La lengua hablada en Madrid
Corpus PRESEEA-Madrid XXI (distrito de Salamanca)

UAH MONOGRAFÍAS
HUMANIDADES 101

La lengua hablada en Madrid
Corpus PRESEEA-Madrid XXI (distrito de Salamanca)

Nicolás Moreno Martín de Nicolás
Sara Engra Minaya
Gema Cuesta Chorro

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 MICIU/AEI/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): 978-84-10432-50-5

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

Hemos de expresar nuestro agradecimiento a los propietarios y autores del corpus PRESEEA-Madrid, Ana María Cestero Mancera, Isabel Molina Martos, Francisco Moreno Fernández y Florentino Paredes García, por la oportunidad de publicar este sub-corpus. Muy especialmente, queremos dar las gracias a Ana María Cestero Mancera e Isabel Molina Martos, que, en todo momento, nos han proporcionado ayuda, materiales y opiniones para la redacción y revisión de este texto.

Presentación

1. El proyecto PRESEEA[1]

En el X Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina (ALFAL), de abril de 1993, la Comisión de Sociolingüística decidió iniciar un proyecto para el estudio sociolingüístico de las ciudades iberoamericanas y españolas, cuyo objetivo sería la puesta en marcha de un servicio de documentación sociolingüística y, especialmente, la creación de corpus sociolingüísticos sincrónicos del español y del portugués[2]. De esta iniciativa nace el "Proyecto para el Estudio Sociolingüístico del Español de España y de América" (PRESEEA), abierto a la participación de todos los investigadores y centros de investigación que lo deseen[3].

Desde el punto de vista teórico, el PRESEEA se basa en una serie de principios sociolingüísticos y sociológicos, relacionados con aspectos tales como 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[4].

---

[1] El corpus PRESEEA-Madrid XXI se inscribe en el marco de las actividades científicas 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, Innovación y Universidades-Agencia Estatal de Investigación/10.13039/501100011033. Nicolás Moreno Martín de Nicolás y Sara Engra Minaya son, en el momento de esta publicación, investigadores en formación en la universidad de Alcalá, en el departamento de Filología, Comunicación y Documentación. También son estudiantes en el programa de doctorado en Estudios Lingüísticos, Literarios y Teatrales de este departamento y disponen respectivamente de una ayuda del Ministerio de Ciencia e Innovación de referencia PRE2020-095311 y un contrato FPU del Ministerio de Universidades (referencia FPU19/05346).

[2] Esta introducción se desarrolla más detalladamente en el artículo de F. Moreno Fernández de 1996 y en el de A. M. Cestero de 2012a. La presentación general del proyecto de Madrid, recogida en gran parte en estas páginas introductorias, se anticipó en Moreno, Cestero, Molina y Paredes 2000. Véase, también, Moreno, Cestero, Molina y Paredes 2001, 2002, 2005a, 2005b y 2007. Es necesario señalar que en estas páginas seguimos el modelo de introducción que plantearon Cestero, Molina y Paredes (2020), Molina, Paredes y Cestero (2020a) y Paredes, Cestero y Molina (2020) para los corpus de PRESEEA-Madrid (distritos de Vallecas).

[3] En la gestación de PRESEEA fue decisivo el apoyo de Carmen Silva-Corvalán, que en repetidas ocasiones había señalado la necesidad de un proyecto de estas características (Silva-Corvalán 1994).

[4] Algunos de estos principios pueden formularse de la siguiente manera (Moreno Fernández 2009): Principio del dialecto: "el dialecto se manifiesta en una comunidad"; Principio de la comunidad de habla: "una comunidad

En cuanto a la metodología, aunque el proyecto pretende ser muy abarcador, se aconseja seguir una serie de requisitos mínimos (Moreno 2021a y b):

- Los hablantes seleccionados pueden ser monolingües o bilingües; en este segundo caso, deben tener un dominio del español similar al de un monolingüe[5].
- Los núcleos poblacionales de los que se reúnen materiales deben ser núcleos de población asentada, de forma que se pueda hablar de una conciencia de comunidad de habla. En este sentido, los estudios se ajustarán a núcleos urbanos concretos, aunque la comunidad de habla a la que se refieran pueda superar los límites de dicho núcleo[6].
- El muestreo debe ser representativo del universo que sirve de base al estudio sociolingüístico. A este efecto, se plantea trabajar con muestras por cuotas con afijación uniforme. La pre-estratificación de la población se hará teniendo en cuenta estas tres variables: sexo, edad y nivel de instrucción. Los investigadores de cada equipo pueden añadir otras variables en la post-estratificación de la población en función de las características del núcleo estudiado[7].
- El tamaño de la muestra debe ajustarse a los niveles de representatividad que se consideran adecuados en los trabajos de sociolingüística[8]. El tamaño-tipo con el que se está trabajando es de 54, 72 o 108 informantes.
- La recogida de datos se hará por medio de entrevistas de al menos 45 minutos de duración. Las conversaciones serán semidirigidas y grabadas mediante una cámara de vídeo digital que se deja a la vista, cuando la entrevista se haga en persona, y con la herramienta Microsoft *Teams* cuando se realice de manera virtual.

---

de habla está formada por un conjunto de hablantes que comparten al menos una variedad lingüística, unas reglas de uso, una interpretación de ese uso, unas actitudes y una misma valoración de las formas lingüísticas"; Principio de la variación lingüística: "los usos lingüísticos variables –fonéticos, gramaticales o léxico–semánticos– pueden covariar con otros elementos lingüísticos o extralingüísticos"; Principio de la cuantificación: "para establecer relaciones entre variables lingüísticas y variables sociales se requieren métodos cuantitativos"; Principio de la representatividad: "el habla de una muestra debe ser representativa de un grupo social o de una comunidad".

[5] El bilingüismo que se tiene en cuenta es el derivado del uso habitual de dos lenguas dentro de una comunidad. Esto quiere decir que se podría trabajar con individuos bilingües si el proyecto se realizara, por ejemplo, en ciudades como Barcelona (España) o Asunción (Paraguay). De hecho, en España se ha trabajado sistemáticamente con individuos bilingües en Valencia, Barcelona, Lérida, Palma de Mallorca y Santiago de Compostela.

[6] Acerca de las muchas dificultades, teóricas y metodológicas, que presenta el concepto de "comunidad de habla" pueden consultarse los trabajos de H. López Morales (1994: 47-52) y de S. Romaine 1980.

[7] Existe una propuesta de experimentar el rendimiento sociolingüístico, desde la metodología de PRESEEA, del concepto (variable) "modo de vida", propuesto por T. Højrup. Los modos de vida responden a un modelo en el que los grupos sociales se consideran entidades internamente estructuradas y relacionadas con otros grupos. La conducta lingüística responde más a la capacidad de determinación de las redes y de las estructuras en que se mueven los hablantes que a los atributos sociales que caracterizan a determinados grupos (Milroy 1992: 206-220). Sobre el manejo del factor "modo de vida", véase el trabajo de Lastra y Martín Butragueño 2000.

[8] Como es bien sabido en estadística, el tamaño dependerá más de la heterogeneidad del universo que de su dimensión cuantitativa.

- La transcripción y codificación de los materiales se realiza mediante una adaptación del *Standard Generalized Markup Language* (SGML) (Van Herwijnen)[9].

