



METODOLOGÍA DEL
"PROYECTO PARA EL ESTUDIO SOCIOLINGÜÍSTICO
DEL ESPAÑOL DE ESPAÑA Y DE AMÉRICA"
(PRESEEA)

VERSIÓN REVISADA - OCTUBRE 2003

1.- INTRODUCCIÓN

En el mes de abril de 1993, durante la celebración del X Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina (ALFAL), tuvo lugar una reunión de la Comisión de Sociolingüística de esta Asociación, en la que se decidió iniciar un proyecto para el estudio sociolingüístico de los núcleos urbanos de Iberoamérica y de la Península Ibérica.

La Comisión de Sociolingüística de ALFAL decidió poner en marcha un proyecto que incluyera tres actividades:

- 1ª.- Creación de un Servicio de Documentación Sociolingüística para el ámbito iberoamericano y de la Península Ibérica (lenguas española y portuguesa).
- 2ª.- Creación de un Corpus Sociolingüístico del español (PRESEEA).
- 3ª.- Creación de un Corpus Sociolingüístico del portugués (PRESOPO).

La finalidad del proyecto es coordinar las investigaciones sociolingüísticas de Iberoamérica y de la Península Ibérica para facilitar la comparabilidad de los estudios y el intercambio de información básica.

La adhesión al proyecto de la Comisión de Sociolingüística por parte de los investigadores y los centros de investigación vinculados a ALFAL es voluntaria. El fundamento del proyecto es la colaboración: se trata de aportar información propia para recibir información sobre las actividades de otros investigadores y, además, de aportar materiales procedentes de un territorio y reunidos de acuerdo con un método pre-determinado, para recibir materiales recogidos en otras zonas con el mismo método.

La coordinación del proyecto de la Comisión de Sociolingüística de ALFAL requiere el funcionamiento de un "Centro de Información y Materiales Sociolingüísticos" (CIMAS), que tiene su sede en la Universidad de Alcalá de Henares (España). Las

universidades e instituciones que aportan información y materiales al proyecto son centros asociados.

De las actividades propuestas por la Comisión de Sociolingüística, sin duda la más ambiciosa es la creación del Corpus Sociolingüístico y sincrónico de la lengua española, por un lado, y de la lengua portuguesa, por otro. El proyecto para la lengua española se denomina "Proyecto para el Estudio Sociolingüístico del Español de España y de América". El nombre formado por las siglas (PRESEEA, *preseea*) quiere expresar las intenciones generales del proyecto: llegar a ser algo tan *valioso* para el futuro conocimiento de la lengua española, como *útil* para las personas que se ocupan de ella.

El Corpus estaría formado por los materiales que proporcionaran los centros que se asociaran al proyecto y que siguieran sus directrices generales. Para la creación de un Corpus con las suficientes garantías, sería necesario realizar las siguientes tareas:

1ª.- Elaboración de una metodología sociolingüística básica. Todos los centros asociados al proyecto que se comprometieran al envío de materiales deberían ajustarse a ella; sólo de esta forma se podría garantizar la reunión de unos materiales homogéneos y, por lo tanto, comparables.

2ª.- Reunión del Corpus. Los materiales recogidos por los centros asociados serían enviados al CIMAS en una forma acordada previamente.

En las páginas siguientes presentamos una propuesta metodológica para la formación, la distribución y el estudio del Corpus Sociolingüístico de PRESEEA.

2.- ANTECEDENTES Y JUSTIFICACIÓN.

El conocimiento de la lengua española, en lo que se refiere a su uso en España, en el continente americano o en los territorios hispanohablantes de África y de Asia, ha alcanzado durante los últimos veinte años una dimensión insólita hace tan sólo medio siglo. En 1964 afirmaba Lope Blanch lo siguiente:

No creo que resulte excesivamente exagerado afirmar que el "español de América" sigue siendo un ilustre desconocido. Ya el mismo nombre, de aplicación global, con que se designa indiscriminadamente a tantas y tan diferentes modalidades del habla hispánica — *español de América* — es buena prueba del estado de ignorancia relativa en que nos encontramos.

Pero los lamentos que varias generaciones de lingüistas han proferido, por no disponer de información de primera mano sobre América, Guinea, Filipinas o sobre la hablas urbanas españolas, han comenzado a hacerse más débiles porque algunos de esos lingüistas decidieron pasar de la queja a la acción, de los anaqueles de las bibliotecas a los cuestionarios, las grabaciones y los documentos. Gracias a estos trabajos, la geolingüística y la sociolingüística hispánicas han experimentado en los últimos años un avance digno de reconocimiento.

En el campo de la geografía lingüística hispánica, la publicación de atlas ha sido continua a lo largo de la segunda mitad del siglo XX, a pesar de las enormes carencias que se han arrastrado desde la primera mitad del siglo. En España, tras el primer

volumen del *Atlas Lingüístico y Etnográfico de Andalucía*, aparecieron algunos materiales del *Atlas Lingüístico de la Península Ibérica*, el *Atlas Lingüístico y Etnográfico de las Islas Canarias*, el *Atlas Lingüístico y Etnográfico de Aragón, Navarra y Rioja*, el *Léxico de los marineros peninsulares*, el *Atlas Lingüístico y Etnográfico de Cantabria*; y se encuentran en distintas fases de elaboración el *Atlas Lingüístico de España y Portugal* y el *Atlas Lingüístico (y etnográfico) de Castilla-La Mancha*. Deben añadirse a estas labores las aportaciones al *Atlas Linguarum Europae*, al *Atlas Lingüístico del Mediterráneo* y al *Atlas Lingüístico Románico*.

