

Léxico disponible de los jóvenes preuniversitarios de la Comunidad de Madrid

Florentino Paredes García

Luis Guerra Salas

M.^a Elena Gómez Sánchez



Léxico disponible de los jóvenes preuniversitarios de la Comunidad de Madrid

**UAH MONOGRAFÍAS
HUMANIDADES 90**

Comité científico

- JAIME ALVAR ÉZQUERRA
Catedrático de Historia Antigua de la Universidad Carlos III.
- MANUEL ALEJANDRO RODRÍGUEZ DE LA PEÑA
Profesor titular de Historia Medieval de la Universidad San Pablo-CEU.
- CARLOS ALBERTO GONZÁLEZ SÁNCHEZ
Catedrático de Historia Moderna de la Universidad de Sevilla.
- FERNANDO MARIAS FRANCO
Catedrático de Historia del Arte de la Universidad Autónoma de Madrid.
- RAFAEL BONILLA CEREZO
Profesor titular de Literatura de la Universidad de Córdoba.
- CELIA CASADO FRESNILLO
Profesora titular de Filología Hispánica de la Universidad Nacional de Educación a Distancia.
- JULIO DE LA CUEVA
Profesor contratado doctor de Historia Contemporánea de la Universidad de Castilla-La Mancha.

Léxico disponible de los jóvenes preuniversitarios de la Comunidad de Madrid

Florentino Paredes García
Luis Guerra Salas
M.^a Elena Gómez Sánchez



Universidad
de Alcalá

EDITORIAL
UNIVERSIDAD DE ALCALÁ

La colección de Monografías de Humanidades de la Editorial de la Universidad de Alcalá ha sido distinguida con el Sello de Calidad en Edición Académica - Academic Publishing Quality (CEA-APQ).



El contenido de este libro no podrá ser reproducido, ni total ni parcialmente, sin el previo permiso escrito del editor. Todos los derechos reservados.

- © De los textos: sus autores.
- © De las imágenes: sus autores.
- © De la ilustración de portada: Begoña Moreno *
- © Editorial Universidad de Alcalá, 2022

Plaza de San Diego, s/n
28801 Alcalá de Henares
www.uah.es

I.S.B.N.:978-84-18979-97-2
<https://doi.org/10.37536/SZPS9864>

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

* La portada es una representación infográfica de asociaciones léxicas obtenidas en el *CI 14. Los animales*, superpuesta a una imagen generada mediante Dall-e 2, una inteligencia artificial desarrollada por OpenAI.

Esta obra se inscribe dentro de las actividades científicas de los siguientes proyectos de investigación: *AGENDA 2050. Procesos de variación y cambio espaciales y sociales (VARES-AGENDA 50; PID2019-104982GB-C51)*, del Ministerio de Ciencia, Innovación y Universidades, y *La población migrante de la Comunidad de Madrid: factores lingüísticos, comunicativos, culturales y sociales del proceso de integración y recursos lingüísticos de intervención (INMIGRA3-CM; H2019/HUM-5772)*, financiado por la Comunidad de Madrid y el Fondo Social Europeo.

ÍNDICE

Mi palabra es prenda de oro. PRÓLOGO, POR FRANCISCO MORENO FERNÁNDEZ	11
---	----

PRIMERA PARTE

LA INVESTIGACIÓN SOBRE EL LÉXICO DISPONIBLE DE MADRID

1. INTRODUCCIÓN. LA DISPONIBILIDAD LÉXICA.....	15
1.1. Los estudios de disponibilidad léxica.....	15
1.1.1. Fundamentos teóricos y metodológicos.....	16
1.1.2. El Proyecto Panhispánico de Disponibilidad Léxica (PPHDL)	18
1.2. El proyecto de estudio del léxico disponible de Madrid (LDM)..	21
2. ASPECTOS METODOLÓGICOS	23
2.1. El territorio	23
2.2. La población estudiada.....	24
2.3. La muestra	25
2.3.1. Criterios de selección	25
2.3.2. Configuración de la muestra.....	29
2.4. La obtención del léxico disponible: la encuesta	30
2.4.1. El diseño de la encuesta	30
2.4.2. La aplicación de la encuesta.....	32
2.5. La edición de los materiales	33
2.5.1. Consideraciones previas.....	33
2.5.2. Criterios generales de edición del LDM	35
2.5.3. Criterios específicos de edición para cada centro de interés	40
2.6. Tratamiento estadístico de los materiales.....	92
2.6.1. El índice de disponibilidad léxica estandarizado (DISP_st)	93
2.6.2. El índice de compatibilidad estandarizado (COMP_st)...	94
2.6.3. El índice de descentralización léxica (IDD).....	95
2.6.4. El índice de contigüidad de los vocablos (ICV).....	96

3. RESULTADOS DE LA INVESTIGACIÓN	101
3.1. Resultados cuantitativos generales	101
3.1.1. Número de palabras (<i>tokens</i>).....	101
3.1.2. Número de vocablos (<i>types</i>).....	103
3.1.3. Índice de cohesión (ICOH) e índice de densidad (DENS).....	105
3.1.4. Rango de dispersión, medias y desviación típica.....	107
3.2. Comparación con otros proyectos vinculados al PPHDL	108
3.3. Las variables sociológicas	112
3.3.1. La variable sexo	112
3.3.2. La variable tipo de educación recibida.....	116
3.3.3. La variable localidad de residencia	120
3.3.4. La variable nivel sociocultural	124
3.3.5. La variable área territorial en la Comunidad de Madrid	129
4. ASPECTOS CUALITATIVOS Y DESARROLLOS DIDÁCTICOS.....	133
4.1. Las categorías gramaticales.....	134
4.2. Los adjetivos.....	137
4.3. Neologismos y creaciones léxicas	139
4.4. Notas ortográficas.....	143
4.5. Relaciones asociativas.....	145
4.6. Propuestas didácticas derivadas del LDM.....	147
5. CONCLUSIONES.....	151
REFERENCIAS BIBLIOGRÁFICAS	155

SEGUNDA PARTE EL LÉXICO DISPONIBLE DE MADRID

REPERTORIOS LÉXICOS

A.1. Diccionarios ordenados por disponibilidad.....	167
CI 01. El cuerpo humano	167
CI 02. La ropa	177
CI 03. Partes de la casa	186
CI 04. Los muebles de la casa.....	201
CI 05. Comidas y bebidas	211
CI 06. Objetos colocados en la mesa para la comida.....	229
CI 07. La cocina y sus utensilios.....	242
CI 08. La escuela: muebles y materiales.....	258
CI 09. Calefacción, iluminación y ventilación.....	278
CI 10. La ciudad.....	297

CI 11. El campo.....	323
CI 12. Medios de transporte.....	351
CI 13. Trabajos de campo y de jardín.....	366
CI 14. Los animales.....	390
CI 15. Juegos y distracciones.....	404
CI 16. Profesiones y oficios.....	430
CI 17. Los colores.....	452
CI 18. Sensaciones y sentimientos.....	461
CI 19. Adjetivos.....	493
CI 20. Medios de comunicación.....	537
A.2. Diccionarios ordenados alfabéticamente.....	569
CI 01. El cuerpo humano.....	569
CI 02. La ropa.....	574
CI 03. Partes de la casa.....	578
CI 04. Los muebles de la casa.....	586
CI 05. Comidas y bebidas.....	591
CI 06. Objetos colocados en la mesa para la comida.....	600
CI 07. La cocina y sus utensilios.....	606
CI 08. La escuela: muebles y materiales.....	615
CI 09. Calefacción, iluminación y ventilación.....	625
CI 10. La ciudad.....	634
CI 11. El campo.....	647
CI 12. Medios de transporte.....	662
CI 13. Trabajos de campo y de jardín.....	669
CI 14. Los animales.....	681
CI 15. Juegos y distracciones.....	688
CI 16. Profesiones y oficios.....	701
CI 17. Los colores.....	712
CI 18. Sensaciones y sentimientos.....	717
CI 19. Adjetivos.....	733
CI 20. Medios de comunicación.....	755
ANEXOS.....	771
Anexo 1. Protocolo de aplicación de la encuesta y cuadernillo de respuestas.....	771
Anexo 2. Test de ANOVA de un factor para las variables sociales....	781

