

SOCIOLINGÜÍSTICA ESPAÑOLA metodología

JOSÉ ANTONIO GONZÁLEZ SALGADO
Madrid

1. CONSIDERACIONES PREVIAS

Manuel Alvar, en su trabajo sobre los niveles socioculturales en el habla de Las Palmas de Gran Canaria¹, se lamentaba de la escasa atención que han dedicado los dialectólogos al estudio de las hablas urbanas. Once años más tarde, Martínez Martín se hacía eco de la misma afirmación en su tesis sobre fonética y sociolingüística en la ciudad de Burgos². Desde entonces, se han llevado a cabo diversos proyectos que han contribuido a aumentar la nómina de publicaciones que tienen como objeto el estudio del habla de las ciudades. Pero, a pesar de todo, el panorama sociolingüístico español sigue siendo precario; aún quedan multitud de comunidades que necesitan una investigación exhaustiva para conocer la estructura y características de su habla. La sociolingüística —que se ha ido haciendo con un método riguroso— está poco desarrollada en nuestro país.

Para Martínez Martín, cuatro son las causas por las que no se ha prestado interés en los estudios dialectológicos a las hablas urbanas³: la complejidad del

¹ M. Alvar, *Niveles socioculturales en el habla de Las Palmas de Gran Canaria*, Cabildo Insular, Las Palmas de Gran Canaria, 1972, pág. 15

² F. M. Martínez Martín, *Fonética y sociolingüística en la ciudad de Burgos*, CSIC, Madrid, 1983, pág. 30.

³ F. M. Martínez Martín, *loc. cit.*, págs. 31-32.

habla de las ciudades, la imposibilidad de cartografiar los materiales, el tiempo que se invertiría en las encuestas en una gran ciudad y la inadecuación entre método y finalidad de la dialectología tradicional cuando se aplica a núcleos urbanos.

Para Etxebarria Arostegui, el motivo por el que las hablas urbanas no han despertado interés dentro de la dialectología hispánica es distinto⁴. Según esta autora, los estudiosos consideraban que la ciudad carecía de diversidad al tratarse de un núcleo centralizador. Sin embargo, el problema no es que no se investigara el habla urbana, sino que cuando se hacía no se utilizaba la metodología adecuada.

Otro asunto distinto es que no se estudiaran las ciudades por considerarlas menos conservadoras que los núcleos rurales. Como afirman Chambers y Trudgill, los dialectos urbanos «tendían a ser relativamente nuevos, a menudo surgían de la inmigración de las áreas rurales adyacentes, y eran, por tanto, menos interesantes para los filólogos»⁵.

Este ser *menos interesante para los filólogos* hay que entenderlo en su dimensión histórica: la dialectología nació para la investigación de las variedades lingüísticas rurales, y la geografía lingüística se dedicó a elaborar atlas que permitieran el estudio conjunto de las hablas populares.

En el presente artículo revisaremos la metodología aplicada en una selección de obras de sociolingüística española publicadas desde 1972, atendiendo tanto a la descripción social de la lengua urbana como a la de la rural. Aunque con anterioridad a 1972 ya se había presentado algún estudio bajo el rótulo de *Sociolingüística*⁶, consideramos que es con la publicación, por parte de Manuel Alvar, de los *Niveles socioculturales en el habla de Las Palmas de Gran Canaria* cuando las investigaciones de la lengua en la sociedad cobran una especial relevancia en el mundo hispánico.

Los problemas de orden metodológico que vamos a tratar son los que atañen a la selección de las variables lingüísticas y de las variables sociales, el tamaño de la muestra, la utilización de instrumentos para recopilar materiales (entrevista, cuestionarios...), las técnicas de muestreo y el análisis estadístico.

2. VARIABLES LINGÜÍSTICAS

Una variable lingüística se define como cada unidad que posee dos o más variantes condicionadas socialmente. En palabras de Chambers y Trudgill, «podemos considerar a las variables lingüísticas frecuentemente como modos socialmente diferentes, pero lingüísticamente equivalentes de hacer o de decir lo mismo, y se dan en todos los niveles del análisis lingüístico»⁷.

⁴ M. Etxebarria Arostegui, *Sociolingüística urbana: El habla de Bilbao*, Universidad de Salamanca, 1985, pág. 68.

⁵ J. K. Chambers y P. Trudgill, *La dialectología*, Visor, Madrid, 1994, pág. 83.