El proyecto aglutina a más de cuarenta equipos de investigación que han recopilado materiales de lengua hablada en España y en América en un corte sincrónico que abarca desde 1995 hasta 2010[10], en una primera fase, y a partir de 2011, en la segunda y actual (Moreno y Cestero 2020). Durante estos periodos, se han trabajado los corpus correspondientes a las siguientes comunidades urbanas hispanohablantes: Alcalá de Henares (España), Arequipa (Perú), Asunción (Paraguay), Barcelona (España), Barranquilla (Colombia), Bogotá (Colombia), Buenos Aires (Argentina), Cádiz (España), Cali (Colombia), Caracas (Venezuela), Cartagena de Indias (Colombia), Chicago (EE. UU.), Cipolletti (Argentina), Cuenca (Ecuador), Gijón (España), Granada (España), Guadalajara (México), Guatemala (Guatemala), La Habana (Cuba), La Paz (Bolivia), Las Palmas de Gran Canaria (España), Lérida (España), Lima (Perú), Madrid (España), Málaga (España), Medellín (Colombia), Mérida (México), Mérida (Venezuela), Mexicali (México), Ciudad de México (México), Monterrey (México), Montevideo (Uruguay), Nueva York (EE.UU.), Oviedo (España), Palma de Mallorca (España), Pamplona (España), Pereira (Colombia), Puebla (México), Quito (Ecuador), San Juan de Puerto Rico (Puerto Rico), Santander (España), Santiago de Chile (Chile), Santiago de Compostela (España), Sevilla (España), Tijuana (México), Tunja (Colombia), Valencia (España), Valledupar (Colombia), Valparaíso (Chile) y Zaragoza (España)[11].

1.1. El PRESEEA en Madrid: PRESEEA-Madrid XXI

El corpus PRESEEA de la Comunidad de Madrid se planificó para que incluyera materiales de habla tanto de Alcalá de Henares como de la ciudad de Madrid. Tras investigar el habla de la ciudad complutense (Moreno, Cestero, Molina y Paredes 2002, 2005b y 2007), en 2001 el equipo de la Universidad de Alcalá[12] comenzó el

---

[9] Véanse los trabajos de la TEI (*Text Encoding Initiative*) (Ide y Sperber-McQueen 1995). Ello no impide que, con distintos fines, los materiales se puedan presentar con otras formas o codificaciones. Para más información, véase el documento PRESEEA de trabajo *Marcas y etiquetas mínimas obligatorias para materiales de PRESEEA* (Moreno Fernández 2021c).

[10] Para ampliar la información relativa a estos proyectos, así como para consultar muestras de los materiales y otros trabajos de investigación puede verse https://preseea.uah.es/.

[11] PRESEEA debe mucho al poder de convocatoria y al trabajo de organización e investigación de Humberto López Morales, embajador entusiasta del proyecto, así como a la difusión que se ha realizado, impulsada por Francisco Moreno Fernández, coordinador general, desde los Congresos Internacionales de la ALFAL y desde los Encuentros Internacionales del Proyecto PRESEEA.

[12] Los corpus de Alcalá de Henares y de los distritos de Salamanca, Puente de Vallecas y Villa de Vallecas de Madrid que, junto con el corpus PRESEEA-Madrid XXI, conforman el corpus general PRESEEA-Comunidad de Madrid fueron diseñados, organizados y recogidos por Ana M.ª Cestero, Isabel Molina, Francisco

estudio sociolingüístico de la capital. En los años posteriores, este trabajo dio como resultado la creación del corpus PRESEEA-Madrid, conformado por los sub-corpus del distrito de Salamanca (Cestero, Molina y Paredes 2012; Molina, Paredes y Cestero 2014; Paredes, Cestero y Molina 2015) y los distritos de Vallecas (Cestero, Molina y Paredes 2020; Molina, Paredes y Cestero 2020a; Paredes, Cestero y Molina 2020). Justificar el estudio sociolingüístico de Madrid requiere de pocas argumentaciones: la ciudad es uno de los focos de irradiación lingüística más importantes del español y la consideración de su habla como modelo idiomático hace que el análisis sociolingüístico deba tenerla presente.

En 1561, la corte del antiguo Reino de Castilla se trasladó a la villa de Madrid, lo que supuso una considerable expansión demográfica de la nueva capital: los 4060 habitantes con los que contaba Madrid en 1530 llegaron a ser 175 000 en el siglo siguiente (Moreno 2005b: 141). La llegada de los Borbones a principios del siglo XVIII dio lugar a la centralización, en favor de la villa, de las leyes, usos y costumbres de los nuevos territorios bajo gobierno borbónico. Con la entrada del siglo XX, la creciente relevancia política y económica de Madrid culminó en un notable aumento demográfico, como consecuencia de un modelo de distribución caracterizado por movimientos migratorios desde el interior de la Meseta central hacia la capital y las costas. Es durante este siglo cuando se produce el impulso poblacional definitivo para España, al pasar de 18,5 millones de hablantes en 1900 a unos 40 a finales del siglo XX. Madrid fue beneficiaria de este crecimiento, al atraer a grandes masas campesinas de las dos Castillas y así llegó a convertirse en la única ciudad castellana que, por factores políticos y demográficos, era capaz de actuar como centro socioeconómico de referencia. Las consecuencias lingüísticas de esta expansión poblacional consistieron en un impulso a los procesos de convergencia de la periferia con el centro, es decir, a la uniformización lingüística a semejanza de la norma castellana madrileña.

Las décadas de 1990 y posteriores trajeron consigo una nueva redefinición de la personalidad sociolingüística de Madrid, al incorporarse a la población española enormes contingentes de migrantes de procedencia diversa, debido al desarrollo económico de España a finales del siglo XX y comienzos del XXI. La demanda de trabajadores generada por el crecimiento económico nacional afectó significativamente a las cifras de población migrante en España: de un 1,3% de población inmigrante en el total de la población española en 1996, se pasó a un 10% en 2007. Estos más de 4,5 millones de migrantes provenían principalmente de Marruecos, Rumanía y Latinoamérica. Cabe la posibilidad de que el impacto económico y social que estos

---

Moreno y Florentino Paredes (Moreno, Cestero, Molina y Paredes 2002, 2005b y 2007; Cestero, Molina y Paredes 2012; Molina, Paredes y Cestero 2014; Paredes, Cestero y Molina 2015; Cestero, Molina y Paredes 2020; Molina, Paredes y Cestero 2020a; Paredes, Cestero y Molina 2020), con la colaboración, en ocasiones, de personal en formación de la Universidad de Alcalá.

grupos tienen sobre España lleve a una cierta modificación de la lengua común, uniéndose a la difusión centrífuga de rasgos propia de los centros de pestigio (Penny 2000: 57). En la tabla 1 se reflejan las cifras de estos procesos migratorios, especialmente interesantes entre 2000 y 2010, cuando se recogió el PRESEEA-Madrid.

Tabla 1. Madrid: población

	1998	2000	2005	2010	2015	2018	2022
Comunidad de Madrid	5 091 336	5 205 408	5 964 143	6 458 684	6 436 996	6 578 079	6 750 336
Madrid	2 881 506	2 883 860	3 155 359	3 273 049	3 141 991	3 223 334	3 280 782
Población española en Madrid	2 789 880	2 703 507	2 709 198	2 704 835	2 747 504	2 799 459	2 768 564
Población extranjera en Madrid	91 626	80 718	446 161	568 214	380 707	423 875	512 218

Fuente: Instituto Nacional de Estadística (INE)
[http://www.madrid.org/iestadis/fijas/otros/estructu_demodosnivel.htm]
Fecha de consulta: marzo de 2024

Los datos poblacionales de la tabla también evidencian cambios demográficos posteriores a la década de 2010. La crisis económica iniciada en 2008, que perjudicó gravemente a España en los siguientes años e impulsó cambios políticos y sociales, también supuso un cierto declive en la población migrante en el país: los datos de 2015 indican que casi 200 000 migrantes abandonaron España, si bien se aprecia una recuperación poblacional de este colectivo en los cortes demográficos de 2018 y 2022.

