Y a su/ tu médico? ¿A los vecinos con los que no tiene(s) mucho contacto? ¿Y cómo te/le gusta que te/le traten a ti/usted? Si una persona más joven te/le trata de tú qué te/le parece? (Preguntar por si tratan de evitar el trato asimétrico o en qué situaciones les parece mejor).

Es un problema, yo a veces no sé qué hacer.

Bueno, a mí trátame de tú ¿le/te parece?

¿Cómo estás/está? ¿Estás animado/a?

En estos días todos estamos un poco raros, yo creo que es por el tiempo, ¿verdad?

2. EL TIEMPO

¡Qué calor/frío ha hecho hoy!

A mí no me gusta el verano/invierno ¿Tú cuál prefieres?⁴

Este año ha hecho más calor/ha llovido más que el año pasado, ¿verdad?

Yo pienso que está cambiando el tiempo, al menos por esta zona, ¿tú qué opinas?

¿Por qué será?

¿Recuerdas el tiempo que hizo el año pasado por estas fechas, lluvia, frío, viento, tormentas...? ¿Y en invierno/verano?

Dicen que está cambiando el clima en la Tierra ¿qué crees que va a pasar si deja de llover y continúa la sequía en los próximos años?

⁴ Según lo respondido en el bloque 1, se procederá a emplear la forma de tratamiento “pactada”, aunque de ahora en adelante se use únicamente el “tú” para facilitar la lectura.

3. LUGAR DONDE VIVE

¿Dónde vives? ¿Cómo es tu casa? Descríbemela un poco. ¿Qué es lo que más te gusta de ella? ¿Por qué? ¿Ha estado siempre así o has hecho reformas? ¿Cómo era antes?

¿Llevas mucho tiempo viviendo allí? ¿Dónde vivías antes? (Descripción del lugar).

¿Te gusta vivir aquí? ¿Dónde te gustaría vivir? ¿Por qué?

¿Cómo crees que sería vivir en Madrid?

¿Conoces a mucha gente por aquí? ¿Qué tal te llevas con ella? ¿Cómo te gustaría que fueran tus vecinos? ¿Qué relación te gustaría tener con ellos? ¿Qué haces tú para mantener o mejorar esa relación?

Supongo que este barrio habrá cambiado desde que viniste a vivir aquí. El mío también ha cambiado, la gente, el barrio en sí, recuerdo cuando era pequeña...

¿Cómo ha cambiado el tuyo? ¿Qué tiene ahora que no tenía antes? ¿Te gusta más ahora o antes? ¿Qué recuerdas tú de cuando eras pequeña?

Si cada barrio ha cambiado, la ciudad ha cambiado mucho más. ¿Recuerdas cómo era? ¿Qué diferencias ves ahora? ¿Crees que son buenos los cambios acontecidos? ¿Por qué? ¿Cómo te hubiera gustado que fuera? ¿Cómo crees que será dentro de unos años?

Ahora al menos hay más sitios para salir y divertirse un poco, ¿no? ¿Qué haces tú cuando sales? ¿Qué te parece tu ciudad con respecto a las diversiones? ¿Qué te gustaría que tuviera? ¿Qué harías si pudieses organizar cosas en Alcalá para que la gente se divirtiera?

¿Y la gente de otras edades, tus padres, tus hijos..., qué suele hacer cuando sale? Sin embargo, la gente cada vez sale menos por la noche debido al miedo a que les ocurra algo desagradable (atracos, violaciones, peleas...). ¿Cómo ves la delincuencia en la ciudad? ¿Y en tu barrio? ¿Qué crees que se puede o debe hacer para acabar con ella?

¿Has oído algo sobre algún acto de delincuencia de tu barrio o ciudad? ¿Qué se dice de lo que pasó?

4. FAMILIA Y AMISTAD

¿Estás casada/o? ¿Cómo conociste a tu marido/mujer? ¿Tienes hijos/piensas tener hijos? Cuéntame cómo fue. ¿Qué crees que hubieras hecho si corriera peligro tu vida o la del niño? (En caso de ser hombre, preguntar qué habría aconsejado a su mujer, hermana...). ¿Qué piensas de problemas actuales como, el aborto, la anorexia de los jóvenes o la eutanasia?

¿Cómo es tu marido/tu hijo/tu padre...? (Descripción física)

¿Cómo te hubiera gustado que fueran?

¿Quién es tu mejor amigo? ¿Qué es para ti un amigo? ¿Cómo debe ser una persona o qué debe hacer para que la consideres un buen amigo? ¿Son iguales los amigos que se tienen de niños que los que se tienen de adultos? ¿Cómo eran los que tú tenías de niño?

¿A qué te dedicas?

¿En qué te gustaría trabajar/qué vas a estudiar?

(En caso de tener ya una profesión definida) ¿Es eso lo que siempre has querido hacer realmente? ¿Por qué no lo hiciste? ¿Cómo te imaginas tu vida si hubieras sido.../si hubieras hecho...?

¿Qué sueles hacer en un día normal? (Descripción de un día normal de su vida desde que se levanta hasta que se acuesta)

¿Estás contenta con tu forma de vida? ¿Por qué?

¿Qué estará haciendo tu familia/tu marido/tu hijo/tu padre ahora?

5. COSTUMBRES

Ahora se aproximan las vacaciones, ¿qué sueles hacer en navidades/verano? ¿Qué harás las próximas navidades/verano?