⁶ Véase, por ejemplo, M. Alvar, «Sociología en un microcosmos lingüístico (El Roque de las Bodegas, Tenerife)», *Prohemio*, II, 1, 1971, págs. 5-24.

⁷ J. K. Chambers y P. Trudgill, *op. cit.*, pág. 88.

Pese a que la variabilidad se puede estudiar teóricamente en todos los niveles, en el ámbito hispánico los que más rendimiento están ofreciendo son el nivel fónico y el nivel léxico, siendo casi inexistentes los trabajos que se refieren a variables sintácticas, y poco numerosos los dedicados a las morfológicas.

Las tres propiedades que debe presentar una variable lingüística para que logre ser objeto de estudio en una comunidad de habla son, según Labov⁸: presentar una elevada frecuencia, ser una unidad estructural y denotar una distribución estratificada. Estas tres propiedades se cumplen, en mayor o menor medida, en las variables que se estudian en nuestro corpus.

En los trabajos de carácter léxico se han seguido diversas motivaciones para su estudio: Borrego Nieto evalúa el alcance general de la penetración de la lengua estándar en la modalidad lingüística de Villadepera de Sayago⁹. Azurmendi Ayerbe utiliza el léxico para establecer el grado de bilingüismo existente en la comarca de San Sebastián¹⁰. Aguado Candanedo, Etxebarria Arostegui, González Ferrero y Morín Rodríguez realizan una descripción del léxico para comprobar las diferencias sociales existentes en sus respectivas comunidades¹¹. Salvador Salvador se limita a recoger el caudal onomasiológico de un determinado grupo social: la clase culta granadina¹².

El resto de los autores que han investigado sociolingüísticamente algún punto de nuestra geografía se ha centrado con mayor atención en los demás niveles del análisis lingüístico: Labrador Gutiérrez, en uno de los primeros estudios realizados en España sobre un núcleo semiurbano con la aplicación de nuevas metodologías, analiza el fenómeno del yeísmo en Lepe (Huelva)¹³. Martínez Martín aísla dos variables fonéticas en Burgos: el yeísmo y los grupos cultos en posición silábica implosiva¹⁴. Lynn Williams, en sus *Aspectos sociolingüísticos del habla de la ciudad de Valladolid*, trabaja con siete variables, seis de carácter fonético y una de carácter sintáctico: la terminación *-ado* en participios y sustantivos, la reducción de los grupos cultos *xCons*, *-x-*, *-cc-* y *-ct-*, la aspiración de *-s* implosiva en vocablos como *asco*, *esclavo* y *esquina*, y el adverbio *muy* en expresiones

⁸ W. Labov, *Modelos sociolingüísticos*, Cátedra, Madrid, 1983, pág. 36.

⁹ J. Borrego Nieto, *Sociolingüística rural: Investigación en Villadepera de Sayago*, Universidad de Salamanca, 1981.

¹⁰ M. J. Azurmendi Ayerbe, *Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial en la comarca de San Sebastián*, Caja de Ahorros Provincial de Guipúzcoa, San Sebastián, 1983.

¹¹ D. Aguado Candanedo, *El habla de Bercianos del Real Camino (León)*, Diputación Provincial de León, 1984; M. Etxebarria Arostegui, *op. cit.*; J. C. González Ferrero, *La estratificación sociolingüística de una comunidad semiurbana: Toro (Zamora)*, Universidad de Salamanca, 1991; A. Morín Rodríguez, *Actitudes sociolingüísticas en el léxico de Vegueta (Las Palmas de Gran Canaria)*, Universidad de Las Palmas, 1993.

¹² F. Salvador Salvador, *Léxico del habla culta de Granada*, Universidad de Granada, 1991.

¹³ T. Labrador Gutiérrez, *Sociolingüística andaluza: LL-Y en Lepe (Huelva)*, Diputación Provincial de Huelva, 1975.

¹⁴ F. M. Martínez Martín, *op. cit.*, págs. 75-200.

como *muy bueno* o *muy grande*¹⁵. Fernández Juncal se detiene en el estudio de dos interesantes variables en el oriente de Cantabria: la metafonía y el neutro de materia¹⁶. En Quintanar de la Orden (Toledo), Moreno Fernández¹⁷ seleccionó ocho aspectos de la lengua hablada coloquial para analizar «el grado de determinación que emisor, receptor y mensaje ejercen en varios actos de conducta lingüística y social»¹⁸; los ocho actos de habla a que se refiere el autor son: tratamiento, expresiones vocativas, excusas, ofrecimientos, peticiones, respuesta a agradecimientos, presentaciones y saludo. Trigo Cutiño describe y analiza cuatro configuraciones discursivas en el habla de los niños sevillanos: paréntesis o incisos, factor común, enumeraciones y simetrías¹⁹. Por último, Samper Padilla estudia cinco variables fonológicas en el español hablado en la capital de la isla de Gran Canaria²⁰: segmentos fonológicos /s, r, l, n/ en posición silábica implosiva y /d/ en el contexto intervocálico.