Las transformaciones sociales de España en este tiempo justifican la recogida de un nuevo sub-corpus del PRESEEA-Madrid: el corpus PRESEEA-Madrid XXI[13], que nace de la necesidad de profundizar en la teorización sobre patrones sociolingüísticos del español. El corpus general PRESEEA-Madrid ya había dado lugar a numerosas investigaciones sobre el habla madrileña (Cestero, Molina y Paredes 2015a) y había permitido alcanzar un entendimiento profundo de la situación sociolingüística de la región, además de haberse utilizado para investigar tendencias sociolingüísticas y procesos de convergencia y divergencia en el español (Molina, Paredes y Cestero 2020b). Para avanzar en la comprensión de la lengua de Madrid y del español en general, en una nueva fase del proyecto comenzaron a reunirse muestras de

[13] El corpus PRESEEA-Madrid XXI (distrito de Salamanca) que aquí se ofrece es propiedad de Ana M.ª Cestero Mancera, Isabel Molina Martos, Francisco Moreno Fernández y Florentino Paredes García. Las encuestas que lo conforman han sido realizadas por los cuatro investigadores, con la colaboración de Nicolás Moreno Martín de Nicolás, Gema Cuesta Chorro y Sara Engra Minaya.

habla de momentos cronológicos anteriores y posteriores a la recogida de entrevistas del PRESEEA-Madrid, que tuvo lugar a lo largo de la primera década del siglo XXI. Las muestras de estos sub-corpus fueron tomadas aproximadamente 20 años antes y 20 años después que las del corpus PRESEEA-Madrid, por lo que tienen el potencial de proporcionar información muy valiosa en relación con la variación diacrónica del habla de la capital. El sub-corpus que aquí nos ocupa, el PRESEEA-Madrid XXI, está integrado por muestras lingüísticas audiovisuales tomadas de hablantes procedentes del distrito de Salamanca entre 2022 y 2023.

La importancia del sub-corpus PRESEEA-Madrid XXI radica en su valor comparativo con el corpus general PRESEEA-Madrid, puesto que la distancia temporal entre ambos permite realizar estudios sociolingüísticos en tiempo real: de este modo, es posible comprobar la evolución de los cambios lingüísticos ya identificados anteriormente, tales como las realizaciones de -/s/ en coda silábica (Molina 2015), la pluralización del verbo *haber* (Paredes 2016) o los procesos de atenuación (Cestero 2012b, 2015b, 2017), entre otros. Estos nuevos datos serán vitales en la construcción de modelos predictivos del cambio lingüístico generados a partir de datos cuantitativos (Nyce 2007), como los que se usan en sociología, economía o medicina. El PRESEEA-Madrid XXI también puede utilizarse para realizar estudios de panel, ya que se ha procurado mantener, cuando ha sido posible, a los mismos informantes que ya participaron en el proyecto veinte años atrás. Contamos, en esta ocasión, con seis informantes que formaron parte del PRESEEA-Madrid recogido a principios del siglo XXI. De esta manera, estas entrevistas podrán dar lugar a estudios longitudinales en los que se investigue la variación lingüística a lo largo de la vida de un mismo hablante[14].

1.2. Madrid, distrito de Salamanca: la entidad social

Con el objetivo de mantener un nivel de representatividad similar al obtenido en Alcalá de Henares, en la ciudad de Madrid fueron seleccionadas dos áreas: los distritos de Salamanca y de Vallecas (Cestero, Molina y Paredes 2015b). En la tabla siguiente se muestran datos de la población de los sub-corpus que conforman PRESEEA-Comunidad de Madrid, en una comparativa histórica:

[14] Véase la página 19, en la que se proporcionan más detalles sobre los entrevistados.

Tabla 2. Comunidad de Madrid: población de zonas de PRESEEA

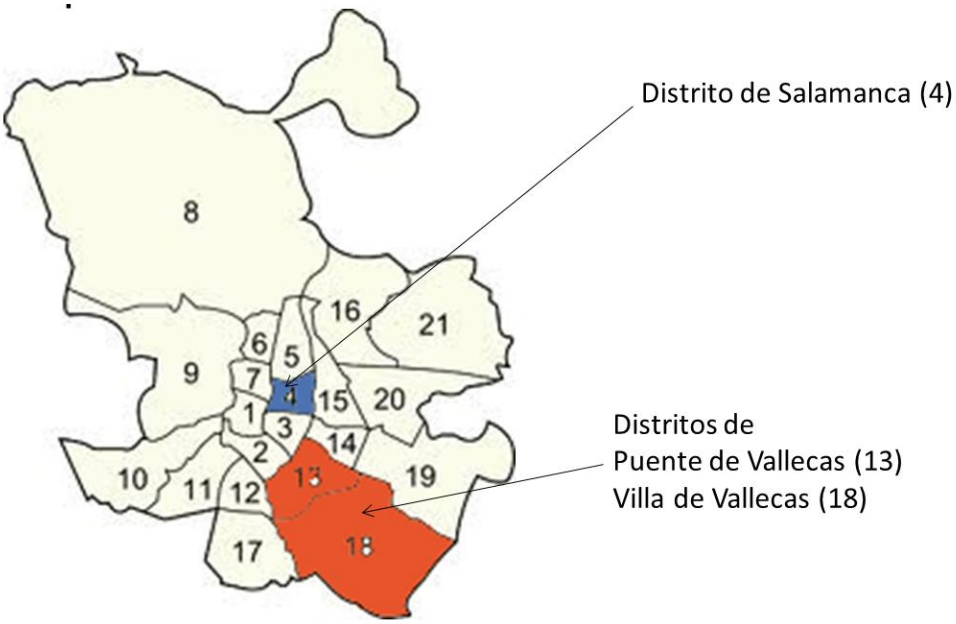
	1991	1996	2000	2005	2010	2015	2018	2022
Comunidad de Madrid	4 947 555	5 022 289	5 205 408	5 964 143	6 458 684	6 436 996	6 578 079	6 750 336
Madrid	3 010 492	2 866 850	2 883 860	3 15 359	3 273 049	3 141 991	3 223 334	3 280 782
Distrito de Salamanca	156 082	142 001	143 090	150 775	147 380	142 388	145 815	145 579
Distritos de Vallecas	284 186	285 265	286 462	307 503	333 2239	325 771	339 651	353 046
Alcalá de Henares	159 355	163 386	166 397	201 380	204 120	198 750	193 751	169 888

Fuentes: Ayuntamiento de Madrid. Subdirección General de Estadística.
[https://servpub.madrid.es/CSEBD_WBINTER/arbol.html]
e Instituto Nacional de Estadística (INE)
[http://www.madrid.org/iestadis/fijas/otros/estructu_demodosnivel.htm]
Fecha de consulta: marzo de 2024

El corpus PRESEEA-Madrid (distrito de Salamanca) se publicó en tres volúmenes entre 2012 y 2015 (Cestero, Molina y Paredes 2012; Molina, Paredes y Cestero 2014; Paredes, Cestero y Molina 2015); mientras que el sub-corpus de los distritos de Vallecas fue publicado en 2020, también en tres tomos (Cestero, Molina y Paredes 2020; Molina, Paredes y Cestero 2020a; Paredes, Cestero y Molina 2020). El PRESEEA-Madrid XXI se ha recogido con informantes procedentes del distrito de Salamanca, por lo que a continuación describiremos esta zona de la ciudad con más detalle y la compararemos brevemente con los distritos de Vallecas.

El distrito de Salamanca se sitúa en el centro de la ciudad, entre las grandes arterias de tráfico: el Paseo de la Castellana por el oeste, la calle de María de Molina y la Avenida de América por el norte, la M-30 por el este y la calle O'Donnell por el sur. Ocupa una superficie de 5,4 km^2 de los 604 km^2 que constituyen la ciudad de Madrid. Es un distrito de trazado muy regular, cuya población se puede caracterizar como de clase media y media-alta, largamente asentada en la zona. En la siguiente imagen se presenta la ubicación de los distritos de Salamanca y Vallecas:

Figura 1. Ubicación de los distritos de Salamanca y Vallecas dentro de Madrid



Fuente: Cestero, Molina y Paredes (2020).

