

INFORMACIÓN BÁSICA SOBRE EL
«PROYECTO PARA EL ESTUDIO SOCIOLINGÜÍSTICO DEL
ESPAÑOL DE ESPAÑA Y DE AMÉRICA» - PRESEEA (1996 - 2010)

FRANCISCO MORENO FERNÁNDEZ
Univ. de Alcalá de Henares - Coordinador General de PRESEEA

En octubre de 2005, en el seno del «XIV Congreso Internacional de la Asociación de lingüística y Filología de la América Latina» (ALFAL), se celebró una reunión de los equipos que integran el proyecto internacional conocido como PRESEEA. Se cumplían 10 años de la presentación pública del proyecto, que se produjo precisamente en el Congreso de esa misma asociación celebrado en Las Palmas de Gran Canaria (1996). El documento en que se sentaban las primeras bases de la investigación sociolingüística apareció publicado en las páginas de la revista *Lingüística* (1996, pp. 257-287). Actualmente, toda la información relativa al proyecto, incluida su metodología y el estado de las investigaciones, puede consultarse en la página electrónica <www.linguas.net/preseea>.

La intención de PRESEEA es reunir un gran corpus oral, técnicamente adecuado y sociolingüísticamente representativo de una amplia muestra de ciudades de todo el mundo hispánico. La intención es que este corpus de lengua hablada se convierta en uno de los de mayor tamaño del español. Asimismo PRESEEA aspira a que sus materiales de lengua hablada permitan realizar estudios desde perspectivas muy diferentes, interesando a especialistas de distintos campos: dialectólogos y sociolingüistas, fonetistas, gramáticos, analistas de la conversación, analistas del discurso, pragmáticos o etnolingüistas.

Los fundamentos metodológicos de este proyecto internacional quedan resumidos en los siguientes puntos:

1. PRESEEA se articula en una red de equipos de investigación sociolingüística, adscritos voluntariamente al proyecto y que se responsabilizan de la recogida de los materiales de lengua hablada procedentes de

una comunidad. Las comunidades estudiadas en el proyecto son núcleos urbanos, monolingües o bilingües, con una población hispanohablante bien asentada, con conciencia de comunidad y suficientemente diversa desde un punto de vista sociológico.

2. Los hablantes cuya lengua oral se recoge y analiza han de ser, por nacimiento o años de residencia, miembros reconocidos de su comunidad. La selección de los informantes para las entrevistas se hace siguiendo los criterios fijados en un muestreo por cuotas con afijación uniforme. Las cuotas se crean a partir de tres variables: sexo/género, edad y nivel de instrucción. También es posible atender a otras variables, como profesión, ingresos económicos, condiciones de alojamiento o modos de vida, pero se tratan como variables de post-estratificación.
3. El estudio de las comunidades con poblaciones de entre medio millón y un millón de habitantes se aborda mediante muestras formadas por 54 informantes. Cuando la ciudad es de un tamaño mayor y de una mayor complejidad sociológica, se recomienda el empleo de muestras de 72 o de 108 hablantes. También se prevé la posibilidad de recolectar materiales de una gran urbe procediendo al tratamiento independiente de sus barrios, colonias o sectores más representativos.
4. Los materiales destinados a formar parte del corpus oral se reúnen mediante conversaciones semi-dirigidas y grabadas con magnetófono a la vista en situación de entrevista. Las entrevistas se realizan proponiendo a los informantes módulos temáticos (saludos, el tiempo, lugar donde vive, familia y amistad, costumbres, peligro de muerte, anécdotas importantes en la vida, deseo de mejora económica) que permiten acceder a diversos tipos de discurso. Las conversaciones tienen una duración mínima de 45 minutos.
5. Los materiales grabados se transcriben siguiendo convenciones basadas en el sistema *SGML* y las normas internacionales de la TEI (*Text Encoding Initiative*).
6. Los equipos que integran PRESEEA se coordinan para la realización de análisis sobre los materiales de lengua hablada recopilados.

La participación de los equipos de investigación de PRESEEA se formaliza mediante la aceptación del núcleo metodológico común, si bien cada equipo puede desarrollar la investigación en otros aspectos específicos de interés para el conocimiento de la lengua hablada en su comunidad.