Las otras grandes vacaciones que todos esperamos son las navidades, ¿verdad? Las navidades se caracterizan por la comida y por la reunión de la familia. ¿Cuál es la comida típica de navidad? ¿Cómo se hace? ¿Eres buen/a cocinero/a? ¿Qué crees que es necesario para ser un buen cocinero? ¿Cuál es la comida típica de Alcalá? ¿Cómo se hace? ¿Por qué empleas tal ingrediente? ¿Cuándo se añade?

Antes la Navidad era una fiesta religiosa, ¿crees que sigue siéndolo o que ya se ha perdido y ahora es solo una fiesta popular?

¿Qué planes tienes para las próximas navidades/verano?

Si pudieras elegir, ¿qué te gustaría hacer/qué harías estas navidades/el próximo verano?

6. PELIGRO DE MUERTE

Siempre da un poco de miedo salir de viaje, por los accidentes que ocurren y las malas noticias que se oyen a diario, ¿verdad? ¿Has estado tú alguna vez en peligro de muerte? ¿Qué ocurrió? (Descripción física de personajes y lugares de la narración). ¿Qué hubiera pasado si...? ¿Y si...? ¿Qué harías si volvieras a estar en una situación parecida?

(Preguntar por lo que dijo alguna de las personas que intervinieron en el suceso).

7. ANÉCDOTAS IMPORTANTES EN LA VIDA

Cuéntame más cosas importantes o curiosas que te hayan ocurrido, algún robo, algún premio, algún viaje especial...

¿Qué crees que se debe hacer en una situación como esa? ¿Qué harías si te volviera ocurrir o si tuvieras la oportunidad de volver al pasado?

8. DESEO DE MEJORA ECONÓMICA

¿Juegas a la lotería o a cualquier otro juego de azar? ¿Por qué?

Supongo que te gustaría que te tocara algún premio gordo ¿verdad? ¿Qué harías si te tocaran 200 millones de pesetas?

¿Qué piensas/haces cuando oyes que a alguien le ha tocado una gran cantidad de dinero?

9. FINAL

En fin... yo creo que todos nos podemos considerar con suerte, ¿verdad? Ahora, a seguir disfrutando los momentos buenos de la vida.

Ahora ya voy para casa ¿cómo voy desde aquí a una boca de metro / cómo voy desde aquí al intercambiador?

Bueno, me ha encantado poder charlar contigo un rato, muchas gracias por todo. Espero que volvamos a repetirlo, ¿eh?

Adiós.

Se pretende, así, que la disposición de estos módulos temáticos proporcione, en las intervenciones de los informantes, las modalidades discursivas que se muestran en la Tabla 2.

TABLA 2. MÓDULOS TEMÁTICOS Y MODALIDADES DISCURSIVAS

Módulo	Modalidad discursiva registrada
1.- Saludos	Discurso dialógico (pregunta-respuesta)
2.- El tiempo	Exposición, descripción
3.- Lugar donde vive	Descripción, argumentación
4.- Familia y amistad	Narración, descripción, argumentación
5.- Costumbres	Narración, descripción, argumentación, exposición, instrucciones
6.- Peligro de muerte	Narración
7.- Anécdotas importantes en la vida	Narración, exposición, discurso referido
8.- Deseo de mejora económica	Discurso dialógico, discurso hipotético
9.- Final	Discurso dialógico

Fuente: extraído de Cestero Mancera, Molina Martos y Paredes García, 2019: 25

2.2.3. La recogida de materiales

Los datos que componen el corpus PRESEEA-Santander están formados por 54 entrevistas de una duración aproximada de 45 minutos cada una. Los datos se recogieron entre agosto de 2014 y enero de 2024. En las grabaciones participaron tanto profesores como estudiantes egresados con intereses investigadores del Centro Internacional de Estudios Superiores del Español Fundación Comillas (CIESE-Comillas)⁵. Para la grabación de las entrevistas, la interacción se desarrolló entre dos tipos de participantes:

1. Entrevistador: el rol de entrevistador ha sido desempeñado por 10 colaboradores del CIESE-Comillas, todos profesores o estudiantes egresados con intereses investigadores en sociolingüística. Para todas las entrevistas se ha seguido el criterio de que el entrevistador y el informante fueran del mismo sexo. Los entrevistadores, además, recibieron formación previa sobre el contenido de la entrevista y los objetivos del proyecto PRESEEA. Así, el entrevistador, en todos los casos, ha podido velar por que en cada entrevista se abordaran, de manera fluida y natural, los módulos temáticos planteados en el guion, para poder así suscitar, en el discurso del informante, aquellos aspectos lingüísticos susceptibles de variación sociolectal.

⁵ Nos gustaría expresar nuestro profundo agradecimiento a todos los informantes, entrevistadores y colaboradores que participaron desinteresadamente en este proyecto, tanto en la realización de las entrevistas como en las transcripciones y en la compleja búsqueda de informantes. Queremos destacar, especialmente, la ingente ayuda proporcionada por el personal y dirección del Centro de Educación de Personas Adultas de Santander (CEPA Escuelas Verdes), sin cuya colaboración hubiera sido imposible concluir el corpus.

2. Informante: los informantes fueron seleccionados de manera previa para cubrir las cuotas fijas en función de dos aspectos:
 - Procedencia geográfica: todos los informantes han tenido que haber nacido y vivido en Santander, haber llegado a la ciudad antes de los 10 años o, en último caso, llevar viviendo en este núcleo urbano más de veinte años (siempre que su origen lingüístico no sea marcadamente diferente).
 - Características sociales: sexo, edad y nivel de instrucción. Para completar las cuotas fijas, se buscaron tres informantes para cada cuota de afijación que cumplieran con las tres características preestablecidas para cada una de ellas.