Aunque se conoce popularmente como barrio de Salamanca, en realidad es un distrito integrado por las áreas de Recoletos, Goya, Fuente del Berro, Guindalera, Lista y Castellana. Se construyó en la segunda mitad del siglo XIX, cuando el marqués de Salamanca, que dio su nombre al distrito, se encargó de financiar y llevar a cabo uno de los proyectos programados por el ayuntamiento para realizar el ensanche de la capital. Las pretensiones del proyecto inicial se mantuvieron, pues continuó siendo un distrito activo, con comercios de lujo y habitado por clases sociales de elevado estatus sociocultural y económico. El carácter burgués del distrito se reflejaba en las características de sus viviendas: ya en el siglo XIX contaba con los adelantos tecnológicos más modernos en materia de vivienda, como el alumbrado eléctrico, el teléfono, los ascensores o los termos de agua caliente (Gil Peña 2005: 31).

Considerado paradigmático de la clase alta, desde mediados del siglo XX el distrito de Salamanca se ha convertido en un arquetipo lingüístico con capacidad de influir en el habla de la ciudad (Molina 2022a). Desde sus orígenes, el distrito se ha identificado con la clase alta –cuyos miembros, especialmente los jóvenes, son también llamados "pijos"– conocida e imitada por el resto de los madrileños en los hábitos de vida, en la forma de vestir e incluso en la manera de hablar, que muchas veces se considera afectada.

En cambio, los distritos de Vallecas tienen un perfil demográfico más humilde, compuesto por clases medias y bajas e influido por la inmigración nacional e inter-

nacional de los siglos XX y XXI. Los distritos de Vallecas son, en realidad, dos distritos distintos, contiguos, periféricos y sociodemográficamente similares: Puente de Vallecas y Villa de Vallecas. El primero, surgido como extensión del anillo externo de la capital y poblado en las olas de inmigración rural del siglo XX, se caracteriza por su notable densidad poblacional y su actividad comercial. Por su parte, el segundo nace de la unión entre el barrio de Santa Eugenia y el casco histórico de la Villa de Vallecas, que anteriormente era un municipio independiente de Madrid: estos orígenes heterogéneos determinan también la diversidad sociológica de su población, puesto que Villa de Vallecas acoge tanto a la población arraigada en la zona como a migrantes de dentro y fuera de España. A su vez, el Ensanche de Vallecas define al distrito como una zona en expansión.

La distinta composición sociológica de los distritos de Salamanca y Vallecas está parcialmente reflejada en la tabla 3, que incluye las cifras de población española y extranjera en cada distrito.

Tabla 3. Madrid y los distritos de Salamanca y Vallecas: población española y extranjera

	2003	2005	2010	2015	2018	2022
Madrid	3.098.065	3 15 359	3 273 049	3 141 991	3 223 334	3 280 782
Distrito de Salamanca	152.403	153.196	148.130	143.152	145.815	145 579
Población española	132.733	131.676	127.491	127.100	127621	124 016
Población extranjera	19.670	21.520	20.639	16.052	18.194	21 563
Puente de Vallecas	242.070	244.385	244.886	227.263	231.524	238 577
Población española	210.589	206.417	196.099	194.570	193.661	189 978
Población extranjera	31.481	37.968	48.787	32.693	37.863	48 599
Villa de Vallecas	65.221	66.383	88.748	100.039	108.127	114 469
Población española	58.469	57.941	73.930	32.693	95.842	98 592
Población extranjera	6.752	8.442	14.818	10.920	12.285	15 907

Fuente: Ayuntamiento de Madrid. Subdirección General de Estadística.
[https://servpub.madrid.es/CSEBD_WBINTER/arbol.html]
Fecha de consulta: marzo de 2024

En las últimas décadas, se han producido nuevos cambios demográficos en la sociedad madrileña: a la caída y recuperación de la población migrante en la década de 2010 se une el envejecimiento poblacional, tendencia general en España. La tabla 4, elaborada a partir del padrón de 2020 (cuando se planteó el proyecto), refleja la distribución de población en Madrid y el distrito de Salamanca por edad y sexo.

Tabla 4. Distrito de Salamanca: población clasificada por sexo y grupos etarios (2020)

	Total	Entre 0-19 años	Entre 20-34 años	Entre 35-54 años	Más de 55 años
Madrid	3 334 730	584 622	622 786	1 033 701	1 093 621
Hombres	1 553 899	298 425	301 039	499 321	455 114
Mujeres	1 780 831	286 197	321 747	561 082	638 507
Distrito de Salamanca	147 841	20 936	31 206	40 889	54 810
Hombres	65 158	10 484	14 146	18 967	21 561
Mujeres	82 683	10 452	17 060	21 922	33 249

Fuentes: Ayuntamiento de Madrid. Subdirección General de Estadística.
[https://servpub.madrid.es/CSEBD_WBINTER/arbol.html]
Fecha de consulta: marzo de 2024

Del gran dinamismo social que caracteriza la ciudad de Madrid cabe esperar un reflejo en lo lingüístico: es modelo de la norma del centro peninsular y, en consecuencia, se ha convertido en foco desde el que irradian innovaciones lingüísticas que los medios de comunicación contribuyen a difundir. La convergencia con la norma madrileña por parte de la población del área metropolitana y las actitudes compartidas ante los mismos hechos de lengua han llevado a considerar el "Gran Madrid" como una comunidad de habla muy extendida[15].

Por otro lado, si bien en sus niveles socioculturales altos los madrileños mantienen el carácter conservador del castellano, la presencia de inmigrantes meridionales debe de haber dejado su impronta, según puede deducirse del avance geográfico hacia el norte de desarrollos lingüísticos considerados meridionales[16]. La hipotética influencia de la inmigración española sobre el habla madrileña podrá estudiarse poniendo en contraste los materiales de Vallecas con los del distrito de Salamanca. Es posible que el contacto entre los diferentes niveles sociales que conviven en la capital permita el trasvase de influencias lingüísticas, no solo por imitación de la norma, sino también aceptando rasgos dialectales. La elaboración y publicación de los corpus sociolingüísticos de Madrid permitirá entrar a dilucidar este tipo de cuestiones.

[15] Martín Butragueño (2004: 100).
[16] Cf. García Mouton y Moreno Fernández 1994.

2. Metodología

2.1. Criterios de selección de informantes

Siguiendo los criterios metodológicos del PRESEEA expuestos en (§1), Cestero Mancera y Moreno Martín de Nicolás diseñaron y organizaron la recogida de la muestra para el estudio sociolingüístico de Madrid (distrito de Salamanca) en la década de los años veinte del siglo XXI. Se ha partido, como en el caso de los anteriores corpus PRESEEA de la capital, de una preestratificación de la población de acuerdo con los tres factores anteriormente señalados: "edad", "sexo" y "nivel de instrucción". Por "sexo", la población queda agrupada en Mujeres y Hombres, mientras que en la estratificación por "nivel de instrucción", los informantes se incluyen en uno de los grupos que siguen: Grado de Instrucción 1, Enseñanza Primaria; Grado de Instrucción 2, Enseñanza Secundaria; Grado de Instrucción 3, Enseñanza Superior[17]. En relación con la estratificación por "edad", los hablantes se distribuyen en una de estas tres variantes: en la Generación 1, los de 20 a 34 años; en la Generación 2, los de 35 a 54 años; y en la Generación 3, los de más de 55 años.

En cuanto al problema metodológico, ya clásico en sociolingüística, de la representatividad de la muestra que ya se había planteado en los corpus PRESEEA-Madrid (distritos de Salamanca y de Vallecas), no ha sido necesario lidiar con ello antes de la recogida de datos. Esto se debe a que PRESEEA-Madrid XXI es un sub-corpus, una muestra, de PRESEEA-Madrid y no un corpus independiente.