En América, se han concluido o puesto en marcha, de forma ininterrumpida, numerosos proyectos geolingüísticos: a la colección de mapas de Navarra (*El español de Puerto Rico*), hay que sumar el volumen del *Atlas Lingüístico y Etnográfico del Sur de Chile*, el *Atlas Lingüístico y Etnográfico de Colombia* y el *Atlas Lingüístico de México*, además de otros trabajos en curso de realización, como el *Atlas Diatópico y Diastrático del Uruguay* o el *Atlas Lingüístico de Cuba* Pero, sin duda alguna, el proyecto que permitirá obtener por primera vez una visión lingüística de conjunto de toda la América hispanohablante será el *Atlas Lingüístico de Hispanoamérica*, que quisiéramos ver concluido en un plazo razonable de tiempo.

En lo que se refiere a la investigación sociolingüística, desarrollada en los últimos treinta años, hay que señalar que, si tomamos el concepto de sociolingüística en su sentido más amplio, esto es, dejando un lugar para todos los trabajos que se preocupan por las relaciones entre la lengua y la sociedad, la sociolingüística hispánica incluiría las corrientes de estudio que a continuación se exponen:

a) Dialectología social. Ha sido la corriente de más temprano desarrollo. Ha surgido como prolongación de unos estudios de dialectología que gozan de una gran tradición y prestigio dentro del mundo hispánico. En líneas generales, se trata de investigaciones que atienden a diversos factores sociales y en las que se utilizan técnicas de recogida de datos habituales en geolingüística .

Dentro de este ámbito pueden incluirse los trabajos realizados por Manuel Alvar sobre el habla de Las Palmas de Gran Canaria y los estudios derivados de los materiales sociolingüísticos de los atlas regionales. Del mismo modo, podemos incluir algunas investigaciones de José Pedro Rona. Un autor y otro han contribuido de forma importante al posterior desarrollo de la sociolingüística en el mundo hispánico.

Desde nuestro punto de vista, una de las principales expresiones de lo que podríamos denominar "Dialectología social" es el "Proyecto para el estudio de la norma culta de las principales ciudades de Iberoamérica y de la Península Ibérica", promovido por Lope Blanch a mediados de los años sesenta. Este gran proyecto ha propiciado la publicación de materiales de la norma culta de numerosas ciudades: Madrid, Sevilla, Lima, Santiago de Chile, San Juan de Puerto Rico, etcétera. Debe tenerse en cuenta, no obstante, que, a pesar del enorme mérito que supone la recopilación de una gran masa de materiales comparables de todo el mundo hispánico y de los muchos trabajos que se han derivado de ellos, los datos que ofrece este magno proyecto se ciñen a la norma culta y tienen, por lo tanto, un valor sociolingüístico muy limitado.

b) Sociología del lenguaje. Los trabajos de sociología del lenguaje de Fishman y Gumperz, así como la ideas de Charles Ferguson han tenido una amplia repercusión en el mundo hispánico, especialmente en lo que se refiere al estudio de las situaciones bilingües. En Hispanoamérica, la sociología del lenguaje ha centrado su interés, prioritariamente, en las relaciones entre el español y las lenguas indígenas. A sabiendas de que cualquier selección es injusta, creemos que merecen una mención

especial los estudios realizados por Y. Lastra sobre situaciones bilingües de México o por A. Escobar sobre el bilingüismo de las regiones andinas. Permítasenos no comentar las muchas publicaciones sobre sociología del lenguaje, en general, y sobre bilingüismo, en particular, que se han ocupado del español en Estados Unidos.

c) Sociolingüística variacionista. Esta corriente de investigación sigue las directrices generales marcadas por la sociolingüística de William Labov y de sus colaboradores, tanto estadounidenses como canadienses. Entre los pioneros hispánicos de este campo tienen un lugar destacado los nombres de H. Cedergren, H. López Morales, C. Silva-Corvalán y B. Lavandera. Ahora bien, actualmente la nómina de investigadores ha crecido de forma importante: destacan aquéllos que se han centrado en el español del Caribe (Orlando Alba), en el español de Argentina y Uruguay (por ejemplo, Adolfo Elizaincín) y en el español de Perú (Rocío Caravedo).

La sociolingüística hispánica ha surgido, en términos generales, de las escuelas de dialectología y tiene entre sus rasgos más característicos el cultivo de la dialectología social. Más tarde, la influencia de la investigación norteamericana ha marcado el rumbo de un buen número de estudiosos que se reparten entre la sociolingüística de la variación y la sociología del lenguaje.

Todo lo que se ha reseñado hasta este momento, especialmente lo referido a los campos de la geolingüística y de la sociolingüística, nos lleva a extraer las siguientes conclusiones:

1ª.- La segunda mitad del siglo XX está siendo testigo de la aparición de numerosos estudios y proyectos de investigación que están llevando — y van a llevar — por vez primera a un conocimiento serio y suficientemente amplio de la realidad del español en España y en América.

2ª.- Los proyectos de investigación concluidos y los que están en desarrollo constituyen un conjunto de antecedentes que permiten pensar, desde unos sólidos fundamentos, en nuevas líneas de investigación. El mundo hispánico universitario e investigador cuenta ya con una experiencia que le permite afrontar con garantías otros proyectos ineludibles. Es cierto que no todas las regiones del mundo hispanohablante cuentan con nutridas nóminas de investigadores o con unos recursos de investigación de última generación; pero también es cierto que en otros muchos lugares hay personas, medios y recursos suficientes para realizar tareas de apoyo y de asesoramiento.