Las grabaciones se realizaron bien con una grabadora digital (modelo (Olympus WS-81), bien con los dispositivos móviles de los entrevistadores. En todos los casos, el informante conocía que iba a ser grabado y el dispositivo de grabación permanecía a la vista y cerca de los interlocutores. En todas las entrevistas se ha observado que, si bien se inician con cierto grado de formalidad y tensión, lo que motiva cierto cuidado del registro en el discurso, los informantes, gracias a las intervenciones de los entrevistadores, van adquiriendo mayor naturalidad y terminan prácticamente olvidando la presencia de la grabadora.

Las entrevistas se llevaron a cabo en lugares heterogéneos: casa de los informantes, lugar de trabajo, salas de trabajo en la biblioteca de la Universidad o parques, entre otros. En todo caso, fueron lugares consensuados con los informantes de manera que ellos no se sintieran coartados y el devenir conversacional permitiera un progresivo relajamiento del ambiente y de la conversación, con el fin de hacer aflorar el idiolecto natural del entrevistado.

2.3. Transcripción y etiquetado del corpus

La fase de grabación, transcripción y etiquetado del corpus de PRESEEA-Santander, con sus 54 entrevistas, ya ha sido completada. En una primera etapa, se completó la recogida, transcripción y etiquetado de la muestra del corpus (formada por 18 entrevistas), es decir, de una entrevista para cada una de las cuotas de informantes, tal y como se señala en la Tabla 3:

TABLA 3. CONFIGURACIÓN DE LA MUESTRA DEL CORPUS PRESEEA-SANTANDER

	Generación 1		Generación 2		Generación 3	
	H	M	H	M	H	M
Nivel de instrucción 1	1	1	1	1	1	1
Nivel de instrucción 2	1	1	1	1	1	1
Nivel de instrucción 3	1	1	1	1	1	1

Esta fase fue terminada en abril de 2020 y los materiales fueron puestos a disposición de la comunidad investigadora a través del portal web oficial del proyecto PRESEEA⁶. Se priorizó esta tarea porque es el compromiso que se establece con el proyecto global, la entrega de la muestra de 18 informantes. A continuación, se procedió con la transcripción y etiquetado del resto del corpus.

Todo el proceso de anonimización, transcripción y almacenamiento de las entrevistas ha seguido de manera directa las indicaciones establecidas por Moreno Fernández en las instrucciones originales de 1996 y en todas sus actualizaciones, la última, en Moreno Fernández (2021a, 2021b).

De este modo, las grabaciones se han transcrito⁷ de manera íntegra empleando ortografía ordinaria con el procesador de textos *Word*™. Para ello, el equipo de transcriptores ha reproducido las grabaciones en formato digital, bien manejándolas con un pedal de transcripción, bien valiéndose de *software* que facilita la gestión de la grabación para la transcripción como *Express Scribe*.

Para reflejar los diferentes rasgos de los niveles lingüísticos y su variabilidad, la transcripción de las entrevistas ha respetado las normas internacionales de la TEI (*Text Encoding Initiative*); por otro lado, se han empleado una serie de marcas y etiquetas consensuadas por los diferentes equipos del proyecto PRESEEA (las cuales pueden encontrarse en Moreno Fernández 2021b y serán reproducidas en el siguiente epígrafe).

Para garantizar que las transcripciones reflejen en el mayor grado posible lo registrado en las grabaciones, cada texto ha sido revisado y contrastado con la grabación por dos personas diferentes.

3. LOS MATERIALES DE LA LENGUA HABLADA EN SANTANDER

3.1. Presentación de los materiales

Los materiales que conforman el corpus PRESEEA-Santander se encuentran custodiados informáticamente por la coordinación del equipo. No obstante, como ya se ha indicado líneas más arriba, una muestra representativa del corpus, formada por 18 encuestas, está alojada en el portal electrónico del proyecto PRESEEA y puede consultarse y descargarse en <https://preseea.uah.es/corpus-preseea>.

Pese a que el resto de los materiales no se halla en acceso abierto, el corpus completo del español de Santander se pone a disposición de aquellos investigadores

⁶ Puede consultarse en la dirección <https://preseea.uah.es/corpus-preseea>

⁷ Nos gustaría dar las gracias a los transcriptores que han hecho posible el que los materiales de la muestra PRESEEA-Santander se encuentren a disposición de la comunidad científica internacional. De manera especial, queremos agradecer a la coordinadora del proyecto, la profesora Ana María Cestero, la inmensa ayuda recibida para concluir esta tarea.

miembros de otros equipos del proyecto PRESEEA que deseen consultarlos para llevar a cabo estudios de acuerdo con las líneas estratégicas y metodológicas del macroproyecto. Sin embargo, guiados por el mismo espíritu que el equipo de Madrid (Cestero Mancera, Molina Martos y Paredes García, 2019) y gracias a esta publicación, los materiales de PRESEEA-Santander se ponen a disposición de otros perfiles de profesores e investigadores que necesiten acceder a material lingüístico sociolectal de esta área geográfica.

En la edición del corpus PRESEEA-Santander se siguen los mismos criterios presentes en la publicación de los corpus predecesores de Alcalá de Henares (Moreno, Cestero, Molina y Paredes 2002, 2005 y 2007), Madrid (distrito de Salamanca) (Cestero, Molina y Paredes 2012; Molina, Paredes y Cestero 2014; Paredes, Cestero y Molina 2015) y Madrid (distrito de Vallecas) (Cestero, Molina y Paredes 2019), si bien se han adaptado las convenciones pertinentes a la actualización de los criterios y documentos proporcionados por la coordinación del proyecto PRESEEA.