Otro de los criterios para la selección de informantes tenido en cuenta para PRESEEA-Madrid XXI ha sido su colaboración en los corpus anteriores de PRESEEA-Madrid (distrito de Salamanca). Para la primera selección de hablantes, Ana M. Cestero partió de aquellos informantes que participaron en el corpus recogido hace veinte años con el fin de conseguir una misma muestra en dos intervalos de tiempo bien diferenciados. No se consiguió ningún informante de la tercera generación (más de 55 años) del corpus previo que se encontrara en condiciones de participar; no obstante, se pudo recoger, veinte años después, las muestras de seis informantes que habían participado en PRESEEA-Madrid: cuatro de la primera generación, que ahora son de la generación 2 (35-54 años), y dos de la segunda generación, que corresponden a la tercera en este sub-corpus. En total, la muestra de Madrid (distrito de Salamanca) está formada por 18 informantes, distribuidos por cuotas fijas según las variables ya mencionadas, como refleja la siguiente tabla. En ella se indican, con un asterisco, aquellos informantes que participaron también en PRESEEA-Madrid (distrito de Salamanca).

---

[17] En los corpus PRESEEA-Comunidad de Madrid se considera que un hablante pertenece al Grado 2 o al Grado 3 cuando ha cursado más de un año completo en cada uno de esos niveles.

Tabla 5. Distribución por cuotas de los informantes

	Generación 1		Generación 2		Generación 3	
	H	M	H	M	H	M
Grado de instrucción 1	1	1	1	1	1	1
Grado de instrucción 2	1	1	1*	1*	1*	1*
Grado de instrucción 3	1	1	1*	1*	1	1

2.2. Metodología para la recogida de materiales

2.2.1. *Diseño de la encuesta*

El diseño adecuado de una encuesta encierra varias cuestiones a las que se debe prestar especial atención, entre las que destaca una especialmente, relacionada con su objeto de estudio: una encuesta debe ser conveniente a lo que se ha delimitado como objetivo de la investigación (Moreno Fernández 2021a). En el proyecto sociolingüístico de Madrid, como habían hecho previamente en el de Alcalá de Henares[18], Cestero, Moreno, Molina y Paredes se plantearon, de forma general, la recogida de materiales que proporcionaran suficiente información sobre el habla de los estratos sociales de la ciudad. La representatividad de los distintos estratos queda aparentemente confirmada por la configuración por cuotas de la muestra, aunque resulta conveniente determinar con claridad el tipo de encuesta que aportará suficiente información sobre el habla de los informantes.

Desde sus inicios, la sociolingüística variacionista ha realizado análisis sobre distintos niveles de la lengua, especialmente en el fonético-fonológico, el léxico-semántico y el morfosintáctico y, en cada uno de ellos, se han encontrado con un obstáculo al definir la "variable sociolingüística". No nos ocuparemos aquí de tratar esa polémica sobre la dificultad mencionada, pues ya es bien conocida (vid. Moreno Fernández 2009); por el contrario, sí se mantiene la división por niveles del lenguaje para justificar la validez de la metodología.

La variación que menos problemas metodológicos plantea en el ámbito de la sociolingüística es la fonético-fonológica, por lo que es una consecuencia lógica que haya sido la más estudiada e investigada (López Morales 2004). La obtención de datos fonético-fonológicos resulta bastante sencilla en cualquier tipo de encuesta por tratarse de un número limitado y escaso de unidades que pertenecen a inventarios cerrados y que, por lo tanto, presentan unos índices de frecuencia altos en el habla espontánea. Esto asegura, en cualquier encuesta que se pueda reproducir para su transcripción o revisión, la presencia

[18] Consúltese Moreno, Cestero, Molina y Paredes 2002, 2005b y 2007, así como Cestero, Molina y Paredes 2012, 2020; Molina, Paredes y Cestero 2014, 2020a; Paredes, Cestero y Molina 2015, 2020.

de suficientes datos para el análisis, al menos en un registro de habla.

Esto no sucede así en otros niveles, como el gramatical o el léxico. La aplicación de cuestionarios es el método más efectivo de afrontar el estudio de la variación léxica. Dada la condición de comunidad urbana de Madrid, la aplicación de cuestionarios dialectales como los utilizados en comunidades rurales no es posible, y el objetivo general del proyecto PRESEEA no lo convierte en una necesidad. Los materiales recogidos a través de la entrevista semidirigida resultan adecuados y pueden posibilitar el estudio del léxico desde diversos puntos de vista; por ejemplo, un estudio interesado en el análisis cuantitativo de diversos elementos del caudal léxico de una comunidad (préstamos de otras lenguas, unidades neológicas, cultismos, etc.) podría utilizar la información comprendida en cualquier muestra de habla espontánea de cierta extensión, como las de los corpus PRESEEA.

El estudio del nivel sintáctico es el más complejo, dada la naturaleza lingüística de las unidades sintácticas y las dificultades de conseguir un buen rendimiento de las técnicas más empleadas. Por esta razón, se centra la atención en aquellos aspectos de la gramática del español que, aparentemente, precisan de un examen sociolingüístico. Al igual que para la fonética y la fonología hay un consenso en el mundo hispánico sobre qué variación de segmentos presenta un correlato social y lingüístico, no existe tal consenso para las variables en el nivel de la morfosintaxis[19]. En cualquier caso, se continúa prestando especial atención a las variables morfosintácticas que se detallan a continuación, establecidas por Cestero y Molina[20] para la recogida de los corpus PRESEEA-Comunidad de Madrid (Moreno, Cestero, Molina y Paredes 2002, 2005b y 2007; Cestero, Molina y Paredes 2012; Molina, Paredes y Cestero 2014; Paredes, Cestero y Molina 2015; Cestero, Molina y Paredes 2020; Molina, Paredes y Cestero 2020a; Paredes, Cestero y Molina 2020):

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.

---

[19] Esto se debe, en gran parte, a que la naturaleza de la variación sintáctica no es análoga a la de la variación fonológica (Moreno Fernández 2009: 28; Silva-Corvalán 2001: 129-132). Además, son escasos los autores que se han encargado de identificarlas y formularlas de manera sistemática, a pesar de la abundancia de trabajos que se han centrado en la variación en el nivel morfosintáctico desde principios del siglo XXI. Entre los que lo han hecho, destacan los trabajos de C. Silva-Corvalán (1994) y de P. Martín Butragueño (1994), en los que se apoyaron Ana M. Cestero e Isabel Molina para la selección de variables sintácticas con las que trabajaron para el diseño de la encuesta de los corpus PRESEEA-Comunidad de Madrid.

[20] Aunque se han publicado interesantes trabajos sobre variación gramatical desde la publicación del primer sub-corpus de PRESEEA-Madrid (Fernández-Ordóñez 2011; Veiga Rodríguez 2019), se han mantenido las mismas variables con el propósito de permitir la coherencia con las muestras anteriores y la comparabilidad con todo el corpus PRESEEA.

   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*; se trata de examinar sus significados, usos e interacción con marca de tiempo y aspecto.
6. Variación en el uso de construcciones de tipo "reflexivo". Se trata de examinar cuáles son los patrones de concordancia, así como la extensión del patrón 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.

Para posibilitar la aparición de distintas variables morfosintácticas como las que se acaban de listar, es necesario desarrollar una entrevista semidirigida, con cierto grado de estructuración, que promueva la aparición de diferentes tipos de discurso (narración, exposición, descripción, argumentación, discurso hipotético...). Así, la obtención de diversas variables gramaticales quedaría posibilitada por la entrevista semidirigida; estos materiales también permitirían conseguir muestras en los otros niveles descritos anteriormente.