3. VARIABLES SOCIOLÓGICAS

Como se ha afirmado en numerosas ocasiones, la diversificación lingüística existe hasta en la más pequeña comunidad. Todos los asentamientos poblacionales (aldeas, pueblos, barrios o ciudades) contienen una diversidad social que condiciona el comportamiento lingüístico de los habitantes. Por todos es conocido el hecho de que no siempre hablan igual los hombres y las mujeres, pero esa diferencia básica se complica extremadamente si observamos las divergencias que se producen entre los hablantes de distintas generaciones, y más aún si nos fijamos en la categoría social (nivel cultural+nivel económico) de esos hablantes.

Las variables sociológicas con las que suelen trabajar los investigadores españoles son: sexo, edad, nivel educativo (estudios) y nivel económico (profesión). Las dos primeras se pueden considerar más variables biológicas que sociales, ya que no están condicionadas por el ambiente, sino por la propia naturaleza de la persona.

¹⁵ L. Williams, *Aspectos sociolingüísticos del habla de la ciudad de Valladolid*, Universidad de Valladolid, 1987.

¹⁶ C. Fernández Juncal, *Variación y prestigio: Estudio sociolingüístico en el oriente de Cantabria*, CSIC, Madrid, 1998.

¹⁷ F. Moreno Fernández, «Análisis sociolingüístico de actos de habla coloquiales», *Español Actual*, 51, 1989, págs. 5-51.

¹⁸ F. Moreno Fernández, *loc. cit.*, pág. 5.

¹⁹ J. M. Trigo Cutiño, *El habla de los niños de Sevilla*, Alfar, Sevilla, 1990.

²⁰ J. A. Samper Padilla, *Estudio sociolingüístico del español de Las Palmas de Gran Canaria*, Imprenta Pérez Galdós, Las Palmas de Gran Canaria, 1990.

3.1. Sexo

La división de la población en hombres y mujeres es utilizada como variable por todos los autores de nuestro corpus en una proporción del 50 % aproximadamente. El objeto que se persigue al operar con esta variable es descubrir posibles diferencias en el habla de los hombres y las mujeres. Sin embargo, la mayoría de los autores no encuentra marcas significativas propias de uno u otro sexo. Para Borrego Nieto, Aguado Candanedo, Etxebarria Arostegui o Trigo Cutiño, el sexo no es factor determinante de las diferencias lingüísticas²¹. Williams, no obstante, sí encuentra esas diferencias en el habla de los jóvenes pucelanos²², aunque quizá habría que buscarlas, como afirma Alvar²³, no en el distinto sexo, sino en el modo de vida llevado por las mujeres en la comunidad.

3.2 Edad

La edad, en el nuevo enfoque del estudio de las variedades lingüísticas, ha adquirido una importancia extraordinaria. Ya no tiene sentido analizar un habla sin tener en cuenta esta variable sociológica, a no ser que la investigación se circunscriba a un determinado grupo generacional. La sociolingüística contempla como posible informante a cualquier habitante de la comunidad, sea cual sea su edad, formación académica o nivel económico.

En cuanto a la edad, dos son los criterios que se pueden seguir para trabajar en sociolingüística: tomar como núcleo de investigación un grupo de edad o estudiar el habla a lo largo del devenir generacional.

3.2.1. Un grupo generacional

Trigo Cutiño analiza el habla de los niños sevillanos dividiéndolos en tres grupos²⁴: a) Niños de 5 años (comienzo de la escolaridad); b) Niños de 8 años (3º de escolaridad obligatoria); y c) Niños de 11 años (final del segundo ciclo). Azurmendi Ayerbe también se centra en el habla infantil, aunque contempla solamente dos divisiones²⁵: Alumnos de 5º de EGB, de 10-11 años, y alumnos de 8º de EGB, de 13-14 años. Lynn Williams estudia el habla de los jóvenes vallisoletanos de 18 a 26 años, coincidiendo con los límites de la enseñanza superior²⁶.

3.2.2. Varios grupos generacionales

Borrego Nieto elige tres edades —que no generaciones— suficientemente alejadas entre sí²⁷: los informantes más jóvenes tienen en torno a 20 años (antes

²¹ J. Borrego Nieto, *op. cit.*, pág. 342; D. Aguado Candanedo, *op. cit.*, pág. 260; M. Etxebarria Arostegui, *op. cit.*, pág. 518; J. M. Trigo Cutiño, *op. cit.*, pág. 205.

²² L. Williams, *op. cit.*, pág. 141.

²³ M. Alvar, *Lecturas de sociolingüística*, Edaff, Madrid, 1977, pág. 69.

²⁴ J. M. Trigo Cutiño, *op. cit.*, págs. 37-43.

²⁵ M. J. Azurmendi Ayerbe, *op. cit.*, págs. 206-218.

²⁶ L. Williams, *op. cit.*, pág. 48.

²⁷ J. Borrego Nieto, *op. cit.*, pág. 50.

del servicio militar), los del segundo grupo en torno a 45 (situación familiar estable) y los más ancianos en torno a 70 (pensionistas y jubilados)²⁸.

El sistema de dos generaciones tan sólo ha sido utilizado por un autor en el ámbito hispánico: Labrador Gutiérrez en su estudio del habla de Lepe²⁹. Labrador distingue una generación de adultos (Generación A) y una generación de niños, adolescentes y jóvenes (Generación B). Esta última se distribuye en tres grupos diferentes: B1 (niños menores de 7 años); B2 (niños y adolescentes de 7 a 15 años); y B3 (jóvenes entre 15 y 20 años).

La división de la sociedad en tres generaciones es la más utilizada en los estudios sociolingüísticos, aunque no siempre coinciden los límites cronológicos en todos los trabajos:

Martínez Martín y Samper Padilla hacen una división de acuerdo con el siguiente esquema³⁰: a) Primera generación: de 20 a 35 años; b) Segunda generación: de 35 a 55 años; c) Tercera generación: mayores de 55 años. Morín Rodríguez y Salvador Salvador restringen en cinco años la edad de la primera generación, pero contemplan de forma idéntica las demás³¹. Etxebarria Arostegui adelanta la edad de los informantes de la primera generación (de 15 a 30 años) y amplía el grupo de la segunda (de 30 a 60 años)³².

El sistema de cuatro generaciones tiene la desventaja de complicar los resultados finales, ya que las variables se multiplican al combinarse entre sí, pero quizá sea el más representativo de lo que ocurre en la comunidad.

Gómez Molina utiliza los siguientes parámetros en Sagunto³³: a) Joven: de 16 a 25 años; b) Maduro: de 26 a 45 años; c) Adulto: de 45 a 65 años; y d) Anciano: mayores de 65 años. Moreno Fernández se fija en los siguientes grupos³⁴: a) Habitantes de hasta 20 años; b) Habitantes de una edad comprendida entre los 21 y los 35 años; c) Habitantes de una edad entre 36 y 50 años; y d) Habitantes de más de 51 años. Por último, González Ferrero divide la población de Toro (Zamora) en cuatro grupos representativos³⁵: a) Primera generación: 18-30 años; b) Segunda generación: 40-45 años; c) Tercera generación: 60-65 años; y d) Cuarta generación: 75-80 años.

El único trabajo donde se contemplan más de cuatro generaciones es el que realizó Alvar en 1972 en Las Palmas de Gran Canaria³⁶. Este autor dividió la población en bloques de diez años, lo que dio como resultado la configuración de ocho grupos distintos.

²⁸ Esta es, exactamente, la misma división que hace Aguado Candanedo en su estudio del habla de Bercianos del Real Camino (León).

²⁹ T. Labrador Gutiérrez, *op. cit.*, pág. 42.

³⁰ F. M. Martínez Martín, *op. cit.*, págs. 55-62; J. A. Samper Padilla, *op. cit.*, pág. 33.

³¹ A. Morín Rodríguez, *op. cit.*, pág. 32; F. Salvador Salvador, *op. cit.*, pág. 12.

³² M. Etxebarria Arostegui, *op. cit.*, pág. 174.

³³ J. R. Gómez Molina, *Estudio sociolingüístico de la comunidad de habla de Sagunto (Valencia)*, Instituto Alfons el Magnánim, Valencia, 1986, pág. 18.

³⁴ F. Moreno Fernández, «Análisis sociolingüístico de actos de habla coloquiales», pág. 21.

³⁵ J. C. González Ferrero, *op. cit.*, pág. 94.

³⁶ M. Alvar, *Niveles socioculturales*, pág. 36.

3.3. Nivel sociocultural

Uno de los factores que más influye en el comportamiento lingüístico de los hablantes es el grado de instrucción que poseen. Sin embargo, esta variable no puede considerarse independiente del nivel económico de los sujetos, pese a que en las comunidades modernas no es raro que se den casos de individuos con un alto grado de preparación cultural que ocupan un puesto profesional que les acarrea pocos ingresos. Según esto, lo más lógico es considerar que la variable denominada nivel sociocultural está compuesta por tres parámetros que intervienen en su determinación: 1. educación; 2. profesión; y 3. ingresos. Con la interacción de estos tres parámetros podemos conseguir una división estratificada de la población como la que postuló Humberto López Morales en San Juan de Puerto Rico³⁷.

Las escalas que se han creado en los estudios españoles atienden desde uno hasta seis grados distintos de instrucción, aunque lo más habitual es utilizar una triple partición.

Samper Padilla divide la población de Las Palmas de Gran Canaria en seis niveles socioculturales³⁸: a) Sin estudios; b) Primer grado (1ª etapa de EGB); c) Segundo grado, primer ciclo (2ª etapa de EGB); d) Segundo grado, segundo ciclo (BUP, COU, FP); e) Tercer grado, nivel A (escuelas universitarias o equivalentes); y f) Tercer grado, nivel B (facultades y equivalentes).

Cinco niveles distingue Gómez Molina en Sagunto³⁹: a) Analfabeto; b) Primarios incompletos; c) Primarios completos; d) Secundarios; y e) Medios y superiores.

Los cuatro niveles de que se valen Alvar, Williams y González Ferrero se esquematizan de la siguiente forma⁴⁰: a) Estrato alto: estudios universitarios; b) Estrato medio-alto: estudios medios; c) Estrato medio-bajo: estudios primarios y secundarios sin terminar; d) Estrato bajo: estudios primarios y analfabetos.

La división de la sociedad en tres niveles es —como quedó antes dicho— la más usual. En España se han servido de ella Borrego Nieto, Aguado Candanedo, Martínez Martín, Etxebarria Arostegui y Calero Fernández⁴¹, según los siguientes parámetros: a) Nivel superior (estudios universitarios); b) Nivel medio (estudios medios); y c) Nivel inferior (estudios primarios o sin estudios).

Carmen Fernández Juncal establece dos categorías extremas en su estudio sociolingüístico del oriente de Cantabria⁴²: a) Analfabetos, sin estudios, con estudios primarios o bachiller elemental; b) Con bachiller superior o estudios universitarios.

³⁷ H. López Morales, *Estratificación social del español de San Juan de Puerto Rico*, Universidad Nacional Autónoma de México, 1983.

³⁸ J. A. Samper Padilla, *op. cit.*, pág. 36.

³⁹ J. R. Gómez Molina, *Estudio sociolingüístico*, pág. 19.

⁴⁰ M. Alvar, *Niveles socioculturales*, págs. 35-36; L. Williams, *op. cit.*, pág. 39; J. C. González Ferrero, *op. cit.*, págs. 86-87.

⁴¹ J. Borrego Nieto, *op. cit.*, págs. 50-51; D. Aguado Candanedo, *op. cit.*, pág. 22; F. M. Martínez Martín, *op. cit.*, págs. 55-62; M. Etxebarria Arostegui, *op. cit.*, pág. 177; M. A. Calero Fernández, *Estudio sociolingüístico del habla de Toledo*, Pagès editors, Lérida, 1993, pág. 84.

⁴² C. Fernández Juncal, *op. cit.*, págs. 41-44.

Otra posibilidad para analizar una variedad lingüística es ceñirse a un solo grupo sociocultural. Este es el caso de Salvador Salvador, que analiza el léxico de la clase culta granadina con estudios universitarios⁴³.

3.4. Nivel socioeconómico

Ya hemos dicho en el punto anterior que esta variable no es independiente del nivel sociocultural, sino que ambas forman el nivel social de los individuos. El nivel socioeconómico se calcula mediante los ingresos que reciben los sujetos, y se suele estudiar atendiendo a la profesión que desempeñan.

Existen cuatro posibilidades para establecer el nivel social de la población, todas ellas usadas por los investigadores españoles:

3.4.1. El nivel social se concreta según el nivel cultural de los informantes. En los trabajos de Borrego Nieto y Aguado Candanedo no se tiene en cuenta la variable *nivel socioeconómico*, sobre todo porque se trata de estudios aplicados a comunidades rurales, donde las diferencias sociales se muestran más claramente en el nivel de instrucción que en el poder económico.

3.4.2. El nivel social se establece según el nivel cultural y el nivel económico, considerados independientemente. Etxebarria Arostegui, tras indicar los tres grados de instrucción, estratifica la sociedad bilbaína en tres grupos según la profesión de los individuos:⁴⁴ a) Nivel medio-alto (industriales, altos directivos, gerentes, etc.); b) Nivel medio (comerciantes, vendedores, personal de servicios, etc.); y c) Nivel medio-bajo (obreros, peones, pequeños agricultores y ganaderos, etc.).

Aunque ambas variables (nivel cultural y nivel económico) son consideradas independientes, existe una relación clara entre los estratos. Los sujetos que pertenezcan al nivel cultural inferior se situarán, en su mayoría, en el nivel económico medio-bajo; los que posean una formación superior ocuparán mayoritariamente los niveles medio y medio-alto.

3.4.3. El nivel social se establece atendiendo al status social, poder económico y grado de cultura del informante. Tanto Moreno Fernández como Gómez Molina consideran que el nivel cultural no puede dissociarse del poder económico⁴⁵. Ambos autores dividen la sociedad en cuatro clases: alta, media-alta, media-baja y baja.

3.4.4. El nivel social se calcula teniendo en cuenta únicamente la variable nivel socioeconómico. Labrador Gutiérrez contempla dos grupos en su estudio del yeísmo en Lepe (Huelva)⁴⁶: los campesinos y los profesionales liberales. Aunque no se fija en la cultura de los hablantes, se puede afirmar que —por regla general—

⁴³ F. Salvador Salvador, *op. cit.*

⁴⁴ M. Etxebarria Arostegui, *op. cit.*, pág. 176.

⁴⁵ F. Moreno Fernández, «Análisis sociolingüístico de actos de habla coloquiales», pág. 24; J. R. Gómez Molina, *Estudio sociolingüístico*, pág. 19.

⁴⁶ T. Labrador Gutiérrez, *op. cit.*, pág. 42.

los integrantes del primer grupo poseerán una instrucción menor que los del segundo, por lo que implícitamente la variable nivel cultural también está presente en este tipo de trabajos.

3.5. Otras variables

3.5.1. Variable viajes

En las comunidades rurales, los viajes que efectúan los informantes son un factor que determina el contacto con otras normas. Borrego Nieto, en Villadepera de Sayago⁴⁷, y Aguado Candanedo, en Bercianos del Real Camino⁴⁸, han establecido tres grados: a) Mínimo: informantes que no han salido nunca, o que han salido muy esporádicamente, del lugar; b) Restringido: informantes con salidas poco frecuentes; y c) Amplio: informantes con salidas abundantes y residencias prolongadas en el exterior.

Los dos autores han unificado las variables nivel de instrucción y viajes en una sola denominada *contacto con la norma*⁴⁹.

3.5.2. Variable origen

El lugar de nacimiento de los informantes suele ser el de la comunidad sobre la que se realiza el estudio. No obstante, Etxebarria Arostegui considera tres variantes en su investigación del habla de Bilbao, según los individuos hayan nacido en la propia capital, en otros lugares de las provincias vascas o en el resto de provincias españolas⁵⁰.

El lugar de residencia dentro de la comunidad es también tenido en cuenta por otros autores. Gómez Molina selecciona dos variantes según el barrio donde habitan los informantes: Sagunto ciudad y Sagunto puerto⁵¹.

3.5.3. Variable lengua materna

En comunidades bilingües es de trascendental importancia para evaluar los resultados. Azurmendi Ayerbe, en San Sebastián, clasifica a sus informantes según tengan como lengua materna el euskera, el español o ambas lenguas⁵². Gómez Molina utiliza el mismo sistema respecto al valenciano⁵³.

⁴⁷ J. Borrego Nieto, *op. cit.*, pág. 51.

⁴⁸ D. Aguado Candanedo, *op. cit.*, págs. 22-23.

⁴⁹ Para Fernández Juncal, el contacto con la norma incluye el grado de instrucción, los estudios del cónyuge, los estudios de los hijos, el nivel de lectura, la afición a escuchar la radio y ver la televisión, y los viajes realizados (C. Fernández Juncal, *op. cit.*, págs. 41-46).

⁵⁰ M. Etxebarria Arostegui, *op. cit.*, pág. 175.

⁵¹ J. R. Gómez Molina, *Estudio sociolingüístico*, pág. 18.

⁵² M. J. Azurmendi Ayerbe, *op. cit.*, págs. 206-218.

⁵³ J. R. Gómez Molina, *Estudio sociolingüístico*, pág. 21.

4. TAMAÑO DE LA MUESTRA

La teoría sociolingüística no proporciona la cifra exacta de informantes que son necesarios para que los datos manejados arrojen conclusiones definitivas. A todo lo más que se ha llegado ha sido a afirmar, como hace Sankoff⁵⁴, que con un número de 150 informantes como máximo se puede realizar un trabajo completo en la sociedad más enmarañada. Sin embargo, el problema no es tanto el número de individuos que son necesarios para que los resultados sean fiables, como las variables sociológicas que operan en la comunidad objeto de estudio.

Al igual que para la geografía lingüística un idiolecto refleja —aparte de las idiosincrasias particulares— las actitudes lingüísticas generales de un dialecto, ese mismo idiolecto reflejará los patrones principales del sociolecto en que se incluye.

El tamaño de la muestra depende, no solo del tipo de comunidad sobre el que se trabaja, sino también de los objetivos o los resultados que se pretenden conseguir. El número de informantes seleccionados no debe ser, teóricamente, el mismo en una comunidad rural que no sobrepasa los mil habitantes, que los de una gran urbe que cuente con varios cientos de miles. De igual manera, por razones prácticas, no es lo mismo trabajar con una sola variable lingüística que con varias decenas, ni tampoco seleccionar una variable social como el sexo —para lo que se pueden utilizar varios representantes masculinos y femeninos— que seleccionar cinco o seis variables distintas, que con sus combinaciones entre sí multiplicarían el tamaño de la muestra innecesariamente si se eligiera más de un representante de cada variante.

Según la observación de Labov⁵⁵, si una muestra está estratificada adecuadamente, basta seleccionar 25 hablantes para una población de 100.000 habitantes, es decir, el 0,025 % del universo. No obstante, esta afirmación es completamente relativa, ya que hay comunidades en las que ese número sería suficiente; pero en otras, que presentan una gran complejidad estructural (ciudades bilingües o con varias razas y culturas), con esos 25 informantes sería imposible el estudio social de la lengua.

En España, hasta ahora, no parece haber sido tenida muy en cuenta la observación de Labov, y las muestras se mueven entre el 7,73 % del universo en el trabajo de Aguado Candanedo y el 0,012 % en la investigación de Etxebarría Arostegui.

5. TÉCNICAS DE MUESTREO

Toda comunidad que va a ser objeto de estudio necesita ser parcelada para simplificar el trabajo. En una investigación de tipo lingüístico-social se tienen

⁵⁴ G. Sankoff, «A Quantitative Paradigm for the Study of Communicative Competence», en Bauman and Sherzer (eds.), *Explorations in the Ethnography of Speaking*, CUP, Cambridge, 1974, págs. 18-49.

⁵⁵ W. Labov, *op. cit.*, pág. 259.

que realizar dos tareas previas a la recogida de materiales. La primera es la división en sectores representativos, es decir, hacer una estructuración de la sociedad según diversas variables (sexo, edad, nivel socioeconómico...). La segunda es tomar una muestra de cada variante. El número de componentes de la muestra será lo suficientemente amplio como para poder establecer conclusiones que afecten también al resto de los individuos no presentes en el muestreo. En todo caso, la selección dependerá del carácter y los objetivos de la investigación, así como del tipo de comunidad a la que se aplica.

Lo que se pretende, al elegir una muestra, es que ésta sea una reproducción exacta, pero en pequeño, del universo.

En España, la técnica de muestreo más utilizada es la censal. Los investigadores recurren al padrón municipal, al censo electoral o a otros recuentos estadísticos de la población para concretar las variables sociológicas que operan y para escoger los informantes que son necesarios para su estudio. Este sistema ha sido puesto en práctica, entre otros, por González Ferrero en Toro, Borrego Nieto en Villadepera de Sayago, Moreno Fernández en Quintanilla de la Orden y Gómez Molina en Sagunto. Etxebarria Arostegui, que también usó el censo para dividir las variables sociales de la muestra, eligió, sin embargo, núcleos familiares para recopilar los datos de su tesis⁵⁶.

6. RECOGIDA DE DATOS: EL CUESTIONARIO

La técnica de recogida de datos, en sociolingüística, no se aparta tanto de la utilizada por la dialectología o la geografía lingüística. El instrumento básico para realizar una encuesta sigue siendo el cuestionario. Su ventaja principal es que permite comparar los resultados de todas las entrevistas realizadas⁵⁷. Aun así, dentro de los trabajos que se presentan bajo el epígrafe de *sociolingüística* conviene hacer una separación atendiendo a la naturaleza de los cuestionarios utilizados:

Por una parte se encuentran los estudios que aplican un cuestionario general procedente de la dialectología sin ningún tipo de modificación, hecho que se documenta con bastante frecuencia en las investigaciones españolas, sobre todo en los trabajos de carácter léxico.

Por otra parte tenemos los estudios que se realizan con cuestionarios elaborados desde una perspectiva sociolingüística de acuerdo con la teoría existente.

Por último, encontramos trabajos que prescinden de la ayuda del cuestionario y que siguen otros métodos de investigación distintos (observación participativa, observación estructurada, etc.).

Alvar, González Ferrero, Etxebarria Arostegui, Salvador Salvador y Morín Rodríguez utilizan cuestionarios que pertenecen al primer tipo. Son cuestionarios de una extensión irregular, pensados para recoger el caudal léxico de la comunidad.

⁵⁶ M. Etxebarria Arostegui, *op. cit.*, pág. 173.

⁵⁷ J. K. Chambers y P. Trudgill, *op. cit.*, pág. 46.

Los más completos son los de Alvar, que utiliza el mismo del *Atlas Lingüístico y Etnográfico de Canarias*, y el de Salvador, que usa el del *Estudio Coordinado de la Norma Lingüística Culta*. Los demás son reelaboraciones parciales de éstos o del *Cuestionario del Atlas Lingüístico de España y Portugal (ALEP)*.

Borrego Nieto utilizó cuatro cuestionarios, dos de carácter léxico y dos de carácter sociolingüístico⁵⁸.

Especialmente interesante es el cuestionario que elaboró Williams para el estudio del habla de Valladolid⁵⁹. Está estructurado en diversos niveles: conversación, lectura de un texto, lectura rápida de una lista de palabras, lectura de una lista de palabras a velocidad normal y lectura de una lista de palabras agrupadas en pares.

Tanto Labrador Gutiérrez, como Samper Padilla y Calero Fernández prescinden del cuestionario al elaborar sus trabajos. El primero recopila sus materiales atendiendo a un diálogo espontáneo con sus informantes; el segundo se basa en la observación participativa; la tercera desarrolla sus encuestas en un *clima de conversación*⁶⁰.

7. ANÁLISIS ESTADÍSTICO

El hecho que ha permitido el extraordinario avance de los trabajos sociolingüísticos en los últimos años es, sin duda, la aplicación del análisis estadístico a las investigaciones sobre la lengua. La estadística se ha convertido en el *alma mater* de esta disciplina científica, ya que contribuye a hacer objetivos los datos obtenidos y permite relacionarlos matemáticamente entre sí. Algunos de los procedimientos utilizados por la sociolingüística —como el test de Pearson— son el núcleo central de muchos trabajos⁶¹. Hasta el momento, los estudios más completos en el ámbito español, en lo que se refiere al análisis estadístico, son los de Etxebarria Arostegui, González Ferrero y Gómez Molina⁶².

⁵⁸ J. Borrego Nieto, *op. cit.*, págs. 45-49.

⁵⁹ L. Williams, *op. cit.*, págs. 51-53.

⁶⁰ M. A. Calero Fernández, *op. cit.*, pág. 85.

⁶¹ Véase la excelente síntesis realizada por F. Moreno Fernández, *Metodología sociolingüística*, Gredos, Madrid, 1990, págs. 122-158; además de sus *Principios de sociolingüística y sociología del lenguaje*, Ariel, Barcelona, 1998.

⁶² M. Etxebarria Arostegui, *op. cit.*; J. C. González Ferrero, *op. cit.*; J. R. Gómez Molina, *Actitudes lingüísticas en una comunidad bilingüe y multilectal. Área metropolitana de Valencia*, Universidad de Valencia, 1998.