En este volumen, se presenta todo el material del corpus transcrito, por lo que pueden encontrarse las entrevistas de los 54 informantes, organizadas de acuerdo con el nivel de instrucción, la edad y el sexo de los entrevistados. El orden que se sigue en la sucesión de las transcripciones se corresponde con el nivel de instrucción de mayor a menor (superior, medio y baja). El perfil de todos los informantes aparece detallado en las tablas 4, 5 y 6 (§ 3.2).

A continuación, siguiendo el modelo de Cestero Mancera, Molina Martos y García Paredes (2019: 28-29), se listan las etiquetas presentes en los textos marcados y sus significados, disponibles en Moreno Fernández (2021b):

TABLA 4. ETIQUETADO EN PRESEEA-SANTANDER

ETIQUETADO	
Ortografía y puntuación	
[¡!]	Enunciados exclamativos
[¿?]	Enunciados interrogativos
[/]	Pausa mínima
[//]	Pausa
Ruidos	
<ruido = « »/>	Ruido, con especificación del tipo de ruido del que se trate
<ruido_de_fondo> </ruido_de_fondo>	Emisión que se produce con ruido de fondo
<registro_defectuoso> </registro_defectuoso>	Fragmento de grabación de mala calidad
<interrupción_de_grabación/>	Interrupción de la grabación
<risas = “ ”/>	Risas, con la especificación del interlocutor que la produce
<entre_risas> </entre_risas>	Emisión que se produce entre risas.
Etiquetado fónico	
<ininteligible/>	Discurso que no se ha podido entender
<énfasis> </énfasis>	Fragmento emitido con pronunciación enfática
<alargamiento/>	Alargamiento de sonidos
<silencio/>	Ausencia de emisión durante un segundo o más
<palabra_cortada>	Palabra no se ha emitido completamente
<vacilación/>	Vacilación en la emisión
<sic> </sic>	Emisión del informante transliterada: no hay descuido en la transliteración
Etiquetado léxico	
<término> </término>	Lexía utilizada como uso especializado
<extranjero> </extranjero>	Extranjerismos
<siglas = []> </siglas>	Siglas. Incluye la pronunciación entre corchetes
Etiquetado de dinámica discursiva	
<simultáneo> </simultáneo>	Superposición de discurso del entrevistador y del informante. Emisión simultánea.
<cita> </cita>	Cita o producción en estilo directo
Etiquetado de transcripción	
<transcripción_dudosa> </transcripción_dudosa>	Fragmento de discurso dudoso para el transcriptor y los revisores
<tiempo = “ ”/>	Proporciona información del minuto y segundo de grabación (se constata cada dos minutos aproximadamente).
<observación_complementaria = “ ”/>	Recoge cualquier observación que no haya sido prevista en el resto de etiquetas y que el transcriptor o revisor considere pertinente.

En las transcripciones, los interlocutores se identifican a través de una letra por su función: el entrevistador será identificado por una E y el informante a través de la I. De esta manera, cada vez que uno de los interlocutores inicie turno (cada vez que tome la palabra, aunque solo sea un turno de apoyo), se comenzará en párrafo aparte y se indicará quién habla con la E o la I seguida, de dos puntos.

Para garantizar el anonimato, se sustituye de forma sistemática y congruente los nombres propios de los informantes, así como los de terceras personas o lugares que posibiliten en alguna medida la identificación de los entrevistados. Se sustituyen por iniciales, sin puntos detrás.

Otras consideraciones prácticas tenidas en cuenta para la transcripción han sido las mismas contempladas por el grupo de PRESEEA-Madrid (distrito de Vallecas) (Cestero, Molina y García, 2019):

- No se emplean mayúsculas, salvo en nombres propios, iniciales y siglas.
- No se emplean signos de puntuación. Las pausas y silencios se indican con las marcas y etiquetas ya comentadas.
- Los nombres propios se transcriben con su grafía original.
- Los elementos cuasi-léxicos, como las interjecciones, onomatopeyas o sonidos varios (*ay, zas, mm, eeh, pff*, etc.) se recogen en escritura ortográfica.

Cada una de las entrevistas está identificada por un código que sigue los estándares del proyecto PRESEEA. Así, para identificarlas, cada una de ellas va encabezada por un apartado de presentación, en el que figura el número de la entrevista en la publicación (ejemplo: Entrevista N.º 01), seguido de una línea con los datos de caracterización del informante (sexo, edad y estudios) y la fecha de grabación.

Por su parte, la versión con etiquetas de cada una de las entrevistas va precedida de una cabecera. En ella, siguiendo el modelo de Cestero, Molina y Paredes (2019), se recogen datos del archivo de la grabación, de la grabación de la entrevista, de la transcripción y su revisión y de los hablantes que participan en la entrevista (entrevistador, informante y audiencia si la hubiera).

Al igual que en los códigos identificativos, en las cabeceras, la edad se codifica como 1 para los hablantes de edades comprendidas entre 20 y 34 años; 2 para aquellos cuyas edades oscilan entre los 35 y los 54 y 3 para los hablantes mayores de 55. Por último, el sexo se codifica como H (hombre) o M (mujer) y el nivel de instrucción, como 1 (instrucción básica), 2 (instrucción media) y 3 (instrucción superior).