MI PALÁBRA ES PRENDA DE ORO

Francisco Moreno Fernández
Universidad de Alcalá

Palabras para hablar de palabras. Se dice que la palabra «explica los conceptos del ánimo, y es propia solo del hombre». Lo dijo la Real Academia Española en 1737. Se dice que «el concepto es tan incapaz de librarse de la palabra como lo es el hombre de prescindir de sus rasgos faciales» y que «la palabra es la configuración individual del concepto, y si éste quiere abandonarla, lo único que logrará es volver a hallarse en otra palabra». Lo dijo Wilhelm von Humboldt en 1836. Se dice que «las relaciones entre las palabras y los sentidos son numerosas y complejas» y que «se sabe que la historia de las palabras no puede escribirse sin un conocimiento previo y profundo de las cosas que designan». Lo dijo Albert Dauzat en 1922. Se dice que la función léxica es «la estructuración primaria de la experiencia por medio de las «palabras», idealmente anterior a las funciones necesarias para la combinación de las palabras en el discurso». Lo dijo Eugenio Coseriu en 1964. Y se dice que «las propiedades semánticas de las palabras y las construcciones vienen determinadas por las formas en que se constituyen, con una rica contribución innata». Lo dijo Noam Chomsky en 2000. La palabra siempre ha sido ocupación prioritaria de los lingüistas y, en el siglo XXI, seguirá siendo así.

La palabra es un elemento tan importante en la adquisición y el uso de la lengua, en el desarrollo cognoscitivo del hablante, en la enseñanza de la lengua y de la cultura, que todo esfuerzo que se haga, para desentrañarla y explicarla, es poco. Por eso se antojan necesarios los estudios sobre las mil caras de las palabras; por eso es imprescindible conocer cuáles se usan y cuáles no, qué lugar ocupan en los mecanismos de producción lingüística, cuáles asumen valores simbólicos extraordinarios y cuáles no o cómo se ajustan a las nuevas referencias. La finalidad de este esfuerzo explicativo es el crecimiento de nuestros saberes sobre la lengua, la mente, la sociedad, y las relaciones entre todo ello; pero la trascendencia de los conocimientos teóricos no quita un ápice de valor a su posible aplicación práctica porque a menudo se justifican en ella.

El libro que tengo el honor de prologar, *Léxico disponible de los jóvenes preuniversitarios de la Comunidad de Madrid*, es un buen ejemplo de cómo se conjuga el

fundamento teórico con la dimensión aplicada. Estamos ante un trabajo muy pegado al terreno, con datos precisos de hablantes precisos, recogidos mediante técnicas probadas por la vía del empirismo. Desde este punto de vista, la información que se ofrece sobre la disponibilidad léxica en estudiantes de la Comunidad de Madrid proporciona un perfil lingüístico sumamente interesante para su aplicación práctica: cuanto mejor se conocen los rasgos lingüísticos de unos alumnos, mejor se identifican sus limitaciones y mejor se pueden planificar, por tanto, las estrategias de enseñanza. Pero aquí no acaba todo.

Conviene que el lector ajeno a este tipo de investigaciones sepa que el análisis del léxico disponible no es un endemismo madrileño, ni siquiera español. Desde hace bastantes años, se vienen recogiendo muestras léxicas con idénticos fines en numerosos lugares del mundo hispánico: San Juan de Puerto Rico, México, Santiago de Chile, Santo Domingo, San José de Costa Rica, en América; Madrid, Cádiz, Almería, Granada, Zamora, Las Palmas de Gran Canaria, en España, entre otros. El impulsor y coordinador de tan larga serie de investigaciones es Humberto López Morales y el fruto de su empeño se está recogiendo en infinidad de tesinas, tesis doctorales, manuales de enseñanza de español y currículos escolares. Conociendo el individualismo que aqueja a la investigación científica del mundo hispánico, no es un mérito conseguir la coordinación de tantos trabajos en tantos países: simplemente es un milagro.

El equipo que ha preparado este *Léxico disponible de los jóvenes preuniversitarios de la Comunidad de Madrid*, conoce los centros de Educación Secundaria de la Comunidad de Madrid y está vinculado a los proyectos de investigación sobre disponibilidad léxica en el Departamento de Filología, Comunicación y Documentación de la Universidad de Alcalá y en la Universidad Europea de Madrid. La persona que mejor representa tal síntesis es Florentino Paredes, catedrático de la Universidad y buen conocedor de la Enseñanza Secundaria de la zona. Su labor de coordinación ha resultado sencillamente decisiva para el inicio y el buen fin del trabajo. Por lo demás, el mérito de recoger datos de decenas de hablantes, analizarlos adecuadamente, interpretarlos casi uno a uno y presentar unas conclusiones válidas es enorme. Los autores pueden sentirse muy satisfechos, como nos sentimos los que hemos leído el estudio y consultado las palabras que en él se encierran.

Y termino. «Mi palabra es prenda de oro». Dice el *Diccionario de Autoridades* que esta es frase con que se pondera a alguno la seguridad que debe tener en la oferta que se le hace. Hago mía la expresión porque no otra ha sido la idea del prólogo que ahora acaba: ponderar la seguridad en la oferta que nos hacen los autores de este excelente material sobre la disponibilidad léxica en Madrid.

PRIMERA PARTE
LA INVESTIGACIÓN SOBRE EL LÉXICO DISPONIBLE
DE MADRID

CAPÍTULO 1

INTRODUCCIÓN. LA DISPONIBILIDAD LÉXICA

1.1. Los estudios de disponibilidad léxica

La disponibilidad léxica es una corriente de investigación que tiene como objetivo principal identificar el léxico disponible, entendiendo por tal aquel que, por estar almacenado en la mente de los hablantes y ser compartido por los miembros de la comunidad lingüística, está listo para ser usado cuando el tema de la conversación así lo requiera. Los trabajos de esta línea de investigación surgieron en Francia, pero han conocido su mayor auge en el seno de la lengua española. Desde que hace ya más de cincuenta años apareciera el primer trabajo sobre el léxico disponible de un territorio hispanohablante (López Morales 1973), las investigaciones, artículos, tesis doctorales y publicaciones sobre este asunto que se han realizado en España y en América se cuentan por centenas, y la lista continúa aumentando, ampliando los territorios estudiados, profundizando en las líneas de investigación precedentes y explorando nuevas vías interpretativas a partir de los materiales obtenidos. Como se ha afirmado (Paredes García 2012), el desarrollo alcanzado en español por las investigaciones sobre disponibilidad léxica no tiene equiparación en cantidad y calidad con respecto a ninguna otra lengua del mundo.