3ª.- No existe en el mundo hispánico un proyecto de estudio sociolingüístico de intención semejante a la del "Proyecto para el estudio de la norma culta". Precisamente este proyecto, impulsado y coordinado por Lope Blanch, es el antecedente más claro y cercano del proyecto que aquí presentamos, aunque las diferencias entre uno y otro son palmarias: PRESEEA no se va a limitar a la recogida y estudio de la norma culta sino que aspira a conocer las variedades urbanas del español con toda la profundidad que sea posible. El estudio sociolingüístico completo del español de España y de América deberá suponer un paso adelante en la investigación lingüística.

Y un último comentario preliminar. Hispanoamérica es la región más urbanizada entre las menos desarrolladas del mundo. Prácticamente tres cuartas partes de la población vive en las ciudades y se espera que en el 2025 cerca del 85% de la población hispanoamericana sea urbana. Ello supone que el conocimiento detallado del español

de América pasa ineludiblemente por el estudio sociolingüístico de sus núcleos urbanos.

3.- CUESTIONES METODOLÓGICAS.

3.1.- Las comunidades de habla.

Las líneas metodológicas de PRESEEA recogerán los principios teóricos que nos han servido como punto de partida. En nuestra opinión, estas orientaciones han de ser muy generales o, dicho de otro modo, tendrían que ser lo suficientemente amplias como para permitir su aplicación en cualquier núcleo urbano hispánico en el que se dieran cita algunas condiciones. Las condiciones que debería ofrecer un núcleo urbano para que su estudio pudiera acogerse al espíritu de PRESEEA serían muy poco excluyentes: debería ser un núcleo urbano hispanohablante — monolingüe o bilingüe —, con una población — o una parte de ella — tradicionalmente asentada en el lugar y que presentara cierta heterogeneidad sociológica.

Las razones que justifican estos requisitos mínimos son fáciles de explicar. En primer lugar, partimos de núcleos hispanohablantes porque el objetivo general del proyecto es conseguir un corpus sociolingüístico sincrónico de lengua española. Esos núcleos pueden ser monolingües en español o bilingües, pero en este último caso el español debe ser una lengua de uso frecuente en la comunidad y los hablantes bilingües que se puedan seleccionar deben ser capaces de hacer un uso del español funcionalmente similar al de un monolingüe. Es evidente que un estudio adecuado y completo de las ciudades bilingües obliga a tener en cuenta muchos elementos que aquí no se van a considerar: actitudes lingüísticas hacia cada una de las lenguas, distribución socioestilística y funciones sociales de las lenguas implicadas, etc. En estos casos, los investigadores directamente responsables del estudio de cada ciudad bilingüe serán los que determinen en qué modo se pueden combinar nuestros criterios metodológicos con los que se desprendan de las características específicas de cada núcleo urbano.

En segundo lugar, es preciso trabajar con núcleos que cuenten con población asentada desde mucho tiempo atrás para estar seguros de que se ha creado una conciencia de comunidad de habla con una configuración socioestilística conocida y reconocida por los propios hablantes (P-II). Por otro lado, para tener una mínima garantía de que el esfuerzo investigador va a merecer la pena, conviene trabajar con núcleos que ofrezcan una variedad y una riqueza sociológica interna (P-IV).

Las directrices metodológicas de PRESEEA se aplicarán sobre comunidades de habla que se asociarán a núcleos urbanos concretos (P-I). Es posible que en muchos casos una comunidad de habla propiamente dicha sobrepase los límites de una ciudad determinada, pero hemos de trabajar sobre entidades que se puedan delimitar con relativa objetividad. La recogida de datos para PRESEEA puede hacerse sobre cualquier comunidad de habla que se ajuste a las condiciones ya comentadas, si bien parece interesante no dejar en un segundo plano el estudio de las más grandes ciudades de España y de Hispanoamérica. En el Cuadro 1 queda recogido el número de habitantes de las dos ciudades más pobladas de cada país hispánico americano.

Buenos Aires (Ar) (1991) 2 965 403 (conglomerado urbano: 10 911 403)
Córdoba (Ar) (1991) 1 179 067
La Paz (Bo) (1992) 711 036
Santa Cruz de la Sierra (Bo) 694 616
Santa Fe de Bogotá (Co) (1993) 5 726 957
Cali (Co) (1993) 1 783 546
San José (CR) (1992) 302 574
Alajuela (CR) (1991) 158 276
La Habana (Cu) (1989) 2 077 938
Santiago de Cuba (1990) 405 354
Santiago (Ch) (1992) 5 180 757
Concepción (Ch) (1992) 330 448
Guayaquil (Ec) (1990) 1 508 844
Quito (1990) 1 100 847
San Salvador (ES) (1992) 422 570
Santa Ana (ES) (1992) 202 337
Madrid (Es) (1991) 3 084 673
Barcelona (Es) (1991) 1 681 132
Ciudad de Guatemala (Gu) (1995) 1 167 495
Escuintla (Gu) (1995) 123 048
Tegucigalpa (Ho) (1989) 608 100
San Pedro Sula (Ho) (1988) 321 197
México D.F. (Mé) (1990) 8 235 744 (Ciudad de México: 18 747 400)
Guadalajara (Mé) (1990) 2 178 000
Managua (Ni) (1985) 682 111
León (Ni) (1985) 100 982
Ciudad de Panamá (pan) (1990) 584 803
San Miguelito (Pan) (1990) 243 025
Asunción (Par) (1992) 502 426
Ciudad del Este (Par) (1992) 133 893
Lima (Pe) (1993) 6 434 328 (incluye Callao)
Arequipa (Pe) (1993) 633 428
San Juan (PR) (1990) 437 745;
Bayamón (PR) (1990) 220 262
Santo Domingo (RD) (1989) 2 200 000
Santiago de los Caballeros (RD) (1989) 467 000
Montevideo (Ur) (1985) 1 311 976
Salto (Ur) (1985) 80 823
Caracas (Ve) (1990) 1 822 465 (área metropolitana: 2 784 042)
Maracaibo (Ve) (1990) 1 363 873

Cuadro 1.- Ciudades más pobladas y número de habitantes. Fuente: *Almanaque mundial 1996*, Editorial Televisa, 1995.