La entrevista es una conversación semidirigida que, al igual que en los otros corpus PRESEEA-Madrid, tiene lugar entre un investigador y un informante cada vez y, en algunas de las entrevistas cara a cara, cuenta con una tercera persona que colabora con labores de grabación. Tradicionalmente, la forma de recogida de datos ha sido con un magnetófono (Ander-Egg 1987) o una grabadora portátil (Juan y Bürki 2022); no obstante, dada la gran innovación y los avances tecnológicos que han sufrido las herramientas y técnicas de recogida de datos, las entrevistas fueron grabadas en vídeo de dos maneras distintas, ambas novedosas en PRESEEA. En el caso de los seis informantes del tercer grupo de edad, teniendo en cuenta que los informantes podrían estar menos familiarizados con las nuevas tecnologías y tener dificultades para conectar y actuar por videoconferencia, las encuestas se realizaron cara a cara y se grabaron con una cámara de vídeo digital SONY Handycam FDR-

AX53, intentando que el plano abarcara tanto al informante como al entrevistador para facilitar el análisis posterior de elementos no verbales.

En los dos grupos etarios restantes, se decidió[21] emplear la plataforma audiovisual Microsoft *Teams* y realizar las entrevistas por videoconferencia. Las plataformas y aplicaciones que permiten interactuar por videoconferencia y grabar las sesiones íntegramente, cuyo uso se popularizó durante la pandemia de la COVID-19 del 2020 y parte del 2021, son fruto del desarrollo de las Tecnologías de la Información y la Comunicación (TIC). Asimismo, han alterado la naturaleza del contexto en el que tiene lugar la investigación, que ahora incluye el espacio virtual (Juan y Bürki 2022: 266), y traen consigo unas claras ventajas y algunos retos que se deben considerar a la hora de llevar a cabo una recogida de datos con ellas.[22] Por una parte, entre los aspectos positivos, destacan el bajo coste de desplazamiento a un lugar concreto para llevar a cabo la entrevista; desde el punto de vista logístico, son más eficientes que una entrevista tradicional, ya que no resulta necesario determinar un lugar donde realizarla; también permiten mayor flexibilidad y facilidad para los interactuantes a la hora de fijar una fecha y hora para la entrevista y llevarla a cabo, lo que resulta en mayor comodidad para todos los participantes. Esta mayor flexibilidad se mostró, en alguna ocasión, como el factor decisivo para aquellos informantes que tenían dudas en participar en el proyecto. No podemos olvidar, además, otra de las ventajas que tiene este tipo de plataformas: la posibilidad de grabar la imagen y el sonido del desarrollo de la charla, de manera, quizás, no tan intrusiva como una cámara de vídeo enfocando continuamente a los participantes en la interacción. La plataforma que elegimos no solamente tiene la capacidad de grabar directamente las sesiones, sino que, al finalizar la grabación, se genera un archivo MP4 descargable con el contenido de la entrevista.

Por otra parte, esta forma de recogida de muestras no está exenta de algunos problemas, o desventajas, metodológicas. En primer lugar, parece que existe una mayor dificultad para conseguir entendimiento y compenetración y establecer empatía a través de las entrevistas mediadas, debido a la diferencia entre presencialidad virtual y presencialidad física, así como a las dificultades para interactuar con el entorno y el interlocutor[23]. Esto tiene estrecha relación con otro de los retos: la correcta interpretación y lectura de las señales no verbales a causa de problemas de conexión, como la pérdida de señal, las interrupciones de audio y el congelamiento de la imagen, que obstaculizan el proceso de la entrevista y, de ser muy graves y continuas, podrían dificultar las tareas posteriores de análisis (afortunadamente, estos problemas no han

---

[21] Esta decisión fue tomada por Ana M. Cestero Mancera, organizadora del corpus PRESEEA-Madrid XXI.

[22] Para una revisión de los aspectos positivos y negativos de las entrevistas mediadas, véase Hine (2000), Juan y Bürki (2022), Davis (2013), Archibald *et al.* (2019), Ardévol *et al*. (2003) y las referencias que allí se encuentran.

[23] Algunos autores (Juan y Bürki 2022: 267) apuntan que no hay consenso en relación con este aspecto; otras (Hine 2000: 58) afirman que las posibilidades de las interacciones mediadas posibilitan que se repiense el rol de la presencia física como fundamento.

sido muy numerosos en el corpus que presentamos). Otro escollo que encontramos fue la alfabetización electrónica de los participantes. Como hemos mencionado, a pesar de la creciente costumbre de la interacción a través de dispositivos electrónicos y plataformas de comunicación, no todos los informantes conocían la plataforma seleccionada, por lo que fue necesario elaborar un documento[24] con instrucciones detalladas de los pasos que tenían que seguir para realizar una correcta conexión e ingresar en el espacio virtual en la fecha y hora fijadas de antemano con el entrevistador. Por último, las distracciones de los participantes son otra potencial dificultad que hay que considerar en la recogida de datos con entrevistas mediadas por dispositivos electrónicos. Los informantes, al no estar en un espacio físico compartido, gozan de mayor privacidad, lo que puede llevarlos a distraerse más fácilmente con otros dispositivos (por ejemplo, el teléfono móvil) o con interrupciones en el entorno.

La tarea del investigador o entrevistador es, esencialmente, orientar la conversación para asegurarse de que aparecen, en el orden establecido previamente, los tipos de discurso que promueven la manifestación de ciertas variables lingüísticas. Por lo tanto, se espera que su discurso sea muy inferior cuantitativamente si lo comparamos con el del informante (Tagliamonte 2006: 39). Por lo general, la entrevista sociolingüística favorece la recogida de discursos de estilo semiformal, no obstante, debido a su duración (unos 50 minutos como media) y el contenido de ciertos temas, el acceso a estilos de habla más informales resulta posible.

Durante la entrevista, se emplea una gran variedad de estrategias de encuesta, encargadas de hacer aparecer los tipos de discurso y las variables que resulten de interés. Por una parte, hay un conjunto de variables, entre las que se enumeran más arriba, que no necesitan estrategias concretas, pues presentan un alto índice de frecuencia, como ocurre con ciertas unidades fonético-fonológicas. Estas son las variables de la lista propuesta por Molina Martos y Cestero Mancera que pueden aparecer en el transcurso de la conversación: *expresión y posición de sujeto, nominal y pronominal (1); orden de los argumentos (2)*; *variación en el uso de construcciones de tipo reflexivo (6)*; *estilo directo e indirecto (9)*; *deícticos pleonásticos (10)*; *leísmo, laísmo, loísmo (11)*; *duplicación de clíticos (12)*; *posición de clíticos (13)*; *posición de OD y OI (14)*; *perífrasis (18)*.

Las variables restantes se recogen mediante las preguntas que se detallan a continuación:

- *Valores y usos del sistema verbal (4)*
  - *Indicativo-subjuntivo/condicionales (4.1)*: los módulos temáticos contienen preguntas sobre situaciones hipotéticas; se extraerán grados de probabilidad de ocurrencia de distintos hechos o realidades (discurso hipotético).
  - *Presente/pretérito con valor de pasado (4.2)*: en los distintos módulos te-

[24] Este documento fue diseñado y creado por Ana M. Cestero Mancera.

máticos se incluyen preguntas de situaciones pasadas (narraciones).
    - *Presente con valor de futuro (4.3)*: los módulos temáticos incluyen preguntas sobre situaciones futuras.
    - *Indicativo/subjuntivo en oraciones subordinadas introducidas por las partículas para que, cuando... (4.4):* en los distintos módulos temáticos se incluyen preguntas destinadas a la obtención de los diferentes tipos de oraciones subordinadas (sustantivas, adjetivas, adverbiales).
- *Verbos semiauxiliares de modalidad epistémica, dinámica y deóntica*: *haber de/que, tener que, deber (de), poder, ser capaz de (5):* consiste en examinar sus significados, usos e interacción con marca o indicación de tiempo y aspecto: los distintos módulos temáticos contienen preguntas directas para conseguir estas formas.
- *Usos de los verbos ser y estar (7)*: se le solicita al informante una descripción de una persona y un lugar (descripciones).
- *Usos de los verbos haber y estar (8)*: se le pregunta al informante por el interior de ciertos lugares, pidiendo una descripción (descripciones).
- *Concordancia del verbo con usos impersonales (15)*: durante la conversación se introducen preguntas directas mediante las que se incita al informante a opinar sobre cuestiones en las que el sujeto es "los países" u otras expresiones marcadoras de impersonalidad (argumentaciones).
- *Impersonalidad (16) cómo se marca* (*se*, *tú*): en el módulo de "las costumbres" se le pide al informante que describa los pasos para seguir una receta de cocina (instrucciones).
- *Uso de los verbos haber* y *hacer* (*hay/hace niebla*...) *(17)*: se incluye un módulo sobre "el tiempo" (descripciones).