Los resultados de la disponibilidad léxica han sido especialmente relevantes en el ámbito de la enseñanza/aprendizaje, tanto en lengua materna como en segundas lenguas (Samper Padilla, Bellón Fernández y Samper Hernández 2003), en consonancia con el hecho de haber nacido en el seno de la lingüística aplicada a la didáctica de las lenguas. Los listados de unidades léxicas disponibles han servido para realizar estudios y propuestas de actuación didáctica de carácter general, para el tratamiento específico de ciertos aspectos léxicos (*cf.* López González 2014, Paredes García 2015), o incluso para analizar y abordar la situación ortográfica (Paredes García 1999, Galloso Camacho 2002: 209-224, Ávila Muñoz 2007, Sánchez-Sauz Laserna 2011, Blanco 2011, Nalesso 2022, Ortolano Ríos 2005, Santos Díaz, Trigo Ibáñez y Romero Oliva 2019,

Trigo Ibáñez, Romero Oliva y Santos Díaz 2018, 2019, 2020). En el ámbito de la enseñanza de segundas lenguas, están posibilitando acercamientos a asuntos como la interlengua, el análisis de errores, el funcionamiento de los procesos de transferencia e interferencia, los mecanismos de generación léxica en L2, etc. (Ayora Esteban 2004, Serrano Zapata 2006, Taberero Sala 2008, Gómez Devís 2010). Por otro lado, desde hace ya tiempo se vienen realizando análisis que contrastan el léxico disponible y el de los manuales de español como lengua extranjera, donde se ponen de manifiesto carencias notables que deben resolverse cuanto antes en aras de conseguir una adaptación a las necesidades reales de aprendizaje de este tipo de alumnado (Lin 2006, Frey Pereira 2006, Fernández-Merino 2011, Hidalgo Gallardo 2017). Las investigaciones, en definitiva, sirven de gran ayuda para disciplinas diversas, como la dialectología, la sociolingüística, la etnolingüística, la psicolingüística o la lingüística cognitiva (López Morales 1999: 19-26). La disponibilidad léxica sigue abriendo nuevos caminos y ámbitos de aplicación, como puede observarse en los trabajos editados por Ávila Muñoz y Villena Ponsoda (2010) o en publicaciones más recientes, como las contribuciones recogidas en los volúmenes editados por López Morales y Trigo Hernández (2019), Calero Fernández y Serrano Zapata (2021) o Gómez Devís y Gómez Molina (2022).

1.1.1. *Fundamentos teóricos y metodológicos*

La corriente de estudios nació en Francia a mediados del siglo XX con el objetivo de establecer el vocabulario que debía incluirse en los manuales para la enseñanza del francés para extranjeros. Surgieron cuando los investigadores franceses del proyecto para la elaboración del *Français Elementaire* (Gougenheim, Michéa, Rivenç y Sauva-geot 1956) se plantearon qué léxico deberían incorporar al currículo para la enseñanza del francés, sobre todo en los programas destinados a las antiguas colonias. Como la confección de cualquier currículo ha de partir de una selección, los investigadores decidieron adoptar criterios objetivos que justificasen sus elecciones léxicas, y recurrieron inicialmente a la frecuencia, partiendo de la premisa de que las palabras más habituales deben ser las que se enseñen en primer lugar. Pero pronto se hizo patente que ese criterio no resultaba suficientemente abarcador, ya que en los listados de frecuencia o bien no figuraban palabras esenciales para poder comunicarse –que además eran perfectamente conocidas por todos los hablantes–, o bien, cuando aparecían en esos listados, lo hacían en posiciones muy retrasadas, lejos del número de palabras que razonablemente es posible llevar al aula. La necesidad de adoptar nuevas perspectivas de análisis condujo a incorporar a las investigaciones la noción de «disponibilidad» de las palabras. El concepto se debe a Michéa (1953), quien lo define del siguiente modo:

Una palabra disponible es una palabra que, sin ser particularmente frecuente, está siempre lista para ser empleada y acude naturalmente a la mente en el momento en que se tiene necesidad de ella. Es una palabra que, formando parte de las aso-

ciaciones de ideas usuales, existe en potencia en el sujeto hablante en cuanto estas asociaciones entran en juego (Michéa 1953: 338).

Se sentaban así las bases de la línea de investigación sobre la disponibilidad léxica y a partir de estas premisas se establece el método para encontrar las unidades léxicas que los individuos conservan en su mente.

Los estudios de disponibilidad léxica son estudios lexicológicos que se basan en criterios estadísticos. Además de delimitar entre palabras frecuentes, palabras comunes y palabras usuales, Michéa (1953) estableció también la distinción entre palabras «temáticas» y «atemáticas»: las primeras, de un contenido semántico muy concreto, sólo aparecen cuando se trata de un determinado tema; las segundas, aparecen con regularidad independientemente del tema tratado.

Metodológicamente, los estudios de esta corriente de investigación emplean la técnica de la encuesta, en la que se pide a los sujetos que en un tiempo prefijado escriban (o digan) todas las palabras que conozcan acerca de un tema concreto —denominado *centro de interés* en la metodología—. De este modo, los informantes actualizan un conjunto de unidades léxicas, que pueden ser palabras simples, léxias complejas o construcciones sintagmáticas. El orden en que se actualizan estas unidades es importante, pues se entiende que aquellas que acudan antes a la mente del encuestado son también las que están más disponibles.¹ La disponibilidad léxica asume una doble perspectiva, individual y social, respecto al conocimiento léxico de los hablantes. Desde el punto de vista individual, las palabras más disponibles serán aquellas que con más celeridad acudan a la mente del sujeto al evocar un tema; desde el punto de vista social, las palabras más disponibles serán las que colectivamente más hablantes hayan sido capaces de actualizar. De este modo es posible determinar para cada unidad léxica su índice de disponibilidad, que vendrá dado por la combinación de su frecuencia de aparición (variable social) y de su posición relativa en cada listado (variable individual).

Los resultados de la disponibilidad léxica se presentan en forma de «diccionarios», que consisten en listados ordenados en los que cada unidad viene acompañada de datos estadísticos, como el índice de disponibilidad o la frecuencia. La piedra angular de la metodología tradicionalmente ha sido el *índice de disponibilidad* de cada vocablo, un valor numérico que conjuga el orden de aparición del vocablo y su frecuencia (cantidad de sujetos que lo actualizaron) y que permite por ello jerarquizar las unidades léxicas. El valor se obtiene al agrupar las respuestas de todos los individuos y resulta de la combinación entre el número de veces que aparece un término y la posición que ocupa en el proceso de cada actualización. Se trata, por tanto, de un valor que predice un comportamiento social, no individual; de ahí que resulte de suma utilidad cuando

¹ En Ávila Muñoz y Sánchez Sáez (2010: 55-58) se revisa la fundamentación matemática tradicional, basada en el orden de aparición de los términos, y se propone manejar el criterio de accesibilidad a la red léxica, que implica procesos de reentrada hacia el centro de la red, donde se ubican los ejemplares léxicos prototípicos.

se pretende determinar la norma de la comunidad objeto de estudio, así como para realizar estudios contrastivos entre variedades sociales o geográficas. Numéricamente se expresa con un valor entre 1 y 0, que refleja el grado de accesibilidad de un vocablo en el lexicón teniendo en cuenta que 1 equivaldría a la disponibilidad máxima y 0 a la nula disponibilidad (obviamente, este último valor se obtiene por defecto, pues corresponde a todos los vocablos que estén ausentes en el listado).