Este cuadro nos proporciona información acerca de las ciudades cuya habla podría quedar recogida en el corpus, información que nos dará pie, por ejemplo, para hablar de muestreos. Pero antes debemos hacer unos comentarios generales.

Ya nos hemos referido a la importancia de los núcleos urbanos en Hispanoamérica, importancia que se hace especialmente notable en Sudamérica: la tasa de urbanización de Venezuela es del 91%, la de Uruguay del 89% y la de Buenos Aires del 86,3%. Si unimos a esto el hecho de que el ritmo de crecimiento de las grandes ciudades sigue aumentando, y a una gran velocidad, llegamos a la conclusión de que estas ciudades, que recogen casi toda la población del país, han de ser forzosa y llamativamente heterogéneas y albergan individuos que manifiestan un muy diverso grado de integración en la vida urbana.

Ante una situación de esta naturaleza surgen los problemas para identificar lo característico de una ciudad y lo que simplemente se están dando cita en una ciudad, aunque en el futuro pueda convertirse en seña de identidad. La dificultad está en fijar el concepto de hablante bonaerense, hablante caraqueño, hablante limeño o hablante sanjuanero. A este respecto, el interés de PRESEEA se decanta claramente por el español de los núcleos urbanos hispánicos, pero serán los investigadores responsables de cada equipo los que decidan qué es esencial y qué circunstancial o accesorio, cuál es el universo absoluto y cuál el universo relativo que se va a manejar en la investigación, partiendo de nuestras consideraciones metodológicas y de las muchas posibilidades que éstas van a dejar abiertas o por desarrollar. Como orientación de carácter general, cabe apuntar que los hablantes candidatos a informantes deberían haber nacido en la ciudad, haber llegado a ella antes de cumplir los diez años o llevar viviendo allí más de veinte, siempre y cuando su origen lingüístico no fuera marcadamente diferente.

3.2.- El muestreo.

Una muestra debe ser representativa del universo que va a servir de base al estudio sociolingüístico. Dado que los núcleos sobre los que potencialmente puede aplicarse nuestra metodología son muchos y diferentes entre sí, la muestra-tipo que se propone tiene un objetivo principal: permitir la recogida de habla partiendo de unos parámetros comparables desde un punto de vista sociológico y estilístico. Esto quiere decir que proponemos trabajar sobre un universo relativo común a todas las comunidades de habla hispánicas, que garantice la comparabilidad de los materiales. Los investigadores locales podrán acrecentar libremente el contorno de este universo relativo, de acuerdo con sus intereses particulares, y trabajar con universos relativos más amplios.

Proponemos la preparación de muestras por cuotas con afijación uniforme. Consisten en dividir el universo relativo en subpoblaciones, estratos o cuotas — atendiendo a unas variables sociales determinadas— y en asignar igual número de informantes a cada una de esas cuotas. Una razón que nos lleva a preferir este sistema y no una muestra aleatoria o probabilística es que la muestra por cuotas permite una más fácil comparación estadística entre las cuotas internas de la misma muestra y entre muestras diferentes. Además, de esta forma el investigador se obliga a buscar informantes más allá de sus círculos de influencia o de las personas que tienden a dejarse entrevistar más fácilmente.

Consideramos razonable crear las cuotas de la muestra a partir de tres variables sociales: el sexo, la edad y el grado de instrucción. En un proceso de post-estratificación debería atenderse a otros factores: profesión, ingresos económicos y

condiciones de alojamiento. Así mismo se propone tener en cuenta, provisionalmente y de modo experimental, la variable "modo de vida". La información sobre estas variables secundarias se recogería por medio de un cuestionario cumplimentado con cada uno de los informantes y, en el caso del modo de vida, también a partir de la información recogida durante las conversaciones. Además, mediante la combinación de los datos de estas variables sería posible trabajar con otra variable más de post-estratificación: el nivel socio-cultural.

La muestra-tipo que proponemos queda reflejada en el siguiente cuadro:

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

Cuadro 2.- Muestra-tipo por cuotas.

Como ha quedado dicho, las variables sociales a partir de las cuales se propone dividir el universo son el sexo, la edad y el grado de instrucción. Todas ellas permiten un estudio sociolingüístico cuantitativo (P-VII). Acerca de la conveniencia y el interés de trabajar con la variable "edad" poco se puede comentar que no esté recogido en la bibliografía sociolingüística: se trata simplemente de una variable esencial en cualquier trabajo de este campo. A la vista de lo que se ha decidido en otras investigaciones del mundo hispánico y con un deseo de primar la simplicidad sobre la casuística, proponemos distinguir tres generaciones: 1, de 20 a 34 años; 2, de 35 a 54 años; 3, de 55 años en adelante. Hay que valorar que la esperanza de vida al nacer, en Hispanoamérica, se mueve entre los 60 y los 75 años aproximadamente.