2.2.2. *La entrevista*

La entrevista se estructura teniendo en cuenta los módulos temáticos que se exponen a continuación:

Tabla 6. Módulos temáticos para entrevistas

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

A partir de estos componentes, PRESEEA ha empleado una entrevista diseñada para recoger muestras de habla en Madrid a la que se han aplicado pequeñas variaciones en cada sub-corpus para que se adapte de mejor modo al grupo de informantes entrevistado[25]. En este sub-corpus (PRESEEA-Madrid XXI) la entrevista ha seguido el siguiente esquema:

1. SALUDOS

Presentación.
Preguntar por forma de tratamiento que desea que se usE: ¿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 tráteme de tú ¿le/te parece?
Interés por el estado del interlocutor: ¿Cómo estás/está? ¿Estás animado/a?
En estos días todos estamos un poco raros, yo creo que es por el tiempo, ¿verdad?

2. NUEVAS FORMAS DE COMUNICACIÓN

Las nuevas tecnologías nos permiten hacer videoconferencias, comunicarnos en línea, como ahora, con lo que podemos quedar en cualquier momento sin desplazarnos. ¿Qué piensa de esta forma de comunicación que se va haciendo cada vez más frecuente?
¿Qué ventajas cree que tiene? ¿qué inconvenientes le ve?
Y con relación a la mayor o menor comodidad, ¿qué opina?

3. EL TIEMPO

¡Qué calor/frío ha hecho hoy!
A mí no me gusta el verano/invierno ¿Tú cuál prefieres?
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?

[25] Ana M. Cestero Mancera añadió algunas cuestiones relativas a la popularización de los medios de comunicación a distancia, vía online, en la entrevista semidirigida.

4. LUGAR DONDE VIVE

(Mirada alrededor de la casa) ¿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?
¿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 Madrid o en tu barrio 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?

5. 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 que 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?

6. 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 Madrid? ¿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?

7. 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).

8. 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?

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

10. 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.

Las unidades temáticas fomentan el desarrollo de las modalidades discursivas de la siguiente manera:

Tabla 7. Módulos temáticos y modalidades discursivas

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

2.2.3. *La recogida de materiales*

Las entrevistas que componen el sub-corpus de PRESEEA-Madrid XXI (distrito de Salamanca), 18 en total, fueron realizadas entre marzo de 2022 y octubre de 2023 por los investigadores del equipo PRESEEA de la Universidad de Alcalá[26]. Las conversaciones semidirigidas resultaron de la interacción entre los tipos de participantes que se describe a continuación:

1. Entrevistadores: las entrevistas fueron semidirigidas por los investigadores de la Universidad de Alcalá, por medio de los módulos temáticos presentados, diseñados específicamente para reunir distintos tipos de discurso y aspectos morfosintácticos susceptibles de variación, siguiendo la pauta de que ambos interlocutores, entrevistado y entrevistador, fueran del mismo sexo.
2. Informantes: cada conversación se llevó a cabo con informantes diferentes, sobre los que se ejecutó el control de las variables sociales que se han tenido en cuenta (edad, sexo y nivel de instrucción).

Los lugares en los que se realizaron las grabaciones del tercer grupo de edad fueron las casas de los informantes o las aulas de un centro público situado en el distrito de Salamanca[27]. En el caso de los entrevistados del primer y el segundo grupo de edad, se realizó mediante la herramienta de videoconferencias Microsoft *Teams*. En las entrevistas hechas cara a cara se dejó la cámara de vídeo en un lugar visible, situándola a una distancia adecuada de los interlocutores que permitiera abarcar el plano completo del informante[28].

Como ya se ha mencionado, la recogida de datos que se realizó en persona se llevó a cabo utilizando una grabadora de vídeo digital SONY Handycam FDR-AX53 con micrófono incorporado para no dificultar la gesticulación de los participantes, a quienes también se les permitía libertad de movimiento. En todo momento, se buscó propiciar un ambiente de tranquilidad y seguridad para los informantes, procurando que no hubiese nada en el entorno que les pudiera hacer sentir cohibidos o reprimidos. Se parte del supuesto de que, en las entrevistas semidirigidas, los informantes no están

---

[26] El diseño del corpus, la búsqueda y selección de informantes y la organización de recogida de la muestra fueron realizados por Ana M. Cestero Mancera y Nicolás Moreno Martín de Nicolás. Ana M. Cestero, Isabel Molina, Francisco Moreno y Florentino Paredes se encargaron de entrevistar a todos los informantes mayores de 35 años, mientras que Nicolás Moreno, Sara Engra y Gema Cuesta entrevistaron a los informantes más jóvenes, al ser de edades cercanas a estos.

[27] Aprovechamos este espacio para reiterar el agradecimiento del equipo a todos los informantes que han participado en la creación de este corpus, así como a cuantos ayudaron en la tarea de encontrar sujetos dispuestos a colaborar. Asimismo, hacemos constar el agradecimiento del equipo al CEPA Joaquín Sorolla de Madrid, por la ayuda y la colaboración prestada tanto en la búsqueda de informantes como en la realización de las entrevistas para el proyecto con el que se ha recogido y publicado el corpus PRESEEA-Madrid XXI.

[28] Todos los participantes en las interacciones sabían desde un principio que iban a ser grabados.

en una situación natural y que, por lo tanto, hay altas probabilidades de que empleen un registro formal y elaborado. Sin embargo, se ha podido comprobar, al igual que en los corpus anteriores de la comunidad de Madrid (Alcalá de Henares, distrito de Salamanca y distritos de Vallecas), que, según avanza la entrevista, los informantes se implican en la conversación y se olvidan, con ello, de la presencia de la grabadora.

En cuanto a las entrevistas recogidas por medio de Microsoft *Teams*, el proceso es distinto en varios aspectos. Por una parte, al tratarse de una herramienta que ha de ser manipulada por todos los participantes, es necesario asegurarse de que el informante tiene la cámara y el micrófono del dispositivo desde el que accede a la videoconferencia activados. Asimismo, hay que comprobar que la cámara del informante se encuentra en el plano adecuado, encuadrado desde la zona del tórax, como límite inferior, hasta bien pasada la cabeza, como límite superior. De esa manera, el entrevistador se asegura de recoger todo el componente no verbal y los movimientos corporales del informante. Por otra parte, es de suma importancia recordar la activación de la grabación, pues no se inicia automáticamente. Una vez se haya comprobado que el encuadre es el adecuado y que tanto el audio como el video están en funcionamiento, se inicia la grabación. Al finalizar la entrevista, se detiene y, una vez hecho esto, se genera automáticamente un archivo descargable con la grabación de la videoconferencia que acaba de tener lugar.

2.3. Transcripción y etiquetado del corpus

El sub-corpus de datos de PRESEEA-Madrid XXI (distrito de Salamanca) está completo; se ha concluido la recogida de materiales, su transcripción[29] y almacenamiento. Las grabaciones se han transcrito en ortografía ordinaria, empleando el procesador de textos "Word" ™, reproduciendo las grabaciones en formato digital y manejándolas con el aparato manual con el que se dé órdenes del dispositivo que se utilice, siguiendo, como en el resto de los corpus PRESEEA, las directrices definidas por F. Moreno Fernández[30] para los materiales PRESEEA, que han guiado esta etapa de transcripción y almacenamiento. Los archivos correspondientes a cada grabación están conservados en PC, además de en copias de seguridad. Cada una de las transcripciones preparada para el análisis cuenta con una doble corrección, llevada a cabo por dos personas diferentes[31].

---

[29] La transcripción de las grabaciones ha sido realizada por Javier González Ruiz, como parte del contrato asociado al 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) subvencionado por el Ministerio de Ciencia e Innovación-Agencia Estatal de Investigación/10.13039/501100011033.

[30] Moreno Fernández 1996, actualizadas en Moreno Fernández 2021c.

[31] Las revisiones han sido realizadas por Javier González Ruiz, Nicolás Moreno Martín de Nicolás, Sara Engra Minaya y Gema Cuesta Chorro.

Las normas empleadas para la transcripción de los materiales están fundamentadas en el sistema *SGLS* y adoptan las pautas internacionales de la TEI (*Text Encoding Initiative*)[32]. Estas normas o convenciones están compuestas por diversas etiquetas y marcas previamente acordadas por los miembros de los equipos que participan en PRESEEA, y expresan una gran diversidad de rasgos de los textos transcritos.

3. Los materiales de la lengua hablada en PRESEEA-Madrid XXI (distrito de Salamanca)

3.1. Presentación de los materiales

Como hemos mencionado, los materiales que constituyen el sub-corpus PRESEEA-Madrid XXI (distrito de Salamanca) se encuentran transcritos y almacenados informáticamente en la Universidad de Alcalá. Su matriz, que recoge la codificación *SGML*, permite un intercambio informático sencillo y una presentación según las normas internacionales. La muestra de este sub-corpus está disponible para todos los equipos de investigación de otras ciudades del mundo hispánico que formen parte de la red de equipos PRESEEA.

El deseo del equipo PRESEEA-Comunidad de Madrid es siempre facilitar el acceso a los materiales para que puedan emplearse con distintos fines; tanto profesores (de diversos niveles) como especialistas de diversas materias pueden precisar la consulta de muestras de lengua oral. Ya que limitar el acceso de los materiales de lengua hablada de Madrid a los círculos de especialistas en sociolingüística nunca ha sido su deseo y con el fin de hacerlos más accesibles a un rango más extenso de posibles lectores, el equipo PRESEEA de la Universidad de Alcalá resolvió entregar los materiales a la imprenta de la universidad.

Para la edición del sub-corpus PRESEEA-Madrid XXI (distrito de Salamanca) se han adoptado criterios similares a los estipulados en la publicación del corpus de Alcalá de Henares (Moreno, Cestero, Molina y Paredes 2002, 2005b y 2007), del corpus de Madrid (distrito de Salamanca) (Cestero, Molina y Paredes 2012; Molina, Paredes y Cestero 2014; Paredes, Cestero y Molina 2015), así como del corpus de PRESEEA-Madrid (distritos de Vallecas) (Cestero, Molina y Paredes 2020; Molina, Paredes y Cestero 2020a; Paredes, Cestero y Molina 2020), incluyendo alguna adaptación de ciertas convenciones de presentación a los nuevos acuerdos que se han ido adoptando en las reuniones de coordinación del proyecto PRESEEA. Los materiales se presentan en un volumen que ofrece las transliteraciones de las entrevistas mantenidas con 18 hablantes de Madrid (distrito de Salamanca).

---

[32] Véase Moreno Fernández 2021a y 2021c.

Cada entrevista corresponde a uno de los 18 informantes y aparece transcrita en ortografía ordinaria, aunque emplea un conjunto de etiquetas o códigos cuyo objetivo es acercar la lectura a la enunciación oral de los discursos. El sub-corpus se ofrece en una única versión en la que se utilizan las marcas y etiquetas mínimas comunes de PRESEEA (Moreno Fernández 2021c).

A partir de las directrices descritas en Moreno Fernández (2021c), las etiquetas manejadas en los textos marcados son las que se describen a continuación:

ORTOGRAFÍA Y PUNTUACIÓN

- Signos de admiración [¡!] señalan los enunciados exclamativos.
- Signos de interrogación [¿?] señalan los enunciados interrogativos.
- Barra [/] representa pausa mínima.
- Barra doble [//] indica pausa.
- Dos puntos [:] marca el inicio del turno de cada interlocutor

ETIQUETADO DE RUIDOS

- <ruido = “ ”/> marca la producción de ruido, con especificación del tipo de ruido del que se trate.
- <ruido_de_fondo> </ruido_de_fondo> señala la emisión producida con ruido de fondo.
- <risas = “ ”/> marca la ocurrencia de risa y el interlocutor que la produce.
- <entre_risas> </entre_risas> marca la emisión producida entre risas.
- <registro_defectuoso> </registro_defectuoso> indica que el fragmento de grabación que acota no es de buena calidad.
- <interrupción_de_grabación/> señala que se ha interrumpido la grabación.

ETIQUETADO FÓNICO

- <énfasis> </énfasis> señala que la emisión de un fragmento se ha realizado con pronunciación enfática.
- <alargamiento/> indica el alargamiento de sonidos.
- <silencio/> señala que, durante un segundo o más, hay una ausencia de emisión.
- <palabra_cortada> indica que una palabra no se ha emitido completamente.
- <vacilación/> marca un breve titubeo en la emisión.
- <sic> </sic> comunica que no hay un descuido en la transliteración, y que la emisión que aparece transliterada es la del informante.
- <ininteligible/> señala aquella emisión de discurso que no se ha podido reconocer o comprender.

ETIQUETADO LÉXICO

- <término> </término> señala una lexía empleada como uso especializado.
- <extranjero> </extranjero> marca los extranjerismos.

– <siglas = [ ]> </siglas> indica las siglas y su pronunciación.

ETIQUETADO DE DINÁMICA DISCURSIVA
– <cita> </cita> indica una emisión producida en estilo directo, si es una cita.
– <simultáneo> </simultáneo> señala un solapamiento de discurso del informante y del entrevistador –emisión simultánea.

ETIQUETADO DE TRANSCRIPCIÓN
– <transcripción_dudosa> </transcripción_dudosa> marca un fragmento de discurso que resulta dudoso para el transcriptor y los revisores.
– <tiempo = “ ”/> anotación del minuto y segundo de grabación (se constata cada dos minutos aproximadamente, en la intervención del informante).
– <observación_complementaria = “ ”/> señala toda observación que no haya sido anticipada en las otras etiquetas y que el transcriptor o revisores consideren oportuna.

En la transcripción de las entrevistas se reemplazan por distintas iniciales (sin puntos detrás) los nombres correspondientes a lugares y personas que permiten la identificación de los informantes. Se hace con el propósito de garantizar el anonimato de los participantes de las entrevistas y de aquellas personas que han contribuido al trabajo de investigación con su participación. Cada interlocutor es distinguido con una letra, correspondiente a la primera de la función que desempeña en la entrevista: el entrevistador se identifica con la letra E y el informante con la I. Asimismo, como se indica en los documentos de trabajo PRESEEA[33], se han adoptado los siguientes criterios:

a. Cada toma de palabra por parte de un interlocutor se considera un turno diferente, sea turno de habla o sea turno de apoyo, y se transcribe en párrafo aparte.
b. El uso de mayúsculas está limitado a los nombres propios, a las iniciales de nombres de lugares y personas y para las siglas.
c. Los signos de puntuación no se utilizan. Las pausas se señalan de la manera que se ha indicado anteriormente.
d. Los nombres propios se transcriben con su grafía vernácula.
e. Los elementos cuasi-léxicos funcionales (interjecciones, apoyos, onomatopeyas (*ah, ay, bum, ring, aha, mm, ñam, eeh, pff, hm*, *plas*, *zas* etc) se transcriben en escritura ortográfica.

Cada entrevista se identifica con un encabezado, a modo de presentación, que contiene el número de la entrevista en la publicación (ejemplo: Entrevista n.º 03), además de una línea con los datos que caracterizan al informante (edad, estudios y

[33] Consúltese Moreno Fernández 2021c.