Pero este dato, combinado con otros valores, puede servir también para hacer inferencias sobre características de los individuos como tales. Así, López Chávez y Strassburger (1991) toman en cuenta el índice de disponibilidad léxica individual (IDLI), con el que es posible medir la participación de cada individuo en el resultado global de la muestra. Su hipótesis es que si un sujeto actualiza los términos de mayor índice de disponibilidad léxica probablemente tendrá más disponibilidad léxica que otro. De signo contrario es el planteamiento de Ávila Muñoz y Sánchez Sáez (2010: 66-69), quienes consideran que el hecho de que los sujetos compartan comunidad lingüística hace que compartan también el léxico y, en especial, los elementos centrales de los centros de interés. Desde este punto de vista, la actualización de las unidades de bajo índice de disponibilidad por parte de un sujeto supone una mayor capacidad léxica, desde el supuesto de que la facilidad de acceso a los vocablos menos disponibles presupone el acceso a los más disponibles, incluso en el caso de que no se actualicen. En su estudio manejan un nuevo índice, que denominan Índice de Descentralización Léxica, capaz de determinar la capacidad léxica individual, para cuyo cálculo se toman en cuenta las palabras que un hablante no comparte con los demás en los centros de interés.

1.1.2. *El Proyecto Panhispánico de Disponibilidad Léxica (PPHDL)*

En la última década del siglo xx se puso en marcha el Proyecto Panhispánico de Disponibilidad Léxica (PPHDL), al cual, bajo la dirección de Humberto López Morales, se incorporaron numerosos equipos de investigación de España y de América. Las reuniones de coordinación celebradas en Bilbao, en San Millán de la Cogolla y en Comillas sirvieron para establecer las bases teóricas comunes en las que se moverían las investigaciones, fijar las líneas metodológicas compartidas por el macroproyecto y proponer el objetivo último, que consistiría en crear un diccionario del léxico disponible de la lengua española, que había de resultar como fruto conjunto de la intersección de las investigaciones sobre el léxico disponible de cada uno de los territorios estudiados. Se resumen a continuación las principales bases metodológicas del proyecto.

- Asumiendo la propuesta de los investigadores franceses, la encuesta se compone de 16 áreas temáticas –o *centros de interés* según la denominación fijada por Gougenheim, Michéa, Rives y Sauvageot (1964)–, mediante los cuales,

al menos teóricamente, se cubren en el plano léxico los intereses inmediatos de cualquier individuo, independientemente de su condición o situación, y que además tienen el grado de coherencia necesario para servir de estímulo asociativo para los encuestados. Son estos:²

- CI 01. Partes del cuerpo
- CI 02. La ropa
- CI 03. Partes de la casa (sin los muebles)
- CI 04. Los muebles de la casa
- CI 05. Alimentos y bebidas
- CI 06. Objetos colocados en la mesa para la comida
- CI 07. La cocina y sus utensilios
- CI 08. La escuela: muebles y materiales
- CI 09. Iluminación, calefacción y medios de airear un recinto
- CI 10. La ciudad
- CI 11. El campo
- CI 12. Medios de transporte
- CI 13. Trabajos del campo y del jardín
- CI 14. Los animales
- CI 15. Juegos y distracciones
- CI 16. Profesiones y oficios

- La unidad social investigada constituye una comunidad de habla (Martínez Olmos 2007), entendida en un sentido lato como el conjunto de hablantes que comparten los mismos usos lingüísticos y mantienen similares actitudes hacia ellos. En la práctica, en España se ha utilizado la provincia como unidad geográfica de estudio, mientras que en América en unos casos ha sido la ciudad y en otros la región.

Las encuestas se aplican a estudiantes del último curso de bachillerato o recién ingresados en la universidad, esto es, individuos que han completado su formación lingüística general, sin haberse expuesto aún al contacto con el léxico de la especialidad (López Morales 1999: 28).³

² Nada impide, sin embargo, que los investigadores puedan añadir a estos centros de interés comunes aquellos que consideren adecuados para ajustarse a las circunstancias específicas de cada investigación. Puede verse al respecto Samper, Bellón y Samper (2003), Samper y Samper (2006) o Paredes (2012).

³ La aplicación de las pruebas a sujetos con características sociales diferentes, iniciada ya por los propios investigadores franceses, ha tenido continuidad en los estudios hispanos, y se ha extendido a escolares de distinto nivel, desde preescolar (Sánchez Corrales y Murillo Rojas 2006) a primaria y secundaria (Prado, Galloso y Vázquez 2006, Prado y Galloso 2008, Samper Hernández 2009, Jiménez Berrio 2019, Paredes García (dir.) *et al.* 2022). Tampoco han sido ajenos los estudios a otros segmentos de la población, y así, por ejemplo, Ávila y Villena (2010) y Medina Peñate (2018) utilizan una muestra sociológicamente representativa de la ciudad de Málaga a partir de la estratificación propuesta en el proyecto PRESEEA, mientras que Borrego Nieto (2008) se sirve de jubilados.

- Se tiene en cuenta una serie de factores sociológicos de los sujetos: sexo, carácter público o privado del centro en que estudian, tipo de localidad de residencia y nivel sociocultural.
- Se utilizan criterios similares en cuanto a la «edición» de los datos, nombre que recibe el proceso de preparación de los datos para su tratamiento estadístico y que posibilita la posterior comparación de los resultados de las distintas investigaciones (Samper Padilla 1998).
- El PPHDL también ha marcado directrices en relación con el tamaño de la muestra, que abogan por la representatividad de la muestra seleccionada en relación con el universo al que representan. Ciertamente es que sería deseable que en las publicaciones se señalase siempre la representatividad de la muestra; no obstante, el número de informantes que se suele usar en los estudios de DL supera con mucho los criterios de representatividad sociolingüística, fijados en 0,025 % del universo estudiado (Labov 1966). Por debajo de ese nivel de representatividad deberían situarse solo los estudios exploratorios.
- Los estudios de disponibilidad léxica adoptan la estadística como herramienta básica en la que fundamentar los análisis cuantitativos y las interpretaciones de los resultados obtenidos. El desarrollo de herramientas y aplicaciones en este punto ha sido tal que se puede decir que la lingüística estadística aplicada al estudio del léxico se ha constituido como disciplina (Ávila Muñoz 2010: 15).

Los trabajos de recogida de datos están prácticamente concluidos en todas las provincias españolas y publicados los resultados de las encuestas. Para una visión general de este proyecto pueden consultarse los trabajos de López Morales (1995), Carcedo González (1998), Samper Padilla, Bellón Fernández y Samper Hernández (2003), Samper Padilla y Samper Hernández (2006) y Gallego Gallego (2014). Actualmente se trabaja coordinadamente para establecer el *Léxico Disponible de España*, que se realiza a partir de una muestra ponderada de las encuestas realizadas en las distintas investigaciones españolas (Bartol Hernández y Hernández Muñoz 2006, Samper Padilla 2008, Bartol Hernández y Borrego Nieto 2008).⁴

Al margen del macroyecto panhispánico, pero estrechamente vinculadas con él, han surgido otras propuestas de estudio. Algunas investigaciones actuales están manejando criterios más abarcadores sociológicamente, tratando de obtener el léxico disponible a partir de muestras representativas de la comunidad de habla. Así han hecho Ávila Muñoz y Villena Ponsoda (2010) y Medina Peñate (2018). Siguiendo los criterios de estratificación propuestos en el proyecto PRESEEA (Moreno Fernández 1996), Ávila Muñoz y Villena Ponsoda han recogido el léxico disponible la ciudad