Sobre la inclusión de las variables "sexo" y "grado de instrucción" en las muestras, sólo cabe decir que poquísimos estudios han decidido prescindir de ellas, a pesar de que en un buen número de análisis el sexo se ha revelado como una variable de escasa capacidad explicativa. La comparabilidad de nuestros estudios con los resultados de decenas de investigaciones aconsejan mantenerla en la estratificación. Las variantes que se distinguen en la variable "grado de instrucción" son las siguientes:

1. Analfabetos, sin estudios. Enseñanza Primaria (hasta 10-11 años de edad aprox.), 5 años aproximadamente de escolarización;
2. Enseñanza Secundaria (hasta 16-18 años de edad aprox.), 10-12 años aproximadamente de escolarización;
3. Enseñanza Superior (universitaria, técnica superior) (hasta 21-22 años de edad aprox.), 15 años aproximadamente de escolarización.

Las variables que se van a manejar en la post-estratificación harán posible la comparación con los resultados de investigaciones anteriores y servirán de punto de referencia. Las variantes que proponemos para esas variables de post-estratificación se exponen a continuación.

Ingresos económicos (variable de validez local exclusivamente): se recomienda distinguir cinco variantes o categorías.

Condiciones de alojamiento: 1, Vivienda sin comodidades sanitarias y de difícil acceso; 2, Casa o piso modesto; 3, Casa o piso elegante y espacioso, con muchas comodidades.

Profesión: 1, Buhoneros y vendedores ambulantes, obreros no especializados urbanos, obreros campesinos, servicio doméstico, servicios no especializados; 2, Pequeños comerciantes, secretarios y oficinistas, obreros especializados, artesanos, mecánicos, vendedores en tiendas, cobradores, ayudantes técnicos, policías y guardias, soldados; 3, Profesionales universitarios, personal docente de educación media y primaria, pequeños empresarios y productores, mandos intermedios, técnicos, supervisores; 4, Profesionales universitarios de libre ejercicio, gerentes medios del sector público y privado, militares con graduación, medianos empresarios y productores, docentes universitarios; 5, Altos funcionarios del poder ejecutivo, legislativo y judicial, altos oficiales del ejército, grandes empresarios privados, grandes hacendados, altos ejecutivos del sector público y privado.

La variable de post-estratificación "modo de vida" exige una justificación pormenorizada. El concepto de "modo de vida", presentado por Højrup y desarrollado por James Milroy, permite poner en relación las redes sociales de pequeñas dimensiones con otras estructuras o grupos sociales de mayor entidad. Los modos de vida responden a un modelo en el que varios grupos étnicos o de clase quedan representados como elementos internamente estructurados y relacionados con otros grupos. En este modelo, la conducta lingüística obedece más al poder de determinación de las redes y de las estructuras en las que se mueven los hablantes, que a los atributos percibidos como característicos de determinados grupos sociales. A su vez, hay redes que tienen capacidad para imponer sus normas socio-lingüísticas a otras más débiles. Se trata en definitiva de dar prioridad al tipo de actividad laboral y familiar, y a las relaciones que los hablantes mantienen con otros miembros del grupo, sobre ciertas características o atributos clasificatorios. Los grupos son considerados como una consecuencia de las estructuras fundamentales de la sociedad que dividen la población en modos de vida sustancialmente diferentes.

Los modos de vida que proponen Højrup y Milroy, y que proponemos manejar de modo experimental, provisional y absolutamente voluntario para este proyecto, son los siguientes:

Modo de vida 1.- Unidad primaria de producción (agricultura, pesca, pequeños servicios). Relaciones cooperativas entre compañeros de profesión. Familia implicada en la producción. Autoempleo. Escaso tiempo libre: cuanto más se trabaja, más se gana. Redes sociales estrechas.

Modo de vida 2.- Empleo en un sistema de producción que no es controlado por los trabajadores. Se trabaja para ganar un sueldo y poder disfrutar de periodos de tiempo libre. Relaciones laborales separadas del ámbito familiar. Cierta movilidad laboral. Redes estrechas de solidaridad con los compañeros y los vecinos.

Modo de vida 3.- Profesión cualificada, capaz de controlar la producción y de dirigir los trabajos de otras personas. Tiempo de vacaciones dedicado al trabajo. Se trabaja para ascender en la jerarquía y adquirir más poder. Actitud competitiva con los colegas.

Los rasgos ideológicos que caracterizarían a estos modos de vida serían "la familia" para el modo 1, "el ocio" para el modo 2 y "el trabajo" para el modo 3. Debe valorarse, no obstante, que el concepto de "modo de vida" es fundamentalmente estructural; los rasgos definidores de un grupo vienen dados por contraste con los de los demás modos. Por otro lado, las relaciones entre los tres modos de vida y las prácticas culturales asociadas a cada uno de ellos no tienen por qué ser exactamente iguales en todos los países, por lo que, en un estudio contrastivo, es importante describirlas con todo el detalle que se pueda.