3.2. Materiales del corpus

Como ya se ha mencionado más arriba, este volumen contiene las 54 entrevistas del corpus PRESEEA-Santander, conforme a las variables sociolingüísticas de sexo, edad y nivel de instrucción, es decir, hombres y mujeres de todas las edades y de todos los niveles educativos.

A continuación, en la Tabla 4 se detalla el perfil de los sujetos que componen el corpus completo. Comenzamos por el nivel de instrucción superior:

TABLA 5. HABLANTES DE INSTRUCCIÓN SUPERIOR: PERFIL DE LOS INFORMANTES

N.º de entrevista	N.º de informante y código	Sexo	Edad	Estudios
Entrevista 01	SANT_H13_001	Hombre	30	Licenciado en Ingeniería industrial
Entrevista 02	SANT_H13_002	Hombre	25	Licenciado en Matemáticas
Entrevista 03	SANT_H13_003	Hombre	23	Diplomado en Magisterio
Entrevista 04	SANT_M13_004	Mujer	24	Estudiante de 4º del Grado en Ingeniería civil
Entrevista 05	SANT_M13_005	Mujer	25	Estudiante del Grado en Diseño de moda
Entrevista 06	SANT_M13_006	Mujer	22	Estudiante de 3º del Grado en Psicología
Entrevista 07	SANT_H23_007	Hombre	38	Licenciado en Administración de Empresas
Entrevista 08	SANT_H23_008	Hombre	52	Licenciado en Filología Inglesa
Entrevista 09	SANT_H23_009	Hombre	47	Licenciado en Ciencias Físicas
Entrevista 10	SANT_M23_010	Mujer	53	Diplomada en Magisterio
Entrevista 11	SANT_M23_011	Mujer	44	Licenciada en Derecho
Entrevista 12	SANT_M23_012	Mujer	50	Licenciada en Geografía e Historia
Entrevista 13	SANT_H33_013	Hombre	59	Diplomado
Entrevista 14	SANT_H33_014	Hombre	58	Licenciado en Psicología
Entrevista 15	SANT_H33_015	Hombre	59	Licenciado en Medicina
Entrevista 16	SANT_M33_016	Mujer	55	Licenciada en Periodismo
Entrevista 17	SANT_M33_017	Mujer	74	Diplomada en Magisterio
Entrevista 18	SANT_M33_018	Mujer	77	Diplomada en Magisterio

En resumen, de los 18 sujetos entrevistados con nivel superior de instrucción, 10 son licenciados, 5 son diplomados y 3 son estudiantes que se encuentran en los últimos cursos de carrera (tercero y cuarto). La edad oscila entre los 22 y 30 años para la Generación 1, los 38 y 53 para la generación 2, y los 55 y 77 para la Generación 3.

A continuación, en la Tabla 5 detallamos el perfil de informantes que constituyen el nivel medio de instrucción del corpus PRESEEA-Santander:

TABLA 6. HABLANTES DE INSTRUCCIÓN MEDIA: PERFIL DE LOS INFORMANTES

N.º de entrevista	N.º de informante y código	Sexo	Edad	Estudios
Entrevista 19	SANT_H12_019	Hombre	32	Formación Profesional Grado Medio
Entrevista 20	SANT_H12_020	Hombre	31	Formación Profesional Grado Superior
Entrevista 21	SANT_H12_021	Hombre	34	Educación Secundaria Obligatoria
Entrevista 22	SANT_M12_022	Mujer	33	Formación Profesional Grado Medio
Entrevista 23	SANT_M12_023	Mujer	32	Formación Profesional Grado Medio
Entrevista 24	SANT_M12_024	Mujer	30	Formación Profesional Grado Medio
Entrevista 25	SANT_H22_025	Hombre	47	Formación Profesional Grado Medio
Entrevista 26	SANT_H22_026	Hombre	42	Formación Profesional Grado Superior
Entrevista 27	SANT_H22_027	Hombre	42	Bachillerato
Entrevista 28	SANT_M22_028	Mujer	39	Bachillerato (sin terminar)
Entrevista 29	SANT_M22_029	Mujer	46	Formación Profesional Grado Medio
Entrevista 30	SANT_M22_030	Mujer	47	Formación Profesional Grado Medio
Entrevista 31	SANT_H32_031	Hombre	63	Bachillerato
Entrevista 32	SANT_H32_032	Hombre	71	Bachillerato
Entrevista 33	SANT_H32_033	Hombre	64	Bachillerato
Entrevista 34	SANT_M32_034	Mujer	60	Perito mercantil
Entrevista 35	SANT_M32_035	Mujer	58	Bachillerato
Entrevista 36	SANT_M32_036	Mujer	65	Bachillerato

En resumen, de los 18 sujetos entrevistados con nivel medio de instrucción: 2 han cursado la Formación Profesional de grado superior; 7, la Formación Profesional de grado medio; 1 ha completado la Educación Secundaria Obligatoria; 6, el Bachillerato; 1 comenzó el Bachillerato, pero no lo terminó; finalmente, 1 cursó un peritaje mercantil. La edad de estos informantes de nivel medio de instrucción oscila entre los 30 y 34 años para la Generación 1, los 39 y 47 para la generación 2, y los 58 y 71 para la Generación 3.

Por último, insertamos la Tabla 6 con los datos de los informantes de PRESEEA-Santander que integran el nivel de instrucción baja:

TABLA 7. HABLANTES DE INSTRUCCIÓN BAJA: PERFIL DE LOS INFORMANTES

N.º de entrevista	N.º de informante y código	Sexo	Edad	Estudios
Entrevista 37	SANT_H11_037	Hombre	33	Educación Secundaria Obligatoria (sin terminar)
Entrevista 38	SANT_H11_038	Hombre	29	Formación profesional básica
Entrevista 39	SANT_H11_039	Hombre	33	Educación Secundaria Obligatoria (sin terminar)
Entrevista 40	SANT_M11_040	Mujer	31	Educación Secundaria Obligatoria (sin terminar)
Entrevista 41	SANT_M11_041	Mujer	29	Estudios primarios
Entrevista 42	SANT_M11_042	Mujer	31	Educación Secundaria Obligatoria (sin terminar)
Entrevista 43	SANT_H21_043	Hombre	42	Estudios primarios
Entrevista 44	SANT_H21_044	Hombre	51	Estudios primarios
Entrevista 45	SANT_H21_045	Hombre	43	Estudios primarios
Entrevista 46	SANT_M21_046	Mujer	35	Educación Secundaria Obligatoria (sin terminar)
Entrevista 47	SANT_M21_047	Mujer	49	Estudios primarios
Entrevista 48	SANT_M21_048	Mujer	45	Educación Secundaria Obligatoria (sin terminar)
Entrevista 49	SANT_H31_049	Hombre	77	Educación Secundaria Obligatoria (sin terminar)
Entrevista 50	SANT_H31_050	Hombre	57	Educación Secundaria Obligatoria (sin terminar)
Entrevista 51	SANT_H31_051	Hombre	57	Educación Secundaria Obligatoria (sin terminar)
Entrevista 52	SANT_M31_052	Mujer	67	Estudios primarios
Entrevista 53	SANT_M31_053	Mujer	74	Estudios primarios
Entrevista 54	SANT_M31_054	Mujer	57	Educación Secundaria Obligatoria (sin terminar)

En resumen, de los 18 sujetos entrevistados con el nivel bajo de instrucción, 9 no han concluido la Educación Secundaria Obligatoria (ESO)⁸, 8 poseen tan solo estudios primarios y 1 se encuentra cursando una formación profesional básica, dirigida a los que no lograron finalizar en su momento la ESO. La edad oscila entre los 29 y 33 años para la Generación 1, los 42 y 51 para la generación 2, y los 57 y 77 para la Generación 3.

⁸ Al tratarse de personas de 57 años, por razones obvias nos estamos refiriendo a la Educación General Básica de entonces, la EGB.

4. ESTUDIOS SOBRE LOS MATERIALES DE PRESEEA-SANTANDER

PRESEEA-Santander ha servido, a pesar de su juventud, como banco de datos clave para comenzar estudios de carácter empírico que permiten trazar una primera caracterización del habla santanderina. Por un lado, estos trabajos abordan algunas de las variables priorizadas del proyecto PRESEEA, pero también nos acercan, por otro lado, a áreas menos estudiadas como es el caso del léxico.

En el primero de los ámbitos, concretamente en el ámbito pragmático-discursivo, cabe reseñarse el estudio de Gancedo y Martínez (2023), que analiza la atenuación en el sociolecto alto de PRESEEA-Santander. Se descubre aquí que la función fundamental de la atenuación es la autoprotección del hablante. Por otro lado, los procedimientos lingüísticos atenuantes predilectos de los hablantes santanderinos son las fórmulas fáticas de contacto, la expresión de opinión en forma de duda o probabilidad, la modificación morfológica externa y las impersonalizaciones. En este mismo ámbito, en Gancedo y Soler (2023), desde una perspectiva contrastiva, se aborda la expresión de la epistemicidad y su impacto en la imagen social de los hablantes. El análisis sugiere que existen cuatro esquemas de comportamiento habituales en la entrevista de nivel sociocultural alto que permiten reconocer una gradualidad en la incidencia de las imágenes, derivada del foco y orientación que toman los recursos epistémicos. Asimismo, en el plano morfosintáctico, se ha llevado a cabo una aproximación a la caracterización sociolingüística del empleo del diminutivo (Gancedo Ruiz 2024).

En la segunda de las direcciones, Martínez y Ueda (2023a) analizan desde la lexicometría las tendencias léxicas de este corpus para avanzar en los estilos léxicos sociolingüísticos del español en sus diferentes variedades. Los resultados apuntan a que el sexo se revela como una variable importante en el proceso de variación léxica al detectarse, entre otros hallazgos, el uso del estilo nominal sobre el verbal y el empleo preferente de adverbios en *-mente* por parte del hombre; con respecto a la edad, se advierte la tendencia al empleo del truncamiento léxico en la generación de jóvenes y en el género mujer; por último, se aprecia la concentración de uso de *muchísimo* en mujer, joven, de nivel primario de instrucción.

Los mismos autores han abordado la caracterización de la flexión verbal, en términos de frecuencia y marcación, aplicable a otros corpus orales contemporáneos como este. Los resultados señalan el avance teórico en la estructura de la morfología verbal, donde parece existir una realidad bastante distinta a la trazada hasta el momento. Se propone, también en esta publicación, el uso del grado de marcación como herramienta útil para conocer la marca lingüística desde el punto de vista cuantitativo. Asimismo, han desarrollado estudios dirigidos a avanzar en los índices de riqueza léxica, donde los resultados obtenidos muestran una diferencia notable de Novedad léxica en función de la categoría gramatical y escasa en lo que atañe a los criterios sociolingüísticos. También han delimitado las relaciones entre longitud léxica y frecuencia de uso en el español contemporáneo (Martínez y Ueda, 2025) y

han concluido que la longitud de las formas debe observarse desde una perspectiva no física sino lingüística.