⁴ El equipo de Madrid ha colaborado con este proyecto del léxico general del español aportando 400 encuestas de los madrileños, todas ellas extraídas de las 600 que conforman la muestra de la investigación que aquí se presenta.

de Málaga a partir de una muestra de 74 sujetos, y Medina Peñate ha hecho lo propio en Las Palmas de Gran Canaria y Madrid con muestras de 54 y 108 informantes respectivamente, preestratificados según el sexo, la edad y el nivel de instrucción. En el ámbito de la enseñanza superior, se ha estudiado la relación entre disponibilidad léxica y fines específicos (Gómez Sánchez y Guerra Salas 2004; Guerra Salas y Gómez Sánchez 2004, 2005; Guerra Salas 2014), con el propósito de conocer el vocabulario especializado que deben manejar los estudiantes universitarios extranjeros que cursan sus estudios (en español) en las universidades españolas. Por otra parte, los estudios de disponibilidad léxica se han aplicado también a población extranjera que aprende español como segunda lengua (Carcedo González 2000, Samper Hernández 2002, Gallego Gallego 2014, Sánchez-Saus Laserna 2011 y 2016, Hidalgo Gallardo 2017; *cf.* también Paredes García 2012 y 2015).

1.2. El proyecto de estudio del léxico disponible de Madrid (LDM)

En el año 2004, siguiendo las pautas del Proyecto Panhispánico de Disponibilidad Léxica, los investigadores autores de esta publicación presentaron a la Consejería de Educación de la Comunidad de Madrid un proyecto de investigación sobre el «Léxico disponible de la Comunidad de Madrid», que venía a unirse a los 46 que estaban ya en marcha entonces en diferentes zonas españolas y americanas (y de los cuales ya se habían concluido y publicado 18 en aquel momento). Aunque el proyecto no resultó finalmente financiado en aquella convocatoria, podemos considerar ese año como el de inicio de los trabajos del Léxico Disponible de Madrid que el lector tiene ahora en sus manos.

Los promotores del proyecto en la Comunidad de Madrid asistimos a las reuniones internacionales de Bilbao (patrocinada por la Universidad del País Vasco) y San Millán (auspiciada por la Fundación San Millán de la Cogolla), y tuvimos ocasión de intercambiar información, presentar problemas y sugerencias, probar los programas informáticos creados *ad hoc*, etc. con los principales investigadores de las dos orillas del Atlántico. Con cierto retraso con respecto a otras zonas geográficas, se ponían así en marcha las tareas preparatorias del Léxico Disponible de Madrid, cuyos aspectos metodológicos se detallan en el capítulo siguiente.

La investigación que ahora ofrecemos muestra la versión final del LDM. Aunque los diccionarios ya estaban a disposición de la comunidad académica por diferentes vías –y han servido de base a distintos estudios y aplicaciones didácticas, publicadas a lo largo de los últimos años–, consideramos, no obstante, que era necesario presentar los diccionarios y los trabajos basados en ellos en una única obra, que permitiera apreciar las múltiples aplicaciones que se derivan de las listas de disponibilidad. El presente volumen ofrece, pues, junto al LDM, un resumen actualizado de las diferentes investigaciones sustentadas en él, recogidas en el capítulo 4 de esta primera parte.

El trabajo se divide en dos partes: la primera se dedica a presentar la investigación madrileña y la segunda contiene el corpus de datos obtenido en las encuestas.

La presentación de la investigación sobre el LDM se inicia con una breve introducción sobre las bases teóricas y metodológicas de los estudios de disponibilidad léxica. A continuación, se ofrece la metodología seguida en la investigación, presentando los criterios de selección de la muestra y de edición de los materiales obtenidos, así como el tratamiento estadístico a que fueron sometidos los materiales. Esta primera parte se completa con los resultados cuantitativos y cualitativos de los datos. La segunda parte de la obra contiene los diccionarios del léxico disponible de los estudiantes preuniversitarios madrileños. Los repertorios léxicos se ofrecen en dos versiones: en primer lugar, se presentan los vocablos de cada centro de interés ordenados por su grado de disponibilidad; a continuación, se presentan los vocablos ordenados alfabéticamente.

Es de justicia mencionar aquí el respaldo recibido, y la confianza otorgada, desde el inicio hasta la conclusión del proyecto, por el profesor Humberto López Morales, por entonces secretario general de la Asociación de Academias de la Lengua Española y, como se ha señalado en la sección precedente, coordinador del Proyecto Panhispánico y auténtico dinamizador de las investigaciones sobre disponibilidad léxica en el mundo hispanohablante. Asimismo, queremos dejar constancia de nuestro agradecimiento por la ayuda recibida por parte de numerosos investigadores. Vaya pues nuestra gratitud para el profesor Guillermo Rojo y su equipo de colaboradores de la Universidad de Santiago de Compostela, que nos prestaron generosa ayuda en la fase de preparación de los materiales para su tratamiento estadístico. Vaya también para el equipo de sociolingüistas de la Universidad de Málaga, entre los que hay que citar especialmente a Antonio M. Ávila y Juan A. Villena, así como al profesor José María Sánchez, quienes nos facilitaron análisis estadísticos que han posibilitado diversos aspectos de esta investigación. Y, en este apartado de agradecimientos, hay que mencionar especialmente a los equipos directivos de los centros de enseñanza, por las facilidades que nos dieron para acceder y poder aplicar las encuestas, a los profesores, que nos prestaron sus horas de clase para poder llevar a cabo nuestra investigación, y, por supuesto, a todos los alumnos que colaboraron con esta investigación: sin su generosidad esta investigación no hubiera sido posible.⁵

⁵ La redacción final de este trabajo se ha visto muy enriquecida por la evaluación de los dos revisores anónimos del libro, los cuales han ayudado con su atenta lectura y sus atinados comentarios y observaciones a mejorar el resultado final. Por supuesto, cualquier defecto, descuido o error que pueda encontrarse en la obra es responsabilidad exclusiva de los autores.

CAPÍTULO 2

ASPECTOS METODOLÓGICOS

2.1. El territorio

La Comunidad de Madrid es una de las seis comunidades autónomas uniprovinciales del estado español. Situada en el centro geográfico de la península ibérica, limita con las provincias castellano-manchegas de Toledo, Cuenca y Guadalajara, y las castellano-leonesas de Segovia y Ávila. Ocupa una superficie de 8021,80 km² y tiene una población de 6 751 251 habitantes, según los datos del INE a 1 de enero de 2022. Estas cifras convierten a Madrid en la provincia de mayor número de habitantes y a la Comunidad de Madrid en la tercera más poblada de las comunidades autónomas (tras Andalucía y Cataluña) y en la de mayor densidad de población (844 habitantes por km²). No obstante, el reparto de la población es desigual: la altísima densidad alcanzada por la capital y el cinturón metropolitano contrasta con la despoblación que padecen algunas zonas montañosas del norte y oeste de la provincia.

Territorialmente, la Comunidad de Madrid se organiza en 179 municipios, de los cuales el de mayor superficie es la propia capital, que ha ido progresivamente anexionando municipios aledaños, como ha ocurrido con Fuencarral, Barajas, El Pardo, Hortaleza, Vallecas o Carabanchel, entre otros, convertidos actualmente en distritos de la capital.