A la hora de considerar la utilidad que el concepto de "modo de vida" puede tener en un proyecto como PRESEEA, es importante pensar que los tres modos de vida propuestos son lo suficientemente comunes como para esperar encontrarlos prácticamente en todas las comunidades de habla del ámbito hispánico. Además estos modos de vida recogen en esencia algunos caracteres básicos de los niveles socio-culturales y socio-económicos manejados en otros estudios y evitan varios problemas graves, como la virtual inexistencia de "clases medias". Esperamos, por tanto, que sean variables realmente explicativas de la conducta lingüística. Como argumentación favorable a la aplicación del concepto de "modo de vida" en el proyecto, añadiremos que cada equipo de investigación local podrá incluir, si decide trabajar con esta variable, otros modos de vida que quizá no se encuentran en otras comunidades, pero que tal vez no puedan ignorarse en el estudio de ciertas sociedades. Es importante, no obstante, que las características de cada modo de vida — de los tres comunes y de los específicos — y las pautas socio-culturales asociadas a cada uno de ellos queden descritas de forma completa y detallada. La información necesaria para adscribir los informantes a un modo de vida o a otro se recogerá por medio de las hojas de datos personales y de los datos que se aporten durante el desarrollo de las conversaciones.

Para el análisis de los materiales recogidos, los investigadores podrán manejar como variables explicativas bien las variables de post-estratificación de forma independiente, bien estas mismas variables combinadas en niveles socio-culturales o socio-económicos. No se recomienda el manejo de conjunto de todas ellas como variables independientes porque los solapamientos serían inevitables y los análisis cuantitativos se verían afectados.

Pero queda aún un aspecto importante por tratar, en relación con la muestra: su tamaño. Si contamos para cada una de las casillas del Cuadro 2 con una afijación uniforme de cuatro informantes, el tamaño total de nuestra muestra sería de 72 informantes. Esto supondría una proporción 1/25 000 para una ciudad de alrededor de dos millones de habitantes y mucho más holgada para las de menor población. De la relación de ciudades que hemos presentado en el Cuadro 1, sólo media docena quedarían por debajo de ese nivel canónico de representatividad (0,025), si bien es cierto que son éstas precisamente las que disfrutan de un mayor prestigio sociolingüístico y de mayor peso socio-económico (Buenos Aires, Lima, Madrid, México, Santa Fe de Bogotá y Santiago de Chile). Quedaría al criterio de los investigadores locales proceder a la recogida de materiales mediante estudios parciales (los barrios más representativos, los más populosos, etc.) o aumentando a cinco (90 informantes) o seis (108) el número de informantes por cuota de la muestra. De igual modo, en los núcleos con un número de habitantes inferior a 500 000 sería posible reducir a tres el número de informantes por cuota, lo que daría una muestra de 54 informantes (1/9250).

Para concluir este epígrafe, quisiéramos dejar clara una característica importante del proyecto. Los criterios y normas que aquí se proponen son un mínimo metodológico que busca la comparabilidad de los materiales que se recojan en los centros de

investigación asociados a PRESEEA. Se proponen unas bases de uso común, pero no exclusivo; con esto queremos decir que los investigadores locales son libres para ir más allá en las exigencias metodológicas: nada impide, por ejemplo, que se rellenen cuestionarios de actitudes lingüísticas o de otro tipo, que se aumente el número de informantes por cuotas, que también se hagan grabaciones en otros contextos, que se trabaje con un grupo de edad formado por personas con una edad comprendida entre los 14 y los 19 años, que se describan modos de vida diferentes de los que hemos descrito o que se incluyan otras variables de post-estratificación, además de las previstas. Queda fuera de duda, no obstante, que los materiales importantes para PRESEEA son aquellos que responden a unas directrices metodológicas comunes.

3.3. La recogida de datos.

La recogida de los materiales de PRESEEA se hará por medio de conversaciones grabadas, que mantendrán los investigadores con unos informantes de características predeterminadas en contextos propios de cada una de las comunidades de habla. Con el fin de lograr una mínima uniformidad estilística, que haga posible y útil la comparación de materiales de comunidades de habla diferentes, se recomienda que la recogida de datos se haga siempre en un lugar de fácil acceso para los informantes, que sea considerado como representativo de una determinada institución oficial, pero que resulte familiar para los hablantes: pensamos en un despacho o un aula de una escuela, de un centro cultural o de ocio, de un instituto, en un local de una dependencia municipal ... De otra forma, se correría el riesgo de mantener las conversaciones en lugares muy heterogéneos, a menudo con niveles altos de ruido y ante oyentes de muy diferentes características, lo que dificultaría enormemente la interpretación estilística de los materiales. Admitimos que el desplazamiento a lugares "oficiales" o poco frecuentados por los informantes, el alejamiento de su entorno familiar o laboral, puede tener como consecuencia una pérdida o una disminución de la espontaneidad y de la naturalidad de su discurso, pero lo que se pierde en este terreno se gana en el de la calidad de grabación y en el de la homogeneidad. Probablemente es cierto que ninguna conversación sociolingüística puede ser considerada como realmente espontánea, pero en este caso vamos a admitir que el desarrollo natural de las conversaciones permitiría a muchos informantes pasar de un estilo formal o semiformal a un estilo algo más espontáneo. Teniendo todo lo anterior en cuenta y en última instancia, son los responsables de cada equipo los que han de valorar si, para lo fines del proyecto, resulta más adecuada la entrevista en los lugares del tipo que se acaba de describir o en los domicilios o lugares de trabajo de los hablantes.

Los materiales destinados a formar parte del corpus de lengua hablada serán recogidos mediante conversaciones semidirigidas y grabadas con magnetófono a la vista. Las grabaciones deben ser realizadas en el medio de mejor calidad de que se disponga (magnetófono de cinta abierta, magnetófono de cinta casete, DAT, etc.). La entrevista se ha de estructurar teniendo en cuenta los módulos temáticos que se refieren a continuación:

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

9.- Final

El orden de tratamiento de los módulos puede variar según las circunstancias de cada entrevista. El módulo temático 1, por ejemplo, no tiene por qué utilizarse siempre para iniciar la conversación: en ocasiones el tema resulta algo forzado y puede plantear el problema del uso del metalenguaje. Se proponen a continuación algunas orientaciones para el tratamiento de cada módulo. La intención última es que el conjunto de materiales de una comunidad ofrezca muestras proporcionadas de cada uno de los módulos.