Cabe reseñarse, por último, la elaboración de dos bases de datos, el Inventario léxico y el Inventario morfológico, con el objetivo de poner a disposición de la comunidad científica herramientas que permitan localizar mejor la información para estudiar los datos desde parámetros tanto cualitativos como cuantitativos. El inventario léxico (Martínez y Ueda 2021) reúne y ordena el léxico extraído de las entrevistas orales, grabadas y transcritas en el núcleo urbano de la ciudad de Santander, para su uso investigador en estudios lingüísticos y aplicados. El inventario morfológico (Martínez y Ueda 2023b), por su parte, contiene, en esencia, las palabras base ordenadas alfabéticamente, con su categoría gramatical, la frecuencia total de aparición en el corpus oral, así como las frecuencias correspondientes a las tres variables sociolingüísticas establecidas (sexo, edad y nivel de instrucción).

En el momento de la publicación de este volumen, además de continuar con las líneas abiertas por los trabajos anteriormente señalados, se han puesto en marcha otras que abordan, dentro de las variables priorizadas, los fenómenos de laísmo y leísmo, los marcadores discursivos o el comportamiento de la -d- intervocálica en este español del norte peninsular. Estos trabajos se verán enriquecidos a partir de los estudios de carácter cuantitativo que podrán llevarse a cabo a partir del corpus de 54 entrevistas que aquí se presenta.

REFERENCIAS

- Ayuntamiento de Santander (2020). *Plan estratégico Santander 2020*. Consultado en: <https://planestrategicosantander2020.files.wordpress.com/2015/03/plan-estrategico-santander-2020.pdf>
- Ayuntamiento de Santander (2004). *Plan General de Ordenación Urbanística (PGOU). Emplazamiento y Estructura Urbana. Cómo entender Santander*. Consultado en: https://docplayer.es/61854254-I-emplazamiento-y-estructura-urbana-como-entender-santander.html#google_vignette
- Biblioteca Nacional de España (2024). *Digital Periodical and Newspaper Library: Boletín de comercio de Santander*. Consultado en: <https://hemerotecadigital.bne.es/hd/en/card?sid=26368695>
- Calderón Noguera, Donald Freddy y Julia Alvarado Castellanos (2011). “El papel de la entrevista en la investigación sociolingüística”. *Cuadernos de Lingüística Hispánica*, 17, pp. 11-24.
- Cestero Mancera, Ana M. (2012). “El proyecto para el estudio sociolingüístico del español de España y América”. *Español Actual*, 98, pp. 227-234.
- Cestero Mancera, Ana M., Isabel Molina Martos y Florentino Paredes García (2012). *La lengua hablada en Madrid. Corpus PRESEEA-Madrid (distrito de*

- Salamanca*), Volumen I. *Hablantes de Instrucción Superior*. Alcalá de Henares: Universidad de Alcalá, Servicio de Publicaciones.
- Cestero Mancera, Ana M., Isabel Molina Martos y Florentino Paredes García (2015a). “El estudio sociolingüístico de Madrid”, en Ana M. Cestero, I. Molina y F. Paredes (eds.): *Patrones sociolingüísticos de Madrid*. Bern: Peter Lang, pp. 17-62.
- Cestero Mancera, Ana M., Isabel Molina Martos y Florentino Paredes García (2015b). *Patrones sociolingüísticos de Madrid*, Bern: Peter Lang.
- Cestero Mancera, Ana M., Isabel Molina Martos y Florentino Paredes García (2019). *La lengua hablada en Madrid. Corpus PRESEEA-Madrid (distrito de Vallecas), Volumen I. Hablantes de Instrucción Superior*, Alcalá de Henares: Universidad de Alcalá, Servicio de Publicaciones.
- El Diario Montañés. Noticias de Cantabria y Santander. Consultado en: <https://cabeceras.eldiariomontanes.es/resultados-electorales-distritos-santander/mapa-santander.html>
- García Ferrando, Manuel (1994). *Socioestadística. Introducción a la estadística en sociología*. Madrid: Alianza Universidad.
- Gancedo Ruiz, Marta (2024). “El diminutivo en el español hablado en Santander: aproximación sociolingüística y pragmática”. [Comunicación en congreso]. *XX Congreso de la Asociación de Lingüística y Filología de la América Latina*. Concepción, Chile.
- Gancedo Ruiz, Marta e Inmaculada Martínez Martínez (2023). “Análisis sociolingüístico de la atenuación pragmática en el sociolecto alto de PRESEEA-Santander”, *Normas*, 13, pp. 296-315. DOI: <https://doi.org/10.7203/Normas.v13i1.27732>
- Gancedo Ruiz, Marta y María Amparo Soler Bonafont (2023). “Patrones relacionales de la expresión de la epistemicidad y la gestión de la imagen en el español peninsular”. *Spanish in Context*, 20(3), pp. 550-573. DOI: <https://doi.org/10.1075/sic.21016.gan>
- Hernández Campoy, Juan Manuel y Manuel Almeida (2005). *Metodología de la investigación sociolingüística*. Granada: Comares.
- Instituto Cántabro de Estadística (2024). *Explotación del ICANE del Censo de Población 2021*, INE. Comunicación personal por correo electrónico.
- Labov, William (1972). “Some principles of linguistic methodology”. *Language in Society*, 1, pp. 97-120.
- Labov, William (1984). “Field Methods of the Project on Linguistic Change and Variation”. En John Baugh y Joel Sherzer (eds), *Language in Use: Readings in Sociolinguistics*. Englewood Cliffs, NJ: Prentice Hall, pp. 28-66.
- Labov, William (1983). *The social stratification of English in New York City*. Washington, D. C: Center for Applied Linguistics.
- Labov, William (1996). *Modelos sociolingüísticos*. Madrid: Cátedra.
- López Morales, Humberto (1994). *Métodos de investigación lingüística*. Madrid: Gredos.

- Martín Butragueño, Pedro (1994). “Hacia una tipología de la variación gramatical”. *Nueva Revista de Filología Hispánica*, 41 (1), pp. 29-75.
- Martínez-Martínez, Inmaculada y Hiroto Ueda (2021). *Inventario léxico de PRESEEA-Santander*. DOI: <https://zenodo.org/records/10620777>
- Martínez-Martínez, Inmaculada y Hiroto Ueda (2023a). “Métodos de lexicometría sociolingüística: análisis del corpus oral contemporáneo PRESEEA-Santander”. *Círculo de Lingüística Aplicada a la Comunicación*, 94, pp. 227-245. DOI: 10.5209/clac.81206
- Martínez-Martínez, Inmaculada y Hiroto Ueda (2023b). *Inventario morfológico de PRESEEA-Santander*. DOI: <https://zenodo.org/records/10620852>
- Martínez-Martínez, Inmaculada y Hiroto Ueda (2024a). “Marcas de flexión verbal en español. Análisis de frecuencia en el corpus PRESEEA-Santander”, *Spanish in Context*. <https://doi.org/10.1075-sic.23004.mar>
- Martínez-Martínez, Inmaculada y Hiroto Ueda (2025a). La Novedad léxica como medida de riqueza léxica en un corpus oral contemporáneo: el corpus PRESEEA-Santander. *Revista Signos. Estudios De Lingüística*, 58(117). <https://doi.org/10.4151/S0718-09342025011701076>
- Martínez-Martínez, Inmaculada y Hiroto Ueda (2025b). Longitud léxica y frecuencia de uso en el español contemporáneo: un análisis estadístico de corpus. *ELUA: Estudios De Lingüística. Universidad De Alicante*, (43), 161–181. <https://doi.org/10.14198/ELUA.26993>
- Milroy, Lesli y Matthew Gordon (2003). *Sociolinguistics: Method and Interpretation*. Oxford: Blackwell.
- Moreno Fernández, Francisco (1996). “Metodología del ‘Proyecto para el estudio sociolingüístico del Español de España y de América’ (PRESEEA)”. *Lingüística*, 8, pp. 257-287.
- Moreno Fernández, Francisco (2021a). “Metodología del «Proyecto para el estudio sociolingüístico del español de España y de América» (PRESEEA)”. *Documentos PRESEEA de investigación*. Universidad de Alcalá. DOI: 10.37536/PRESEEA.2021.doc1
- Moreno Fernández, Francisco (ed.) (2021b). “Marcas y etiquetas mínimas obligatorias para materiales de PRESEEA”. *Documentos PRESEEA de investigación*. Universidad de Alcalá. DOI: <https://doi.org/10.37536/PRESEEA.2021.doc2>
- Moreno Fernández, Francisco (ed.) (2021c). “Guía PRESEEA para la investigación lingüística”, *Documentos PRESEEA de investigación*. Universidad de Alcalá. DOI: <https://doi.org/10.37536/PRESEEA.2021.doc3>
- Moreno Fernández, Francisco, Ana M. Cestero Mancera, Isabel Molina Martos y Florentino Paredes García (2002). *La lengua hablada en Alcalá de Henares. Corpus PRESEEA-ALCALÁ. I. Hablantes de instrucción alta*, Alcalá de Henares: Universidad de Alcalá.
- Moreno Fernández, Francisco, Ana M. Cestero Mancera, Isabel Molina Martos y Florentino Paredes García (2005). *La lengua hablada en Alcalá de Henares*.

- Corpus PRESEEA-ALCALÁ. II. Hablantes de instrucción media.* Alcalá de Henares: Universidad de Alcalá.
- Moreno Fernández, Francisco, Ana M. Cestero Mancera, Isabel Molina Martos y Florentino Paredes García (2007). *La lengua hablada en Alcalá de Henares. Corpus PRESEEA-ALCALÁ. III. Hablantes de instrucción primaria.* Alcalá de Henares: Universidad de Alcalá.
- Peña Arce, Jaime (2021). “La complejidad dialectal de Cantabria. Diacronía y sincronía del yeísmo regional”. *Zeitschrift für romanische Philologie* 137 (2), pp. 426-450. <https://doi.org/10.1515/zrp-2021-0017>
- Recalde Fernández, Montserrat y Victoria Vázquez Rozas (2009). “Problemas metodológicos en la formación de corpus orales”. En P. Cantos Gómez y A. Sánchez Pérez (eds). *A Survey of Corpus-based Research. Panorama de investigaciones basadas en corpus.* Murcia: AELINCO, pp. 37-49. <https://www.um.es/lacell/aelinco/contenido/pdf/4.pdf>.
- Silva-Corvalán, Carmen (1989). *Sociolingüística: Teoría y análisis.* Madrid: Alhambra.
- Silva-Corvalán, Carmen (1994). “Direcciones en los estudios sociolingüísticos de la lengua española”, *Actas del Congreso de la Lengua Española.* Sevilla: Instituto Cervantes, pp. 399-415.
- Vida Castro, Matilde (ed.) (2007). *El español hablado en Málaga.* Málaga: Editorial Sarriá.