La mayor parte de la población se concentra en la capital y en las ciudades del cinturón de su área metropolitana. La evolución demográfica ha venido marcada por una tasa de crecimiento casi constante desde que la ciudad de Madrid se convirtiese en la capital del reino a mediados del siglo xvi. En los últimos cincuenta años ha habido dos grandes explosiones demográficas en el territorio, una en la segunda mitad del siglo xx y otra en los primeros años del siglo xxi. Ambas supusieron un notable incremento de la población no solo para la ciudad de Madrid, como venía ocurriendo históricamente, sino también para las ciudades del entorno.

Si en 1950 la comunidad no llegaba a los dos millones de habitantes, en 1981 el censo daba ya una cifra de 4.726.986 habitantes. Esta primera migración fue sobre todo interna: los inmigrantes llegados a la Comunidad de Madrid procedían de las regiones limítrofes o de la periferia del Estado, especialmente de Castilla-La Mancha, Extremadura y Andalucía. La segunda explosión demográfica (primeros años del siglo XXI), supuso que en 2007 la población superase los seis millones de habitantes. A diferencia de la anterior, los causantes de este ascenso demográfico del territorio madrileño en esta segunda explosión demográfica fueron los inmigrantes de procedencia internacional, que en conjunto sumaban en esa fecha más de un millón de personas y suponía el 18,83% de la población total. Por nacionalidades, destacaban los rumanos y los ecuatorianos, seguidos de marroquíes, colombianos y peruanos.

2.2. La población estudiada

La población diana de este estudio son los estudiantes preuniversitarios madrileños, que de acuerdo con el sistema educativo español pueden cursar sus estudios en centros de enseñanza pública o en colegios privados, ya sean estos de titularidad estrictamente privada, ya sea en centros que a través de los conciertos con el Estado reciben fondos públicos para su financiación.¹

El número de estudiantes que cursaban segundo curso de Bachillerato en la Comunidad de Madrid durante el curso 2001-2002 ascendía a 43 452, de los cuales la mitad estaba escolarizada en centros educativos de la capital y el resto se distribuía en las cuatro áreas territoriales restantes. En centros de enseñanza pública estaban inscritos 26 360 alumnos, que constituyen el 60,7 % del total; los alumnos que cursaban estudios en centros privados o concertados eran 17 092, el 39,3 % restante.

A efectos de organización administrativa, la Consejería ha ordenado el territorio en cinco Direcciones de Área: Madrid Centro, que se circunscribe al municipio capitalino, Madrid Norte, Madrid Sur, Madrid Este y Madrid Oeste, como refleja el mapa 2.1.

¹ Aunque hay diferencias importantes entre los centros estrictamente privados y los concertados, desde el punto de vista sociológico la mayoría de los estudiantes que acuden a ellos proceden de un estrato sociológico similar, por lo que en esta investigación han sido considerados como un único grupo.

MAPA 2.1. DIVISIÓN ADMINISTRATIVA DE LA COMUNIDAD DE MADRID



La proporción entre alumnos de enseñanza pública y privada es diferente según la zona: mientras que en las áreas periféricas la proporción es favorable a la pública, en la capital el número de estudiantes que cursa sus estudios en centros privados o concertados supera al de quienes estudian en centros públicos. Los alumnos por tipo de enseñanza y por áreas territoriales se reparten del siguiente modo:

TABLA 2.1. DISTRIBUCIÓN DE LOS ESTUDIANTES MADRILEÑOS PREUNIVERSITARIOS

	M.Centro	M.Norte	M.Este	M.Sur	M.Oeste	CM
Pública	9 531	2 091	4 447	7 749	2 497	26 360
Privada	12 303	1 008	227	1 749	1 805	17 092
Total	21 834	3 099	4 674	9 543	4 302	43 452

2.3. La muestra

2.3.1. Criterios de selección

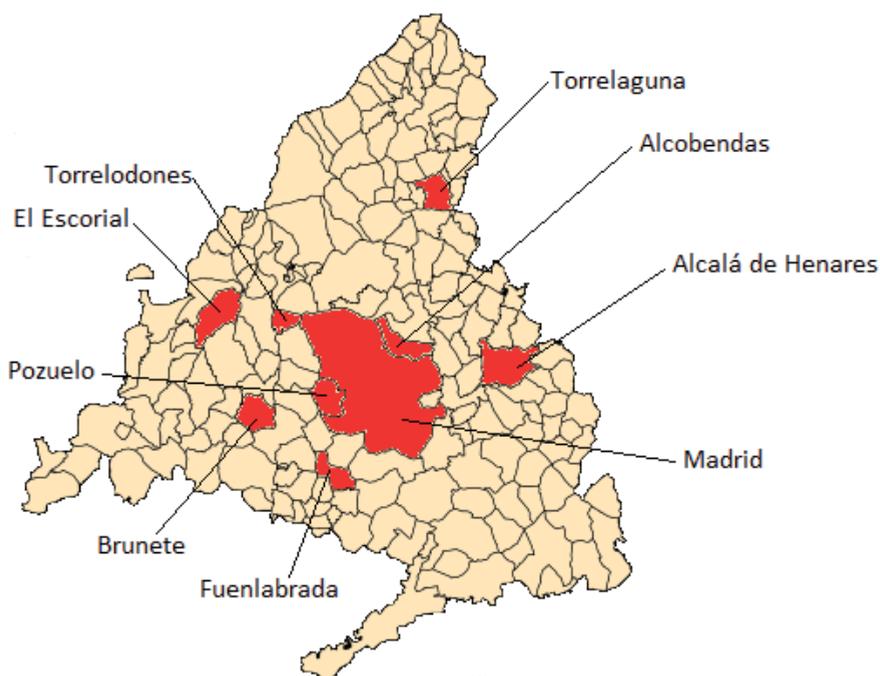
La muestra se compone de 600 encuestas, que representan el 1,38% del total del alumnado madrileño del segundo curso de Bachillerato. El elevado número de encuestas, que se corresponde con el alto número de estudiantes de la Comunidad de Madrid, convierte a la muestra madrileña en la más nutrida de cuantas participan en los proyectos de léxico disponible en España.²

² Para el diseño de la muestra en los estudios de disponibilidad léxica, véase Samper Padilla 2007.

Para conformar la muestra se han seguido criterios de preestratificación y, una vez hechas las encuestas, se agruparon los alumnos atendiendo a otros criterios de posestratificación. En la fase de preestratificación, se ha buscado que la muestra sea representativa territorialmente de la distribución en áreas de la Consejería de Educación y del carácter público o privado de la enseñanza recibida, de manera que se han seleccionado proporcionalmente los alumnos según su adscripción al área correspondiente. En la fase de posestratificación se ha atendido a criterios sociológicos, como el sexo del alumno o el nivel sociocultural de los alumnos, y a criterios demográficos, como el tipo de localidad en la que se ubica el centro de estudios según el número de habitantes.

En la distribución territorial, a cada alumno se le asignó el área en que se ubica el centro de estudios. Las localidades en las que cursan estudios los alumnos encuestados son nueve, seleccionadas de forma que estuviesen representadas todas las áreas territoriales y de modo que la distribución geográfica permitiese un conocimiento del léxico de todo el territorio madrileño (Mapa 2.1.). Las localidades son las siguientes: el municipio de Madrid para el área territorial Madrid-Centro, Alcalá de Henares en el área territorial Madrid-Este, Alcobendas y Torrelaguna del área Madrid-Norte, El Escorial, Pozuelo del Rey, Brunete y Torrelodones en el área Madrid-Oeste, y Fuenlabrada en el área territorial Madrid-Sur.

MAPA 2.2. LOCALIDADES EN LAS QUE SE HAN REALIZADO LAS ENCUESTAS



Los centros participantes fueron 23, que se reparten equitativamente entre once públicos y doce privados o concertados. Por áreas, ocho son de Madrid capital, tres son de la zona Norte, dos de la zona Este, cinco de la zona Sur y otros cinco de la zona Oeste. A continuación se relacionan los institutos y colegios en los que se aplicaron las encuestas.³

CUADRO 2.1. CENTROS EDUCATIVOS EN LOS QUE SE REALIZARON LAS ENCUESTAS

Área	Carácter	Centro	Localidad
Madrid Centro	Público	I.E.S. Mariana Pineda	Madrid
		I.E.S. Pío Baroja	Madrid
		I.E.S. Calderón de la Barca	Madrid
	Privado	Col. Nuestra Señora del Pilar	Madrid
		Col. Las Rosas	Madrid
		Col. Chamberí	Madrid
		Col. S. José del Parque	Madrid
		Col. N.ª S.ª de la Consolación	Madrid
Madrid Norte	Público	I.E.S. Torrente Ballester	Alcobendas
		I.E.S. Alto Jarama	Torrelaguna
	Privado	Col. Base	Alcobendas
Madrid Este	Público	I.E.S. Antonio Machado	Alcalá de Henares
	Privado	Col. Escolapias	Alcalá de Henares
Madrid Sur	Público	I.E.S. Sefarad	Fuenlabrada
		I.E.S. Salvador Allende	Fuenlabrada
		I.E.S. Ítaca	Fuenlabrada
	Privado	Col. Cossío	Fuenlabrada
		Col. NILE	Fuenlabrada
Madrid Oeste	Público	I.E.S. Diego Velázquez	Torrelodones
		I.E.S. Gerardo Diego	Pozuelo
	Privado	Col. S. José de Cluny	Pozuelo de Alarcón
		Col. Arcadia	Brunete
		Col. El Escorial	El Escorial

Respecto al tipo de enseñanza recibida, los alumnos se reparten entre aquellos que han cursado sus estudios en centros de titularidad pública y los que los han realizado en centros privados o de enseñanza concertada.⁴ Se ha tenido en cuenta la posibilidad

³ Reiteramos aquí nuestro agradecimiento a los equipos directivos de los centros educativos, a los profesores que nos facilitaron el acceso a las aulas y a todos los alumnos participantes en las encuestas.

⁴ Siguiendo la metodología del PPHDL y en aras a mantener la homogeneidad con el macroproyecto, para la clasificación de la muestra se han considerado solo dos tipos de centros de enseñanza, los institutos de

de que el alumno hubiese variado de tipo de centro a lo largo de su etapa de formación. Por ello, para clasificar a cada alumno en el grupo de la enseñanza pública o la privada se ha tenido en cuenta no solo el tipo de centro en el que se encontraba el alumno en el momento de realizar las encuestas, sino también su trayectoria académica, de manera que, por ejemplo, si un alumno que en el momento de la encuesta se encontraba en un centro público había cursado todos o la mayor parte de los cursos en un centro de titularidad privada, se codificaba como alumno de este último tipo de centro.

En cuanto a los criterios de posestratificación, se ha atendido al sexo⁵ y al nivel sociocultural de los encuestados. Para la distribución según el género o sexo se ha dividido a los alumnos en los dos grupos poblacionales habituales en los trabajos de corte sociolingüístico, el masculino y el femenino.

Para establecer el nivel sociocultural de cada sujeto, se han tenido en cuenta diversos factores. Por un lado, la profesión y el nivel de estudios del padre y de la madre, por otro, el nivel de ingresos de la familia; en tercer lugar, el grado de implicación personal con su propio desarrollo cultural. A este efecto, se les preguntaba a los alumnos acerca del número de libros que leían habitualmente al año, excluyendo los de lectura obligatoria en la escuela, y el número de veces que solían ir al cine o al teatro. Estos datos han sido ponderados otorgando diferente puntuación a cada uno de estos aspectos. Se asignaban entre 0 y 3 puntos en función de la profesión⁶ y los estudios del padre, y otros tantos según los de la madre; los ingresos familiares anuales se evaluaban en una tabla entre 1 y 4 puntos.⁷ En cuanto a las características personales, en relación con la lectura se le otorgaban 2 puntos si el sujeto leía entre 0 y 4 libros al año, 4 puntos si leía entre 5 y 9, y 6 puntos si sus lecturas superaban los 10 libros anuales; en cuanto a la asistencia al cine y al teatro, no obtenía puntuación si la respuesta era negativa y en caso de ser positiva, se le adjudicaban 2 puntos si la frecuencia era entre 0 y 5 veces al año, 3 puntos si era entre 6 y 10 veces al año y 4 puntos si superaba ese número de veces.

Computados de este modo los datos para cada individuo, la puntuación individual podía oscilar entre los 3 y los 20 puntos. Los sujetos quedaron distribuidos en tres

titularidad pública y los colegios de titularidad privada. No obstante, dentro de este segundo grupo, que incluye centros privados –que no reciben financiación pública– y centros concertados –que sí la reciben–, puede haber diferencias sociológicas notables. No se estudian en este trabajo, pero pueden ser objeto de análisis en investigaciones que desarrollen la que ahora se presenta.

⁵ En este trabajo se usará el término *sexo* para aludir a este factor macrosocial, aunque los autores somos conscientes del debate abierto en torno a la nomenclatura. La opción léxica se ha tomado en consonancia con la mayor parte de los estudios de disponibilidad léxica, pero en ningún caso supone una toma de postura ideológica sobre si conviene más el término biológico *sexo* o el cultural *género*.

⁶ Para catalogar las profesiones se siguieron los criterios de López Morales en su investigación sobre San Juan de Puerto Rico (*vid.* López Morales 1989: 129-130).

⁷ La escala de ingresos que se usó fue esta: 1. de 0 a 13 800 € (para el periodo previo al empleo del euro como moneda oficial, de 0 a 2 300 000 pesetas); 2. de 13 800 a 25 800 € (de 2 300 001 a 4 300 000 pesetas); 3. de 25 800 a 45 000 € (de 4 300 001 a 7 500 000 pesetas); 4. Más de 45 000 € (más de 7 500 000 pesetas).

niveles socioculturales: bajo para quienes obtuvieron puntuaciones entre 3 y 8 puntos; medio si el puntaje se situaba entre 9 y 14 puntos; y alto si la puntuación estaba entre 15 y 20 puntos.

Respecto al carácter de la localidad en la que se ubica el centro educativo, se han distinguido tres tipos teniendo en cuenta la distribución de la población que presenta la Comunidad de Madrid. Se han considerado localidades urbanas aquellas que superan los 100 000 habitantes; localidades semiurbanas, las que tienen entre 10 000 y 100 000 habitantes y las que no llegaban a esta cifra mínima se consideraron localidades rurales.

2.3.2. Configuración de la muestra

De acuerdo con los criterios de preestratificación expuestos, la muestra quedó definitivamente configurada de la manera que se recoge en la Tabla 2.2.

TABLA 2.2. DISTRIBUCIÓN DE LA MUESTRA POR FACTORES SOCIALES

		N	%
Sexo	Hombre	254	42,3
	Mujer	346	57,7
Tipo de centro	Público	340	56,7
	Privado	260	43,3
Tipo de localidad	Urbana	433	72,2
	Semiurbana	149	24,8
	Rural	18	3,0
Nivel sociocultural	Bajo	125	20,8
	Medio	343	57,2
	Alto	132	22,0
Área territorial	Centro	302	50,3
	Norte	43	7,2
	Este	66	11,0
	Sur	130	21,7
	Oeste	59	9,8
Total		600	100,0

En las pruebas de disponibilidad de Madrid han participado 254 alumnos y 346 alumnas. La proporción está en consonancia con el predominio de mujeres entre los estudiantes de este nivel educativo. Por el tipo de enseñanza que ha recibido el encuestado, las encuestas se reparten proporcionalmente al número de alumnos matriculados en la opción correspondiente. El reparto de la muestra según el tipo

de localidad es el siguiente: 433 alumnos pertenecen a una localidad urbana, 149 a poblaciones semiurbanas y 18 a localidades rurales. En cuanto al nivel sociocultural, los tres grupos en los que se distribuye la población presentan el reparto siguiente desequilibrado a favor del grupo medio, pero ajustado a la proporción social que este grupo representa en la sociedad general. Por último, la distribución atendiendo a la zona de escolarización de los alumnos ha respetado la proporción con respecto al total de alumnos matriculados en cada área. De ahí que las encuestas hayan sido más numerosas en la capital, donde se acumula más de la mitad de la muestra, que en el resto de áreas.

Hay que tener en cuenta, además, que la proporción de centros públicos y privados en la Comunidad de Madrid es irregular. La zona centro de Madrid tiene más cantidad de centros privados, mientras que en el resto de zonas geográficas madrileñas la proporción se invierte. De esta manera, la distribución de los 600 alumnos de la muestra según estos dos criterios es la que queda reflejada en la Tabla 2.3.

TABLA 2.3. DISTRIBUCIÓN DE LA MUESTRA SEGÚN EL ÁREA TERRITORIAL Y EL TIPO DE CENTRO

	M.Centro	M.Norte	M.Este	M.Sur	M.Oeste	Total CM
Pública	131	29	61	107	34	362
Privada	171	14	4	24	25	238
Muestra	302	43	65	131	59	600

2.4. La obtención del léxico disponible: la encuesta

Se presentarán a continuación los aspectos metodológicos básicos relativos al modo en que quedó configurada la encuesta para obtener el léxico disponible y la manera en que fue aplicada a los estudiantes madrileños.

2.4.1. El diseño de la encuesta

En la investigación sobre disponibilidad léxica de los estudiantes madrileños se planteó recabar el vocabulario conocido por los sujetos en relación a 20 centros de interés o áreas temáticas. Estos centros fueron enunciados de la manera siguiente:

- CI. 01. El cuerpo humano
- CI. 02. La ropa
- CI. 03. Partes de la casa
- CI. 04. Los muebles de la casa
- CI. 05. Comidas y bebidas
- CI. 06. Objetos colocados en la mesa para la comida
- CI. 07. La cocina y sus utensilios
- CI. 08. La escuela: muebles y materiales
- CI. 09. Calefacción, iluminación y ventilación
- CI. 10. La ciudad
- CI. 11. El campo
- CI. 12. Medios de transporte
- CI. 13. Trabajos de campo y de jardín
- CI. 14. Los animales
- CI. 15. Juegos y distracciones
- CI. 16. Profesiones y oficios
- CI. 17. Los colores
- CI. 18. Sensaciones y sentimientos
- CI. 19. Adjetivos
- CI. 20. Medios de comunicación

Como se puede apreciar, en el diseño se han mantenido las propuestas del Proyecto Panhispánico de Disponibilidad Léxica y se han respetado tanto el número de dieciséis centros de interés comunes, así como el orden en que estos se presentan a los alumnos. Por otra parte, se ha mantenido el *CI 17. Los colores*, que también ha sido empleado con asiduidad en muchas de las investigaciones vinculadas al PPHDL desde que se incluyese en el trabajo pionero de Justo Hernández (1986).

Los tres últimos son específicos de la investigación, aunque algunos de ellos han sido usados en otras investigaciones de una manera idéntica o similar. El *CI 18. Sensaciones y sentimientos* busca registrar léxico abstracto, menos proclive a aparecer en los estudios tradicionales; el *CI 19. Adjetivos*, pretende registrar esta categoría gramatical, también menos frecuente en los estudios, y que además plantea como novedad metodológica en el enunciado no estar asociado a ningún tema específico, por lo que en las listas se registrarán palabras atemáticas. El *CI 20. Medios de comunicación* se añadió por intereses específicos de los investigadores en torno a esta área semántica.

Por otro lado, se elaboró un cuadernillo compuesto por dos partes, la primera destinada a recoger los datos sociológicos de los encuestados, y la segunda para recoger el léxico (Véase Anexo 1). La información sociológica recogida tenía que ver con el sexo/género del sujeto, la edad, el lugar de nacimiento, información sobre la historia

escolar (tipo de enseñanza pública o privada), información sobre el origen, profesión y estudios del padre y de la madre e información sobre los ingresos. También se les preguntaba sobre dos aspectos culturales individuales, el número de libros leídos anualmente y la frecuencia de la asistencia al cine o teatro.

El apartado dedicado a registrar el léxico conocido por los sujetos se componía de varias hojas, cada una de ellas con tres columnas para tres centros de interés. Estas columnas aparecen sin el enunciado del centro de interés –solo el número–, para evitar que los encuestados puedan pensar en asociaciones antes de que se enuncie el tema correspondiente.

2.4.2. *La aplicación de la encuesta*

Las encuestas fueron aplicadas por los propios investigadores, que elaboraron un protocolo con objeto de garantizar la homogeneidad de las pruebas (Véase Anexo 1). Este protocolo básicamente señalaba los siguientes puntos:

- Para cada centro de interés, los alumnos disponían de dos minutos para escribir tantas palabras o expresiones como supiesen pertenecientes al tema. El tiempo fue cronometrado por el responsable de la aplicación de la prueba y, al concluir, se les pedía a los alumnos que dejaran de escribir.
- Para cada centro de interés los alumnos disponían de una columna formada por 45 líneas en blanco y numeradas, pero se les indicaba que podían añadir más en los márgenes o por detrás de la hoja en caso de que fuese necesario. También se les indicaba que escribiesen una palabra o expresión en cada línea, sin dejar espacios en blanco.
- A los alumnos se les recomendaba que escribiesen correctamente las palabras, pero se insistía en que no rechazasen ninguna palabra por dudar sobre su ortografía o por cualquier otra razón (tabú, etc.).
- Cada centro de interés se presentaba con una breve frase, consensuada por los investigadores,⁸ sin dar ningún ejemplo para no condicionar las respuestas y, con ello, invalidar los resultados. A los estudiantes se les pedía que ellos tampoco diesen ejemplos ni hiciesen comentarios o preguntas en voz alta.

Para la aplicación de las encuestas, los investigadores se distribuyeron el territorio objeto de estudio. Florentino Paredes se encargó de las encuestas correspondientes a las áreas territoriales Madrid-Norte, Madrid-Este y buena parte de Madrid-Centro; Luis Guerra y M.^a Elena Gómez, por su parte, se ocuparon de recoger las encuestas de los centros situados en las áreas Madrid-Oeste, Madrid-Sur y parte de Madrid-Centro.

⁸ El protocolo de aplicación de la encuesta se recoge en el Anexo 1.