1. SALUDOS

Presentación

¿Cómo quiere que le trate de tú o de usted? La verdad es que es un problema esto del tratamiento, nunca sabes como tratar a las personas ¿verdad? Por ejemplo, ¿tú cómo tratas a tus amigos de tú o de usted? ¿Y si son personas mayores? ¿Si se trata de alguien joven que no conoces? ¿Y si es una persona mayor, hombre o mujer, al que por ejemplo le preguntas por una calle? ¿Y a tu médico? ¿A los vecinos con los que no tienes 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?

¿Cómo estás/ está? ¿Estás animado/a?

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

2. EL TIEMPO

¡Qué calor/frío ha hecho hoy!

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

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?

3. LUGAR DONDE VIVE

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

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

¿Te gusta vivir aquí? ¿Dónde te gustaría vivir? ¿Por qué? ¿Cómo crees que sería vivir en Madrid?

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

Supongo que este barrio habrá cambiado desde que viniste a vivir aquí. El mío también ha cambiado, la gente, el barrio en sí, recuerdo cuando era pequeña... ¿Cómo ha cambiado el tuyo? ¿Qué tiene ahora que no tenía antes? ¿Te gusta más ahora o antes? ¿Qué recuerdas tú de cuando eras pequeña?

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

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

¿Y la gente de otras edades, tus padres, tus hijos..., qué suele hacer cuando sale?

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

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

4. FAMILIA Y AMISTAD

¿Estás casada/o? ¿cómo conociste a tu marido/mujer? ¿tienes hijos/piensas tener hijos? cuéntame cómo fue. ¿Qué crees que hubieras hecho si corriera peligro tu vida o la del niño? (En caso de ser hombre preguntar que habría aconsejado a su mujer, hermana...)

¿Qué piensas de problemas actuales como 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 tí un amigo? ¿Cómo debe ser una persona o qué debe hacer para que la consideres un buen amigo? ¿Son iguales los amigos que se tienen de niños que los que se tienen de adultos? ¿Cómo eran los que tú tenías de niño?

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

¿Qué sueles hacer en un día normal? (Descripción de un día normal de su vida desde que se levanta hasta que se acuesta) ¿Estás contenta con tu forma de vida? ¿Por qué?

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

5. COSTUMBRES

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

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

¿Qué planes tienes para las próximas navidades/verano?

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

6. PELIGRO DE MUERTE

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

7. ANÉCDOTAS IMPORTANTES EN LA VIDA

Cuéntame más cosas importantes o curiosas que te hayan ocurrido, algún robo, algún premio, algún viaje especial... ¿Qué crees que se debe hacer en una situación como esa? ¿Qué harías si te volviera ocurrir o si tuvieras la oportunidad de volver al pasado?

8. DESEO DE MEJORA ECONÓMICA

¿Juegas a la lotería o a cualquier otro juego de azar? ¿Por qué? Supongo que te gustaría que te tocara algún premio gordo ¿verdad? ¿Qué harías si te tocaran 200 millones de pesetas? ¿Qué piensas/haces cuando oyes que a alguien le ha tocado una gran cantidad de dinero?

9. FINAL

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

Tengo que comprar el periódico camino de la estación ¿sabes dónde hay un puesto o librería? ¿me explicas cómo se va?

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

Adiós.

Módulos temáticos

La conversación con cada informante tendrá una duración mínima de 45 minutos y, durante su desarrollo, los investigadores, que pertenecerán, siempre que sea posible, a la misma comunidad de habla que los informantes, procurarán interrumpir a los hablantes lo menos posible. Después de la conversación grabada, el investigador procederá a comprobar que la calidad de la grabación es adecuada y a cumplimentar un cuestionario para recoger los datos personales del informante; estos datos tendrán que ver con las variables sobre las que se hará la post-estratificación (véase apéndice).

4.- TRANSCRIPCIÓN Y PUBLICACIÓN DEL CORPUS.

Una vez realizadas las grabaciones, se puede iniciar el proceso de transcripción y almacenamiento de los materiales. Para conseguir una transcripción lo más rápida posible y fácil de corregir, recomendamos transcribir en ortografía ordinaria, utilizando el procesador de texto de un ordenador o computadora (se recomiendan los programas *Word* o *Wordperfect*). Es conveniente que la reproducción de las cintas se realice mediante un dictáfono con pedal de control. Las transcripciones deben ser registradas en ASCII (Texto DOS), bien en ordenadores Apple, bien en IBM-PC y deben ser corregidas por un mínimo de dos personas.

En cuanto al sistema que se ha de seguir para transcribir los materiales grabados, parece lógico proponer el uso de un sistema internacional, previsto y admitido en los medios industriales y de investigación de un número considerable de países y que no sea tan complejo como para eternizar la tarea, siempre ardua, de la transcripción. Proponemos que PRESEEA siga las convenciones internacionales de la TEI (Text Encoding Initiative).

La TEI nació en un congreso de la Association for Computers and the Humanities celebrado en Poughkeepsie, Nueva York, en 1987. Se trata del mayor proyecto internacional de estas características y está patrocinado por la asociación ya citada, la Association for Computational Linguistics, la Association for Literary and Linguistic Computing, la National Endowment for the Humanities de los Estados Unidos, la DG XIII de la Comisión de la Unión Europea, el Canadian Social Science and Humanities Research Council y la Fundación Andrew W. Mellon. El objetivo de la TEI es desarrollar y difundir un formato bien definido para facilitar el intercambio de textos entre investigadores interesados en el procesamiento del lenguaje natural.