A partir de estas bases metodológicas, los equipos que integran PRESEEA están reuniendo un corpus de lengua española hablada representativo del mundo

hispanico en su variedad geográfica y social, para un lapso que va de 1996 a 2010, en su primera fase. Hasta el momento (finales de 2005), PRESEEA agrupa a 24 equipos de investigación sociolingüística que están realizando sus investigaciones sobre otros tantos núcleos urbanos y que ya han reunido materiales de ocho países hispanohablantes: Argentina, Colombia, Ecuador, España, EE.UU., Guatemala, México, Puerto Rico y Venezuela (véase la relación de equipos en el cuadro final).⁴³

Como equipo internacional de investigación, PRESEEA coordina los trabajos de más de 60 investigadores permanentes, además de numerosos becarios, encuestadores y personal de apoyo, y hasta hoy ha producido alrededor de 50 publicaciones (libros, capítulos, artículos; véase la bibliografía seleccionada). Además, está permitiendo la realización de tesis doctorales, tesis de licenciatura y otros muchos trabajos académicos de investigación. Todo ello está despertando un interés que se refleja en las más de 500 páginas electrónicas en las que se cita el proyecto y en las más de 50.000 visitas que la página electrónica de PRESEEA ha recibido entre 2003 y 2005.

La página electrónica del proyecto ofrece las siguientes informaciones:

1. Información sobre los equipos participantes y datos personales de todos los investigadores implicados en el proyecto.
2. Documentación básica de la investigación: metodología general; marcas y etiquetas utilizadas en la transcripción; cuestionario para el estudio de las formas de tratamiento.
3. Referencias bibliográficas de todas las publicaciones generadas por el proyecto.

⁴³ Los equipos de investigación están recibiendo ayudas y subvenciones desde distintas fuentes, principalmente de sus respectivas universidades. Desde 1996, el equipo de la Universidad de Alcalá ha recibido financiación externa: «Proyecto para el estudio sociolingüístico del español de la Comunidad de Madrid». Comunidad Autónoma de Madrid. Nº Referencia: 06/0076/1997 (1997-1999); «Proyecto para el estudio sociolingüístico del español de España y América». Ministerio de Educación y Ciencia. Referencia: PB98-0710 (1999-2002); «Proyecto para el Estudio Sociolingüístico del Español de España y América: reunión científica y jornadas sobre sociolingüística hispánica». Referencia: 03-134 A-486.05 (2003). Además se cuenta con la financiación concedida por el Ministerio de Educación y Ciencia de España para un proyecto integrado en el que participan las universidades de Alcalá, Granada, Las Palmas, Lérida, Málaga y Valencia (Referencia: HUM2004-06052-C06). El equipo de El Colegio de México ha recibido financiación para los siguientes proyectos: «Estructura fónica de la diversidad lingüística en México». CONACYT, 27598H; «Diversidad y variación fónica en las lenguas de México». CONACYT, 44731H.

4. Textos de publicaciones generadas por el proyecto.
5. Información sobre noticias generadas durante el desarrollo del proyecto en cada uno de los equipos participantes.
6. Información sobre las reuniones de investigadores del proyecto.
7. Acceso a la lista de correo electrónico del proyecto, mantenida desde El Colegio de México (preseea@colmex.mx).
8. Enlaces con páginas que ofrecen herramientas básicas de investigación sociolingüística (programas informáticos de análisis de la voz, programas para análisis estadísticos, normas internacionales de transcripción y etiquetado de corpus).
9. Enlaces con páginas electrónicas de interés para la sociolingüística y la geolingüística del español (revistas, seminarios y grupos de investigación, información bibliográfica general, entidades que subvencionan proyectos de investigación).

Pero, sin duda alguna, la información más valiosa de esta página electrónica está constituida por las muestras de lengua hablada reunidas por los diferentes equipos de investigación. En la actualidad, el corpus oral que se ofrece abierta y gratuitamente está formado por una muestra de unas 100.000 palabras, muestra que crecerá en la medida en que se vayan concluyendo las transcripciones. En el futuro, a través de la página se podrá acceder a una muestra muy significativa de los materiales de PRESEEA y a las voces originales de los informantes, almacenadas en archivos de sonido con alineamiento automático de voz y de texto, así como a la búsqueda de secuencias de caracteres en todo el corpus oral y a sus concordancias.

Con la mirada puesta en el futuro, este proyecto internacional prevé contar con una red integrada por entre 20 y 30 equipos de investigación, que han de cubrir las más significativas áreas geolectales del mundo hispánico. Por otro lado, se intenta que el corpus PRESEEA de lengua hablada sea lo suficientemente amplio y tenga la suficiente calidad como para utilizarse tanto en la investigación lingüística básica como en las más variadas aplicaciones de la tecnología del habla. Para ello, se trabaja con el horizonte cuantitativo de los 10 millones de palabras, de los cuales ya hay recogidos y transcritos una quinta parte aproximadamente. La representatividad geográfica y social de los materiales de lengua hablada será una de las notas caracterizadoras del corpus. Finalmente, con el fin de proceder al análisis de materiales de forma coordinada, PRESEEA cuenta con una lista de temas de estudio preferentes para cada nivel lingüístico y no se renuncia a abordar el estudio de otros aspectos geolingüísticos y sociolingüísticos de relevancia singular, como pueden ser las formas de tratamiento, para lo cual ya se ha hecho público un cuestionario que está en proceso de experimentación.

Como hemos señalado en otro lugar (Moreno Fernández, en prensa), PRE-SEEA aún no ha alcanzado sus metas principales, pero el hecho de que en este momento estén trabajando de forma coordinada decenas de investigadores, incorporados a más de 20 equipos de investigación de España, México, Puerto Rico, Colombia, Argentina, Guatemala, Venezuela, Ecuador y los Estados Unidos de América y de que ya se haya reunido un corpus parcial de 2 millones de palabras es la esperanza de un futuro prometedor.

SELECCIÓN DE BIBLIOGRAFÍA EMANADA DE PRESEEA

- Casanovas Catalá, M. 2000: «No cale que vengas, porque hoy plegaré tarde: Mecanismos de adaptación léxica en el español de los catalanohablantes», *Analecta Malacitana* 23, 2, pp. 697-709.
- Gómez Molina, J. R. 2001: *El español hablado de Valencia. Materiales para su estudio. I. Nivel sociocultural alto. Cuadernos de Filología*, Anejo 46, Valencia, Universidad de Valencia.
- 2002: «El corpus del español hablado de Valencia: proyecto para el estudio sociolingüístico del español de España y América (PRESEEA)», en Blas, J. L., Casanova, M., Fortuño, S. y Porcar, M., *Estudios sobre lengua y sociedad*, Castellón, Universitat Jaume I, pp. 217-235.
- Martín Butragueño, P. 2000: «Contacto de dialectos y aprendizaje de la variación lingüística», *Anuario de Letras* 38, pp. 309-325.
- 2000: «Los malentendidos naturales en el estudio sociolingüístico de la ciudad de México», *Nueva Revista de Filología Hispánica* 48, pp. 373-391.
- y Lastra, Y. 2000: «El modo de vida como factor sociolingüístico en la ciudad de México», en Martín Butragueño, P., *Estructuras en contexto*, México, El Colegio de México, pp. 13-43.
- Moreno Fernández, F. 1996: «Metodología del 'Proyecto para el Estudio Sociolingüístico del Español de España y de América'», *Lingüística* 8, pp. 257-287.
- 2005: «Project for the Sociolinguistic Study of Spanish from Spain and America (PRESEEA) - A Corpus with a Grammar and Discourse Bias», en Takagaki, T., Zaima, S., Tsuruga, Y., Moreno Fernández, F. y Kawaguchi, Y., *Corpus-Based Approaches to Sentence Structures*, Ámsterdam, John Benjamins, pp. 265-288.
- en prensa: «El español en su variedad geográfica y social. Informe sobre el corpus PRESEEA», *Oralia*.
- 2000: «La sociolingüística de Alcalá de Henares en el Proyecto para el Estudio Sociolingüístico del Español de España y América (PRESEEA)», *Oralia* 3, pp. 149-168.

- 2001: «El Proyecto para el Estudio Sociolingüístico del Español de España y América (PRESEEA): antecedentes, objetivos y estado actual», en Ruiz Miyares, L. y otros, *Actas del VII Simposio Internacional de Comunicación Social*, Málaga, Centro de Lingüística Aplicada - Universidad de Málaga, pp. 45-47.
- 2001: «El Proyecto para el estudio sociolingüístico del español de España y América», en Ruiz Miyares, L., Escobedo Beceiro, D., Ferrer Pérez, J. y Corpas Pastor, G., *Actas del VII Simposio Internacional de Comunicación Social (22-26 de enero de 2001, Santiago de Cuba)*, Santiago de Cuba - Málaga, Centro de Lingüística Aplicada de Cuba - Universidad de Málaga, pp. 45-57.
- 2002: *La lengua hablada en Alcalá de Henares. Corpus PRESEEA - ALCALÁ. I Hablantes de Instrucción Superior*, Vol. I, Alcalá de Henares, Universidad de Alcalá.
- 2005: *La lengua hablada en Alcalá de Henares. Corpus PRESEEA - ALCALÁ. II Hablantes de Instrucción Media*, Vol. I, Alcalá de Henares, Universidad de Alcalá.
- Vida Castro, M. 2002: «Estudio fonético experimental de /-s/ implosiva en el marco del Proyecto PRESEEA. Parámetros acústicos y percepción de la aspiración ante oclusiva sorda», *Actas del IV Congreso de Lingüística General*, Cádiz, Universidad de Cádiz, pp. 2.505-2.514.
- 2004: *Estudio sociofonológico del español hablado en la ciudad de Málaga*, Málaga, Universidad de Málaga.

GUÍA BÁSICA DE PRESEEA

Página electrónica: www.linguas.net/preseea

Equipos asociados hasta 2005 (coordinadores)

PRESEEA-ALCALÁ-ES (F. Moreno, A. Cestero, I. Molina, F. Paredes)

PRESEEA-BARCELONA-ES (A. Torres Torres)

PRESEEA-BARRANQUILLA-CO (Y. Rodríguez Cadena)

PRESEEA-BOGOTÁ-CO (M.C. Henríquez)

PRESEEA-CÁDIZ-ES (M. Casas)

PRESEEA-CARACAS-VE (P. Bentivoglio, I. Malaver)

PRESEEA-CIPOLLETTI-AR (C. Fernández)

PRESEEA-CULIACÁN-MX (M. López, E. Mendoza)

PRESEEA-GRANADA-ES (J. A. Moya)

PRESEEA-GUATEMALA-GU (L. Verdugo)

PRESEEA-LAS PALMAS-ES (J. A. Samper)
PRESEEA-LÉRIDA-ES (M. Á. Calero)
PRESEEA-MADRID-ES (F. Moreno, A. Cestero, I. Molina, F. Paredes)
PRESEEA-MÁLAGA-ES (J. Villena)
PRESEEA-MÉRIDA-MX (L. Rosado)
PRESEEA-MÉXICO-MX (Y. Lastra, P. Martín Butragueño)
PRESEEA-MIAMI-EU (H. López Morales)
PRESEEA-MONTERREY-MX (L. Rodríguez Alfano)
PRESEEA-PALMA DE MALLORCA-ES (M. Romera)
PRESEEA-PEREIRA-CO (R. Areiza)
PRESEEA-QUITO-EC (R. Gómez)
PRESEEA-SAN JUAN DE PUERTO RICO-PR (L. Ortiz)
PRESEEA-VALENCIA-ES (J. R. Gómez Molina, ValEsCo)
PRESEEA-ZARAGOZA-ES (M. A. Martín Zorraquino)

Coordinación del proyecto: Francisco Moreno Fernández
Dpto de Filología. Facultad de Filosofía y Letras. Trinidad, 5. Universidad de Alcalá. 28801 - Alcalá de Henares (Madrid). España (Tlf.: +34 91 885 4000 - Fax: +34 91 885 4413; correo electrónico: francisco.moreno@uah.es)

Lista de correo electrónico: Pedro Martín Butragueño. El Colegio de México. Camino del Ajusco, 20. CP 10740. México D.F. (Tlf.: +525 449 3000 - Fax: +525 645 0464; correo electrónico: pmartin@colmex.mx)