Las normas de este sistema de transcripción se recogen en la guía del "Standard Generalized Markup Language" (SGML). Consiste básicamente en una serie de marcas o etiquetas del tipo < > que reflejan los más diversos rasgos de los textos transcritos. Cada texto transcrito debe ir precedido de una serie de etiquetas identificativas. Las etiquetas de encabezamiento que han de aparecer son de tres tipos. En primer lugar, figuran las que permiten identificar de modo general el texto electrónico. En lo que se refiere a la transcripción, las etiquetas que indican el comienzo de un elemento estructural se inserta dentro de < >; las etiquetas que marcan el final de ese elemento se disponen dentro de </>. La transcripción propiamente dicha debe ir inmediatamente precedida de la etiqueta <texto>. Al final del texto se ha de utilizar la etiqueta </texto>. Aunque son muchas las etiquetas textuales que se pueden utilizar para marcar los materiales de la lengua hablada, se propone manejar solamente las que aparecen en los documentos de marcas y etiquetas que se incluye en la página electrónica del proyecto.

Los materiales transcritos por cada centro asociado a PRESEEA se podrán poner a disposición de los participantes en el proyecto y de la comunidad investigadora por medio de la página electrónica..

La colaboración en el corpus de PRESEEA da derecho a los centros asociados a consultar los materiales equivalentes de otros centros para emprender estudios comparativos. Los materiales serán propiedad de cada uno de los centros. La coordinación del proyecto podrá disponer de estos materiales solamente en la forma autorizada por los centros asociados.

Las funciones del CIMAS, en relación con el Corpus Sociolingüístico del Español, serán las siguientes:

- 1ª.- Establecer contactos con los centros interesados en participar en PRESEEA.
- 2ª.- Distribuir la información sobre la metodología sociolingüística básica que han de seguir los centros asociados.
- 3ª.- Prestar auxilio técnico y metodológico a los centros que lo necesiten.
- 4ª.- Desarrollar los instrumentos necesarios para la comunicación entre los investigadores del proyecto y para el acceso a sus documentos de trabajo y los materiales de lengua hablada.

Los principales de distribución de información y de comunicación entre los equipos participantes serán la página electrónica del proyecto (<http://www.linguas.net/preseea>) y la lista de correo electrónico (preseea@colmex.mx).

5.- ANÁLISIS DE LOS MATERIALES.

Una de las señas de identidad de PRESEEA es que los investigadores locales pueden sentirse completamente libres a la hora de ampliar los objetivos y las técnicas de estudio, siempre y cuando se respeten las directrices comunes. Pero esa libertad es absoluta en el terreno del análisis e interpretación de los materiales lingüísticos. Hacemos nuestros y presentamos resumidos los objetivos que Carmen Silva-Corvalán propuso en 1992 para un proyecto sociolingüístico de las características de PRESEEA:

- 1) La descripción lingüística de los procesos de variación más relevantes en el español.
- 2) La identificación de fenómenos de variación que se correspondan con cambios lingüísticos en marcha.
- 3) La descripción de la estructura y las características de diversos tipos o géneros de discurso: argumentación, planificación, exposición, narración, etc.
- 4) La descripción de elementos pragmáticos característicos de las distintas variedades del español (por ejemplo, marcadores discursivos).
- 5) La redacción de estudios gramaticales del español hablado.

PRESEEA aspira a poner en manos de los investigadores unos materiales comparables, bien transcritos, de origen contrastado, en las mejores condiciones técnicas y de la forma más eficaz posible. Pero más allá de los análisis concretos que de este proyecto puedan emanar, PRESEEA quiere ser un camino para estrechar las relaciones entre los lingüistas de ambos lados del Atlántico en beneficio de un mejor conocimiento del español.

APÉNDICE

PROPUESTA DE CUESTIONARIO

PRESEEA

CIUDAD Código del informante:

DATOS DE LA ENTREVISTA

Fecha de la entrevista:

Lugar de la entrevista:

Entrevistador/a:

Observaciones sobre la entrevista (lugar, audiencia, ruidos, incidencias, etc.):

DATOS PERSONALES DEL INFORMANTE

Nombre y apellidos:

Sexo:

Lugar de nacimiento:

Fecha de nacimiento:

Fecha de llegada a la ciudad (en su caso):

Domicilio (barrio, calle, etc.):

Condiciones de alojamiento (1, 2, 3):

Lugar de nacimiento del padre:

Lugar de nacimiento de la madre:

Lugar de nacimiento del cónyuge:

Modo de vida (1, 2, 3):

Grado de instrucción (indicar si son estudios completos y años de escolarización):

Profesión:

Ingresos (1, 2, 3, 4, 5):

Breve descripción de modo de vida:

Viajes:

Lecturas:

Televisión y radio:

Idiomas (cuáles, L1-L2, funciones y uso):

Observaciones sobre el informante:

DATOS PERSONALES DEL OYENTE

Nombre y apellidos:

Sexo:

Lugar de nacimiento:

Fecha de nacimiento:

Fecha de llegada a la ciudad (en su caso):

Modo de vida (1, 2, 3):

Grado de instrucción (indicar si son estudios completos y años de escolarización):

Profesión:

Relación con informante y con otros oyentes:

Observaciones sobre el oyente: