

METODOLOGÍA SOCIOLINGÜÍSTICA

No cabe duda de que un aspecto de importancia capital en la investigación sociolingüística lo constituye la selección cuantitativa y cualitativa de los informantes cuyas muestras de habla constituirán el corpus. Esto es así por dos razones. En primer lugar, es evidente, y lo es más aún en el caso de comunidades grandes, que no todos los individuos pueden ser encuestados. Y en segundo lugar, porque la sociolingüística se mueve dentro de la estadística inferencial. Lo que verdaderamente le interesa como ciencia y lo que hace que sus resultados sean relevantes es el hecho de que la caracterización y el comportamiento sociolingüístico detectado en un conjunto de hablantes sea extensible a la población total de la que ese grupo se ha extraído. Es decir, la posibilidad de hacer inferencias sobre las características de la población a partir de las características de la muestra.

No es de extrañar, pues, que todo el aparato metodológico de esta rama de la lingüística se experimente y desarrolle con un doble objetivo: por un lado, el de garantizar la representatividad de la muestra, verdadera piedra angular sobre la que se levanta la sociolingüística; y, por otro, el de asegurar la fiabilidad y significación de las conclusiones de la investigación. Contempla, así, dificultades y desafíos metodológicos de carácter general para todos los estudios de esta naturaleza, *-tamaño de la muestra y constitución cualitativa de la misma-*, y, de índole específica, determinados por los propios objetivos con los que la investigación nace, *tipo de muestreo*. Del acierto en el establecimiento de la muestra y de la adecuación entre los objetivos y los instrumentos de obtención de datos es de donde surge la validez de la sociolingüística, la significación de sus conclusiones y el papel que desempeña en el panorama de los estudios de lingüística en general.

Determinar el número exacto de informantes a partir de los cuales la representatividad de la muestra está garantizada, disminuye o se pierde, es un asunto complejo en el que mezclan continuamente aspectos cuantitativos y cualitativos. En primer lugar, porque el tamaño de la muestra depende de la propia comunidad de habla de donde se extrae, de los objetivos de la investigación, y de la complejidad del grupo humano cuyo conocimiento se persigue. Y además, porque se parte de dos hechos admitidos sin discusión: la homogeneidad lingüística y el *umbral de*

*significación*¹. El primero de ellos hace referencia a la recurrencia de fenómenos lingüísticos aun dentro de la variación, y el segundo, término acuñado por Francisco Moreno, a la existencia de un límite numérico antes del cual los datos son significativos pero a partir del cual son una mera repetición de conductas lingüísticas ya registradas. El mismo autor, al tratar este punto, señalaba en una obra imprescindible para la sociolingüística, que la suma de los dos factores explicaba que no fuera necesario un número muy elevado de sujetos, aunque no determinaba cuántos. De hecho afirmaba que "el estudio idóneo contará con un número de hablantes que ronde ese umbral (el de significación). No cabe duda de que sus palabras están perfectamente justificadas desde el punto de vista racional y científico. No hay una razón matemática ni lingüística indiscutible para asegurar que un número concreto de hablantes, y no cifras superiores o inferiores, sea el umbral de significación. Sin embargo, es igualmente evidente que tal conclusión sume al investigador en la perplejidad y en la desorientación precisamente en los primeros pasos de su investigación y especialmente si aborda por primera vez un estudio de corte sociolingüístico.

Por todo ello, los intentos de precisar el tamaño de la muestra han sido un reto constante a lo largo de la sociolingüística y, como apoyos, valiosísimos e imprescindibles por más que todos admitan discusión, matizaciones y reservas. Nos vamos a referir aquí a los dos principales: uno de carácter porcentual, representado fundamentalmente por Labov, que establece en el 0.0025%, o lo que es lo mismo, 25 hablantes de cada 100.000, el número idóneo de informantes; y otro de carácter numérico, que tiene entre sus defensores más ilustres a Gillian Sankoff, y que estima como suficiente y significativa la inclusión de datos de 150 hablantes, incluso para comunidades muy complejas.

Para valorar la rentabilidad de estos métodos, no se puede perder de vista un hecho muy simple. Cuando el investigador aborda el estudio de una comunidad de habla, y muy especialmente si es una comunidad de habla urbana, parte del hecho de que existe una estratificación en ese grupo por motivos biológicos, socioeconómicos y culturales, y de que esos estratos no están uniformemente constituidos, y presentan, por lo tanto, distintos volúmenes.

La utilización del método porcentual ignora por completo el tamaño real de las poblaciones y el volumen de los estratos que la constituyen. Eso significa que se puede aplicar tanto a poblaciones de 300 individuos como de 150.000. Hecho que no es baladí, puesto que este límite porcentual puede dar como válido un tamaño de muestra que no permita –precisamente en contra de lo que se pretende– que todos los grupos representativos de la población estén representados en ella. La técnica de muestreo que debe utilizarse en los estudios de hablas de una comunidad es la de cuotas de afijación proporcional, dado que es la única que garantiza que la

¹ Véase Moreno (1990).

complejidad del entramado social y biológico se reproduzca en la muestra con todos sus rasgos y en la misma proporción en la que se da en la población. Cuanto menor sea el tamaño real de la población más bajo será también el número de los individuos que formen parte de la muestra, y mayores dificultades habrá para que quede plasmado con exactitud el tejido social de la comunidad. Los grupos poco representados, *marginales*, que sin embargo existen y que contribuyen a su caracterización, razón por la que no pueden ser obviados en un muestreo estratificado por cuota, quedarían eliminados o significativamente reducidos. Así, pues, la representatividad de la muestra se vería muy seriamente comprometida, o anulada.

Un ejemplo claro de las distorsiones que podría producir la aplicación de este método en el estudio del habla de una comunidad fue la investigación llevada a cabo en la ciudad de Burgos a propósito de las formas verbales personales (Larrosa Barbero 1999). La población en la fecha del estudio era de 167.104 habitantes. La aplicación, por tanto, del 0.0025% daba como correcto un volumen de muestra de 41 individuos. Un cifra tan baja que no permitía la aplicación de todos los porcentajes resultantes de la estratificación por cuotas y eliminaba, así, estratos representativos cuyos datos resultaron relevantes en sí mismos o por comparación con otros. No significa esto que el sistema porcentual de fijación del tamaño de la muestra deba ser desestimado en todos los casos, sino que, antes de su utilización, deberá tenerse en cuenta si la representatividad de la muestra puede quedar comprometida.

Más adecuada, por tanto, y aunque tampoco exenta de críticas, parece la segunda opción, que establece los límites de la representatividad en una cifra concreta: 150 hablantes, independientemente del volumen real de la población y de la complejidad de su constitución.

La ventaja de este método radica en el hecho de que el límite de hablantes preciso viene dado de antemano y es suficientemente alto como para permitir el establecimiento de las cuotas muestrales sin complicaciones. El que el tamaño de la muestra sea idéntico para distintas comunidades —piénsese en Valladolid, Burgos, León— no elimina la posibilidad de descubrir en cada una de ellas y entre ellas —en estudios comparativos— rasgos coincidentes, discriminatorios o diferenciales, puesto que la constitución socioeconómica, cultural y biológica es propia en cada una de las comunidades de habla, y se reproduce de manera particular a través de las cuotas de afijación proporcional, es decir, de los porcentajes que cada una de las variables representa con relación a la población total.

Sin embargo, y como antes apuntábamos, la imposibilidad de contar con un argumento irrefutable a propósito del número exacto de informantes necesarios permite que el tamaño final de la muestra sea ligeramente superior o inferior a esa cifra. En la investigación sobre las formas verbales personales en el habla de Burgos, se entrevistó a 160 hablantes, (+10 con relación a los fijados por Sankoff) porque, de otro modo, la información aportada por los estratos marginales se perdía. Se opera

con aproximaciones, y del buen criterio del sociolingüista y de sus conocimientos previos sobre la comunidad cuya habla pretende estudiar se espera que establezca el tamaño de su muestra acercándose lo más posible a ese «número idóneo» de individuos del que hablaba Francisco Moreno.

UNIVERSO RELATIVO Y ABSOLUTO. DIFICULTADES METODOLÓGICAS

Por otro lado, todas las precauciones a la hora de determinar el número de hablantes que componen la muestra serían estériles si los propios hablantes no fueran representativos de la población de la que se les ha extraído. Pero los problemas sobre la representatividad cualitativa de los sujetos encuestados no aparecen exclusivamente vinculados a la técnica del muestreo, sino además, y antes incluso que a ella, a la distinción que estableció López Morales (1994: 41) entre todos los censados, *universo absoluto*, y aquellos que verdaderamente pueden considerarse candidatos válidos para la investigación sociolingüística, *universo relativo*, por ser representativos de la comunidad cuya habla se estudia y por cumplir las especificaciones fijadas por el sociolingüista de acuerdo con los fines de su investigación. Una distinción que es muy pertinente, pero que, en la realidad, es prácticamente imposible de llevar a cabo, al menos en lo que se refiere a estudios de variables lingüísticas en comunidades urbanas complejas.

El censo, que es la única forma de acceder al conocimiento objetivo del volumen de la comunidad sobre la que se trabaja, esconde tras su cifras una realidad compleja que no siempre es posible de advertir, y de las que resulta casi imposible restar, matemáticamente, ese número de hablantes que no se ajustan a los requisitos sociolingüísticos que consideramos básicos; aunque sepamos, ciertamente, que existen. En él, se incluyen tanto los hablantes nacidos y residentes en la comunidad un número suficiente de años, como los emigrantes o la inmigración rural, así como nacidos en la comunidad, pero con periodos de ausencia en años claves para la adquisición del habla característica de la zona investigada; y residentes en la ciudad pero trasladados de otras zonas por motivos laborales, de estudios o de otra índole, que llegan a la comunidad de habla que se estudia con hábitos lingüísticos adquiridos en otra. Todos ellos están empadronados y aportan sus datos al censo. El único caso en el que es posible descartar a un grupo de hablantes como miembros de la comunidad no representativos es el de la población que reside en espacios extraurbanos, que aportan sus datos al censo y que en todo momento aparecen identificados bajo esta característica, como población no perteneciente al casco urbano. En este caso el criterio administrativo es también sociolingüístico.

En nuestro estudio sobre las formas verbales personales en Burgos, excluimos a estos hablantes. Aunque no se deben perder de vista comportamientos demoesociales que pueden hacer considerar en un momento determinado también a estos hablantes

como parte de los candidatos. En la actualidad, y no muchos años después del momento en el que se hizo la investigación a la que nos referimos, algunos de estos espacios extraurbanos se han convertido en zonas industriales y otros se están uniendo a Burgos, físicamente incluso, por un factor económico, como es el precio de la vivienda. La distancia a la que estos espacios se encuentran de la ciudad abarata el precio del suelo edificable y los cambios urbanísticos que se están produciendo en esas zonas son profundos. Pero como se trata de una población que trabaja, se abastece y ocupa su tiempo de ocio fundamentalmente en la ciudad, la línea de autobuses de Burgos los incluye en sus recorridos, y lo hará cada vez con más frecuencia, siguiendo el ritmo de poblamiento. Con probabilidad, y sin que medie demasiado tiempo, se puedan llegar a considerar como barrios de Burgos. En cualquier caso, en algunos de estos espacios, se funde la población nacida y residente en la ciudad hasta una edad sociológicamente revelante y la población originaria de la zona, que como Villagonzalo y Villatoro eran pueblos. Todo lo dicho puede ilustrar la complejidad con la que se enfrenta el sociolingüista al decidir cuáles son los candidatos y cuáles no.

Queda claro, pues, que el censo establece un límite de aproximación al conocimiento de la realidad sociológica de la comunidad de habla que el investigador no puede burlar. Es imposible saber en qué zona de la ciudad, sexo, nivel o grupo se localizan los candidatos descartables: emigrantes, población rural y población que no responde a las especificaciones establecidas por el investigador. Por tanto, creemos que la discusión de los sociolingüistas sobre cuáles son los criterios más productivos de exclusión y cuál es el límite máximo de exclusión de candidatos permisible o conveniente², admite algunas críticas al menos en lo que se refiere a los estudios de habla en una comunidad urbana compleja. Y que, asimismo, la distinción entre universo absoluto y relativo es pertinente, indudablemente, pero

² Moreno (1990: 80) daba solución al conflicto señalando un 10% como máximo de individuos que a priori podían quedar fuera del universo del que se iba a extraer la muestra. Sin embargo, Samper (1988: 33), en su estudio sobre el municipio de las Palmas de Gran Canaria, excluyó a 143.849 de un total de 360.099, casi un 40%. El criterio fue la edad, puesto que el autor no consideró como candidatos adecuados a los menores de 20 años. Criterio de exclusión éste usado en numerosas ocasiones por varios investigadores. López Morales (1983: 24-25), en el estudio sobre la estratificación del español de Puerto Rico, eliminó como candidatos a los menores de 20 años juntamente con los extranjeros. Teóricamente es intachable, desde luego; pero, en la práctica al trasladar el modelo a la tesis doctoral sobre las formas verbales personales en el habla de Burgos, Larrosa (1998) surgieron numerosos interrogantes. Así no consideramos como candidatos adecuados a los menores de 14 años por cuestiones metodológicas, por presentar rasgos jergales y por estar en periodo de adquisición de hábitos lingüísticos y porque, indiscutiblemente es preferible para la investigación una muestra de habla más madura, pero sí se les consideró como parte del universo relativo puesto que también se incluía el resto de los habitantes entre los que se sabía que había candidatos no adecuados. En nuestra opinión, no había razón para dar dos tratamientos distintos inclusión / exclusión simultáneamente al mismo problema.

no se puede distinguir el segundo del primero basándose exclusivamente en los datos aportados por el censo.

Bien es cierto que en la entrevista se recogen datos que le indican al investigador la adecuación del sujeto a los propósitos de su investigación, y habrá que descartar candidatos incluso ante la mínima duda de que puedan comprometer la representatividad de la muestra, pero para cuando llega a ese punto, el candidato rechazable ya ha sido *aceptado* en el *universo relativo* al poderse guiar el investigador únicamente por los datos que le ofrece el padrón de habitantes.

Partiendo, por tanto, del hecho de que no hay posibilidad de aislar a los candidatos adecuados de los que no lo son, habrá que considerar los porcentajes de las distintas variables, proporcionados por el censo, como porcentajes válidos y aplicables al tamaño de la muestra para poder establecer así las cuotas muestrales. Técnicas de muestreo³ hay muchas, pero no todas son válidas para los muchos objetivos a los que pueden apuntar las distintas investigaciones. Y es claro que si se busca la representatividad habrá que tener en cuenta que la población está constituida por hombres y mujeres, *variable de sexo*, que tienen distintas edades, *variable de edad*, cuya formación intelectual es diversa, *grado de instrucción*, que desempeñan diferentes actividades laborales remuneradas *parámetro de ingresos*, y que residen en determinadas zonas de la ciudad, *variable de zona*.

Dado que existen todas estas variables demosociales, la única técnica válida para seleccionar a los hablantes que formarán parte de la muestra será la que reproduzca fielmente las características de la población en estos distintos aspectos. Tanto cualitativamente, que queden reflejadas todas esas variables, como cuantitativamente, que mantengan en la muestra la proporción exacta que supone cada una de ellas con relación al total de la población. El censo da información estadística de cuatro de estas variables.

Esta exigencia elimina cualquier técnica de muestreo que no sea la de cuotas de afijación proporcional. Sería impensable que los 150 hablantes, o los 160 de nuestro estudio, se obtuvieran de la suma de todos aquellos que accedieran a la entrevista, y conseguidos, por ejemplo, de una guía telefónica, *los 10 primeros de la cada letra*, encuestándolos por la calle, entre los amigos, recorriendo todos los talleres de mecánica de una ciudad, o en un club de fútbol sin considerar en absoluto su caracterización sociolingüística. Si se quiere que el estudio sociolingüístico sea relevante se debe ser riguroso en todos los aspectos metodológicos, y en éste, muy principalmente.

La aplicación de esta técnica de muestreo es, en esencia, muy simple. En primer lugar, el sociolingüista tiene que establecer las cuotas de instrucción. El método más

³ Tanto López Morales (1994) como Moreno (1990) cuentan con capítulos dedicados a las distintas técnicas del muestreo.

rentable es el que ordena las cuotas muestrales de lo más general a lo más particular: Variable de Zona, Variable de sexo, Variable de edad y Variable de Nivel sociolingüístico.

VARIABLE DE ZONA

Lo primero que debe hacer, pues, es zonificar el plano de la comunidad de habla, incluso aunque la investigación desestime los resultados aportados por ella. La razón es evidente. El sociolingüista debe realizar las encuestas teniendo en cuenta la distribución real de la población en el espacio urbano y no escoger arbitrariamente los puntos de encuesta, ni en cuanto a la localización del foco en el plano ni en cuanto al número de los mismos, por razones de comodidad.

El primer paso para el establecimiento de zonas es contar con unos criterios de segmentación de plano. La observación del mismo y nociones de geografía urbana de la localidad son puntos de referencia claves para establecer un criterio de partición correcto.

Existen líneas naturales –zonas de bosque, ríos, fronteras de otro tipo– y artificiales –desarrollos o movimientos urbanísticos concretos– que delimitan la ciudad y la agrupan bajo determinadas características. Así, pues, se puede hablar de un *criterio de frontera* y un *criterio histórico y de desarrollo socioeconómico afin* (Larrosa Barbero 2000). El primero haría referencia a los límites y obstáculos tanto naturales –ríos, puntos cardinales, zonas arboladas– como urbanísticos, trazados de carreteras, espacios edificados...; y el segundo, a las zonas de la ciudad que surgen en diferentes momentos del crecimiento urbano como consecuencia de la acción de factores externos y cuya impronta queda registrada en ellas frente al resto de las áreas urbanas. Sin olvidar que existen sectores dentro de la ciudad que presentan rasgos muy similares desde el punto de vista socioeconómico y que desaconsejan su división en áreas distintas o más pequeñas.

La aplicación de estos criterios dio como resultado en el estudio del habla de Burgos la zonificación de la ciudad en cinco áreas. El río Arlanzón dividió la ciudad horizontalmente en dos zonas: norte y sur. Este corte resultaba ya rentable puesto que la población se agrupa principalmente al norte del río. Por otro lado, la Autovía de Ronda y la carretera de la Avda de Cantabria segmentaban verticalmente la ciudad. Estas dos fronteras aislaban –y muy claramente mediante la simple observación directa del plano– la zona este de la ciudad. La segmentación era operativa puesto que en la zona noreste se localiza el barrio de Gamonal y Capiscol, cuya unión, pese a ser dos barrios distintos, quedaba plenamente justificada porque ambas poblaciones están fundidas y presentan rasgos económicos y culturales extraordinariamente similares. Así esta zona de la ciudad constituye un claro exponente de lo que nosotros llamamos *criterio histórico y de desarrollo*

socioeconómico afin. El barrio de Gamonal surgió y se desarrolló como consecuencia de la creación del Polo de Promoción Industrial, atrayendo por las expectativas de trabajo a numerosa población de aluvión (*criterio histórico*) y la población de Capiscol se integra perfectamente en la zona sociolingüísticamente delimitada por el *criterio de desarrollo económico afin*.

Por otro lado, otras fronteras naturales, la Plaza de Castilla y la Plaza de la Estación de RENFE, justificaban el segundo corte vertical, que aislaba la zona oeste de la zona centro. Una división que posteriormente se reveló como muy provechosa dado que en la zona oeste se registraba, tanto al norte como al sur del río, muy escasa población, de escaso nivel de instrucción y bajos ingresos; lo que le diferenciaba significativamente de la zona centro, de mayor población y con representación de todos los niveles socioculturales⁴.

Finalmente, en la zona centro habría que distinguir centro norte, el casco antiguo con los monumentos históricos más emblemáticos de la ciudad, y el centro sur, donde se encuentra uno de los puntos más antiguos de asentamiento de la población. Una vez establecida la zonificación más adecuada, la operación siguiente consiste en saber qué calles de la ciudad –compuestas por distritos, secciones y manzanas– pertenecen a las áreas establecidas. Para ello, además del censo, se deben manejar otros documentos: el *callejero* y el *manzanero*. Mucho más rentable para el sociolingüista es este último⁵ puesto que consta de la relación completa de las calles de la ciudad especificando entre paréntesis el sector al que administrativamente se adscribe, y señalando el distrito, la sección y manzana a la que pertenece la calle. Una vez identificadas las zonas con las manzanas, secciones y distritos que las componen, el sociolingüista deberá efectuar sus operaciones con el censo.

En este documento, el censo, aparece la población de derecho agrupada por distritos, secciones y manzanas desglosada en intervalos de edad y en varones y mujeres. De esta manera, el investigador puede, manejando un único documento, conocer la composición de un área según las variables de sexo y edad. El grado de instrucción lo establecerá, siguiendo el mismo mecanismo, pero con otro documento,

⁴ Precisamente estas áreas de baja densidad de población, a cierta distancia del centro histórico, están siendo objeto en el presente de una intensa urbanización que puede en el transcurso de unos años cambiar su perfil sociolingüístico actual, al recoger, por motivos de vivienda, población de niveles socioculturales diferentes, y en cualquier caso, superiores al mayoritario. Lo que da una idea de la atención que el sociolingüista debe prestar a la *historia* de la zona para poder distinguir su población representativa.

⁵ El callejero, de hecho, es una relación de las calles de la ciudad. Su importancia radica en que permite ver qué calles pertenecen a la ciudad y cuáles están localizadas en espacios extrarurbanos. Es importante si el investigador contempla la distinción entre *universo absoluto* y *relativo* y decide que estas áreas que no pertenecen administrativamente al casco urbano, tampoco deben ser tenidas en cuenta en el universo relativo.

el *Padrón de estudios* que incluye igualmente los niveles educativos de la población por distrito, sección y manzana, y desglosados en varones y mujeres.

VARIABLES BIOLÓGICAS: SEXO Y EDAD

Como acabamos de señalar, en el censo aparece información sobre las variables biológicas, sexo y edad. La división simple, dicotómica, de la población en hombres y mujeres ahorra al sociolingüista una operación complicada e ineludible en el resto de las variables: la conversión de modalidades en clases, es decir la organización de los distintos y numerosos intervalos en los que aparece dividida la variable en el censo en grupos metodológicamente operativos y expresivos. Tal y como aparecen no tienen rendimiento para la sociolingüística.

En el caso concreto de la Edad, esta característica se fragmenta en veintiuna modalidades que surgen de la división de la variable biológica en intervalos de cinco años: desde el nacimiento hasta los 100 años. El criterio de agrupación, por otra parte, no está exento de dificultades y será el sociolingüista el que considere, entre las distintas posibilidades, la que más se adecue a los objetivos de su investigación. En el estudio sobre el habla de Burgos, volvimos a utilizar tres tramos generacionales, por el rendimiento que habían probado ya en una investigación anterior sobre el habla del barrio de Gamonal (Larrosa Barbero 1993), barrio de Burgos: G1, hasta los 29 años, G2, de los 30 a los 59; y, finalmente, G3, de los 60 en adelante.

El límite de los 29 años quedaba justificado porque para esa edad el hablante ya había podido adquirir incluso formación académica superior completa y podía haber accedido a algún puesto de trabajo que permitiera caracterizarle por su instrucción y categoría laboral propia y no por la de sus padres, por ejemplo. El segundo grupo generacional absorbía el periodo profesional; y el tercer tramo de edad contemplaba a los hablantes cercanos a la jubilación o jubilados por distintas causas. Esta división no era arbitraria, aunque sea discutible en algunos puntos, porque atendía no sólo al hablante sino a la interacción lingüística del mismo con otros miembros de la comunidad de habla. Es evidente que los hablantes del primer grupo tienen más posibilidades de estar en contacto con un patrón de habla académico, estándar y superior al suyo que los informantes de los tramos de edad siguientes.

VARIABLE SOCIOLINGÜÍSTICA DE NIVEL

En el caso concreto de la variable de Nivel, las dificultades que se presentan al investigador no sólo tienen que ver con los criterios de agrupación de características en clases, sino que van más allá. Esta variable se define a partir de la relación de

varios parámetros base. El documento de Padrón de habitantes. Estudios no es el único documento que debe manejar, aunque sea imprescindible y, por añadidura, el más sencillo de ellos. El grado de instrucción debe completarse con información relativa a la categoría profesional en la que el hablante puede incluirse y datos respecto a sus ingresos económicos. Aunque hay autores que incluyen más parámetros, nosotros consideramos indispensables y suficientes estos tres.

Censo de Estudios. Grado de instrucción del hablante

Respecto al primer documento sobre el grado de instrucción de los hablantes, el censo proporciona en un documento específico datos sobre la formación académica de la población, desglosando los niveles en varones y mujeres y señalando en qué distrito, sección y manzana se encuentran los hablantes con ese grado concreto de formación académica.

Como ocurría en el caso de la variable de Edad, se impone, en lo que se refiere al padrón por Estudios, una reducción de las 15 modalidades⁶ a clases que resulten verdaderamente operativas. En este, como en el anterior, los problemas metodológicos se centran en los criterios de corte.

En nuestra investigación agrupamos en un grado de instrucción inferior las 5 primeras modalidades, desde No sabe leer ni escribir hasta FP 1 Grado; en un nivel de instrucción medio, las tres siguientes, desde FP 2º Grado hasta la marcada con la referencia 08, Otras titulaciones medias, y, finalmente la 3a, agruparía las restantes, que incluirían las diplomaturas, licenciaturas, doctorados y otras titulaciones de postgraduado.

El criterio de división se apoya en dos puntos. En lo que nosotros llamamos nivel límite del grupo, primero, e interacción lingüística del grupo, segundo. Cada uno de los niveles límite marca un cambio importante en el grado de instrucción del individuo; y, por lo que respecta a lo segundo, es más fácil que el individuo que alcance el nivel más alto de su grupo se relacione, interactúe lingüísticamente, con sujetos iguales a él en formación o de nivel inferior que con hablantes de niveles superiores; y que los que no han alcanzado ese nivel de instrucción límite de su propio grupo se relacionen en su vida cotidiana, como máximo, con individuos que, dentro de su propio grupo, sí lo hayan alcanzado. Parámetro de ingresos y categoría profesional

⁶ 1.- No sabe leer ni escribir/ 2.-Sin estudios./ 3.-Estudios primarios/ 4.-Bachiller elemental, Graduado Escolar o equivalentes/ 5.-FP. 1º grado/ 6.-FP 2º grado/ 7.-BUP o Bachiller/ 8.-Otras titulaciones medias/ 9.-Arquitecto, Ingeniero Técnico/ 10.-Diplomado de Escuela Técnica Universitaria y similares/- 11.-Arquitecto e Ingeniero superior/ 12- Licenciado/ 13.-Estudios superiores no universitarios/ 14.-Doctorado/ 15.- Titulaciones de Posgraduado o especialización para todos.

Más complicado resulta establecer el parámetro de ingresos. El sociolingüista debe partir del conocimiento sobre los salarios en España en general y en la comunidad de habla sobre la que centra su estudio. La información más útil a este respecto procede de los convenios colectivos de los distintos sectores profesionales. Convenios que no sólo ofrecen una aproximación a los salarios reales de la localidad sino que permiten al sociolingüista el conocimiento de la división del mundo laboral por categorías con la remuneración económica que por convenio corresponde a su nivel. Siendo esto así, las organizaciones o instituciones que pueden proporcionar estos datos al investigador son los sindicatos y de la Oficina de Información Socio-laboral del Ministerio de Trabajo y Seguridad Social.

Con toda esta información debe establecerse un parámetro de ingresos de seis puntos aproximadamente⁷. Fijar el extremo inferior de este parámetro no es muy complicado. Basta con saber la cuantía económica establecida para las pensiones no retributivas, parados y el salario mínimo interprofesional⁸. En cambio, mayor dificultad presenta establecer el punto máximo. Será el investigador el que fije la cuantía económica, según la información de los convenios para las categorías profesionales superiores. En el estudio realizado por Samper en el 1988, el punto 6º del parámetro aparecía como *Más de 150000 pesetas*. En nuestro estudio, y basándonos en la documentación obtenida, consideramos que era razonable seguir manteniéndolo. Y así, nosotros establecimos el siguiente parámetro: 1.- Hasta 59.999. 2.-De 60.000 -79.999. 3.-De 80.000- 99.999. 4.- De 100.000-119.999. 5.-De 120.000-149.999. 6.-Más de 150.000.

Pero tampoco habría ningún motivo para que, dependiendo de la fecha de la investigación y de la localidad, se considerara como *Más de 200.000* una cantidad aceptable. Por lo que respecta a los puntos intermedios se establecen de acuerdo con los datos de los convenios y en intervalos aproximados de 200.000 pesetas. Aunque como venimos señalando el sociolingüista es quien tiene la última palabra a la hora de establecer sus intervalos y deberá estimar cuidadosamente si resulta adecuado mantener los establecidos por otros lingüistas o sería conveniente introducir modificaciones que aseguren con mayor probabilidad la representatividad.

Por su parte, las categorías profesionales que se especifican en los convenios colectivos de los distintos sectores son realmente numerosas y muy bien puede encontrarse el investigador que un mismo puesto de trabajo aparece en dos categorías diferentes dependiendo del sector industrial. Para eliminar esas

⁷ Las ventajas de operar con un parámetro son considerables. Los informantes suelen mostrarse reticentes a declarar sus ingresos por un cierto complejo de inferioridad ante el investigador al que se presuponen superior en aspectos culturales, sociales y económicos, actitud que suele ser más acentuada en los hablantes de nivel sociolingüístico inferior.

⁸ En la fecha de la investigación sobre las formas verbales personales en el habla de Burgos, era de 60.570 pesetas.

vacilaciones y operar cómodamente con las categorías laborales, redujimos en nuestra investigación todas las modalidades a cuatro clases: 1.- Subalternos y obreros sin cualificar. 2.-Obreros cualificados y administrativos. 3.- Titulaciones medias y pequeños y medianos empresarios. 4.- Licenciados, profesiones liberales de titulación superior y grandes empresarios.

Así de la unión del parámetro de ingresos y de las categorías profesionales surgía una escala que nos permitía adscribir al hablante a un nivel económico. Los subalternos y obreros sin cualificar, más pensiones, se asociaban a los puntos 1 y 2 en la escala de ingresos. Los obreros cualificados, a los puntos 3 y 4. Las titulaciones medias, pequeños y medianos empresarios, al punto 5; y, finalmente, los licenciados y grandes empresarios, al punto 6.

Cuadro 1: Escalas correspondientes a nivel, ingresos y categoría profesional

| Nivel de instrucción 3 | | Ingresos mensuales 5 | | Categoría profesional 6 | |
|------------------------|---|----------------------|---|-------------------------|---|
| Escala | 1 | Escala | 1 | Escala | 1 |
| | 2 | | 2 | | 2 |
| | 3 | | 3 | | 3 |
| | | | 4 | | 4 |
| | | | 5 | | |
| | | | 6 | | |

Realizadas estas operaciones, se está en condiciones de determinar a qué nivel sociolingüístico pertenece el hablante. En sociología cada uno de estos parámetros recibe una puntuación que permite hallar el nivel a través de la suma de los puntos dados. Aunque la suma puede ser o no ponderada, desde un punto de vista técnico, es evidente que en el estudio de comunidades no todos los factores tienen el mismo valor discriminatorio en el establecimiento de una estratificación. Y habrá que resaltar este hecho de una manera matemática. Por eso, Samper en su investigación sobre Las Palmas, H. López Morales (1994: 62), tal y como recoge en su obra, y nosotros mismos empleamos la suma ponderada estableciendo las constantes de 3 para el grado de instrucción, 5, para el nivel de ingresos y 6 para las categorías profesionales. Así que el mínimo de puntuación sería de 14 y se trataría de un hablante con grado de instrucción bajo, obrero sin cualificar y de ingresos no superiores a las 79.999 pesetas; y el máximo de puntuación, 63, correspondería a un hablante de grado de instrucción alto, de actividad laboral acorde con su titulación y con unos ingresos superiores a 150.000 pesetas al mes.

| | | | | | |
|----------------|---|---|---|----------------|--|
| Nivel 1 | 1 | 1 | 1 | 3, 5, 6 = 14 | El nivel 1 está formado por los hablantes cuyas puntuaciones oscilan entre 14 y 29 puntos. |
| | 1 | 2 | 2 | 3, 10, 12 = 25 | |
| | 2 | 1 | 1 | 6, 10, 12 = 17 | |
| | 2 | 2 | 2 | 6, 10, 12 = 28 | |
| Nivel 2 | 2 | 3 | 2 | 6, 15, 12 = 30 | El nivel 2 está formado por hablantes cuyas puntuaciones se encuentran entre 30 y 47 puntos. |
| | 3 | 4 | 3 | 9, 20, 18 = 47 | |
| Nivel 3 | 3 | 5 | 3 | 9, 25, 18 = 52 | Al nivel 3 pertenecen aquellos cuyas puntuaciones se encuentran entre 48 y 63 puntos. |
| | 3 | 6 | 4 | 9, 30, 24 = 63 | |

Con todos estos datos ya disponibles, nada hay más fácil que establecer las cuotas de instrucción. Se obtienen a partir de una *regla de tres*. Si en la zona A la población censada es de 65.188 hablantes y el tamaño de nuestra muestra, 160, supone el 0.001 % de la población total, este porcentaje se aplica sobre el volumen de la zona y el resultado será el número de hablantes que se precisan en ese punto, 65. Como el 47,96% de los hablantes residentes en la zona son varones y el resto, mujeres, 52,03%, aplicamos estos porcentajes al total de los 65 hablantes que componen la cuota y así sabemos que necesitamos para nuestro estudio 31 hombres y 34 mujeres. El resto de las cuotas, edad y nivel sociolingüístico, se establecen de la misma manera.

Es importante por cuestiones metodológicas y para que el propio sociolingüista haga más fácil su labor, realizar una plantilla en la que aparezcan las diferentes cuotas de instrucción, los hablantes que las componen expresados numéricamente y representados mediante casillas que el investigador pueda tachar a medida que realiza las entrevistas y avanza en su investigación. Por lo general las advertencias de este tipo no aparecen en los tratados de sociolingüística más rigurosos y de consulta imprescindible. Parecen labores demasiado sencillas como para merecer unas líneas en manuales. Sin embargo la experiencia indica que los aspectos inicialmente más complicados de una investigación sociolingüística absorben la atención del investigador de tal modo que incluso olvida tareas tan simples, pero de tal importancia, como diseñar plantillas de este tipo antes de lanzarse a la búsqueda de informantes e ir señalando los hablantes de las cuotas a medida que se entrevistan.

Precisamente, la preocupación de dónde encontrar a los hablantes necesarios puede ser uno de esos aspectos complicados a los que hacíamos referencia. Es un capítulo éste, que, como el anterior, no se trata en estudios de metodología. Pero es de vital importancia.

Focos de localización de informantes

Teniendo en cuenta que se parte de población estratificada de acuerdo con unas determinadas características, el sistema de localización del hablante y la vía más rápida de acceso tendrá que ver fundamentalmente con domicilios sociales, lugares de reunión y de trabajo, que tienen un perfil sociolingüístico concreto. Este tipo de locales se distribuyen a lo largo de toda la geografía urbana, y resuelven de paso la localización de hablantes por zonas.

Los criterios de búsqueda son dos: lugares de trabajo y de ocio. Respecto al primero, las únicas precauciones son no violentar al trabajador de cara a sus superiores –no yendo a una hora punta de su trabajo, por ejemplo– para que coopere con tranquilidad y llevar un buen equipo de reportero con micrófono unidireccional para que se recoja su voz y no ruidos de fondo, que podrían invalidar la entrevista. Son productivos los bares cuando hay poca clientela, las floristerías, farmacias, librerías, porterías... que remiten a un nivel sociolingüístico concreto y varios tramos de edad. Igualmente fértiles son los sindicatos –distintas ramas profesionales y todos los niveles–, los institutos de educación secundaria, autoescuelas, asociaciones de mujeres, etc...

Respecto al segundo, el criterio de ocio, hay que tener en cuenta especialmente el perfil de la asociación a la que se une el hablante. Las peñas, las asociaciones de barrio, asociaciones de vecinos unen por lo general a hablantes de ambos sexos, principalmente de tramo de edad medio y de nivel sociocultural mayoritariamente medio-bajo. Hay asociaciones, por otra parte, de intereses culturales variados –teatrales, visitas a museos y exposiciones, recuperación de tradiciones...– que reúnen a miembros de nivel cultural-medio alto.

Las mujeres de nivel sociolingüístico inferior y edad media pueden encontrarse, específicamente, en los Centros de Acción Social del Ayuntamiento (C.E.A.S) y Centros de Educación Permanente de Adultos. Es curioso cómo en estos últimos la mayoría de las matrículas son de mujeres de entre 30 y 59 años con aspiraciones culturales por disponer de tiempo para ellas en el momento actual, por el deseo de equipararse culturalmente a su cónyuge en matrimonios en los que la diferencia de instrucción es grande... por el estímulo que ejercen sus hijos en edad escolar y razones tan variadas como interesantes sociológicamente.

Los hablantes del G3, mayores de 60 años, se encuentran rápidamente en Centros de Jubilados, de la Junta o bien de las Cajas de Ahorros. Otro punto productivo son las parroquias y si el nivel sociolingüístico es medio-alto, las universidades para la Tercera Edad resuelven al investigador esta cuota.

En numerosas ocasiones se ha hablado de la necesidad de un contacto para el éxito de la entrevista. Siendo necesario, no es imprescindible. Pero sí que es importante en el caso de los hablantes de mayor edad contar con el apoyo de la asistente social o trabajadores que conozcan bien a los mayores. Desgraciadamente,

no todos gozan de una salud mental perfecta, o tienen carencias y problemas que echan a perder la grabación. La falta de dentadura o el uso de dentaduras postizas mal adaptadas produce una distorsión en la voz que, aunque no parezca muy grave en el contexto de una conversación natural, sí lo es al escuchar la conversación grabada con los medios técnicos que se emplean. El asistente social presenta al investigador a los candidatos más adecuados, ahorrándose mucho tiempo y molestias para todos.

Instrumentos para la recogida de materiales, entrevistas cuestionarios y tests lingüísticos

Una vez elegidos los informantes, el siguiente paso es la elección y definición del instrumento a través del cual se van a obtener los datos para la realización del análisis sociolingüístico. La sociolingüística pone a disposición del investigador una serie de técnicas –entrevista, cuestionarios y tests lingüísticos– cuya utilización depende directamente de los fines de la investigación⁹.

Si lo que interesa es un fenómeno de tipo fonológico o fonético, conocer el léxico de un determinado grupo o actitudes ante fenómenos lingüísticos concretos o la seguridad de los hablantes ante usos particulares, lo pertinente será emplear una entrevista estructurada –*lectura, lista de palabras, pares mínimos, encuesta rápida*– puesto que éste será el método más rápido y fiable para sacar a la luz el fenómeno buscado. Si por el contrario, como fue nuestro caso, lo que se pretende es acceder al conocimiento de una actuación lingüística, por más que se seleccione un elemento de interés –formas verbales personales–, la única forma de conseguirlo será mediante una entrevista no estructurada de conversación no dirigida. Es decir, mediante una conversación con el informante en la que se suscite un tema de su interés y en el que la intervención del investigador sea mínima y funcional.

Sin embargo, una entrevista de estas características, no estructurada y no dirigida, no significa improvisada o no controlada, sino todo lo contrario. El grado de concentración que requiere por parte del investigador es muy considerable por la imprevisible respuesta del sujeto. La mera presencia física del investigador, su forma

⁹ La limitación en extensión de un artículo de estas características no nos permite hacer un análisis pormenorizado de los instrumentos para la recogida de materiales. De cualquier manera, los estudios más difundidos de esta disciplina cuentan con capítulos enteros dedicados a este aspecto de la sociolingüística. Humberto López Morales y Francisco Moreno, ambos autoridades de la sociolingüística y referente ineludible para cualquier investigador en este campo, tratan en las obras que venimos citando las distintas técnicas con una claridad y rigor a las que poco se puede añadir desde aquí. Nos vamos a centrar por ello en el análisis de la entrevista no estructurada de conversación no dirigida, que fue el método de obtención de datos en nuestro estudio, y en su rendimiento.

de presentarse y abordar al hablante, el modo en el que aquel suscita el tema de conversación, el mismo tema y circunstancias personales y laborales que rodean al candidato pueden determinar la aparición de un estilo *–formal, semiformal, informal–* que afecte directamente al objetivo de la investigación.

Si lo que se busca, y es lógico en este tipo de técnica, es la aparición de una actuación lingüística en estilo informal, la labor del entrevistador será crear un clima de comodidad psicológica para el hablante. Un reto en un contexto que no puede ser más artificial: se escoge a un hablante mientras trabaja u ocupa su ocio, se le extrae generalmente del grupo en el que se encuentra, se le requiere para una investigación sobre cuya naturaleza tampoco tiene una idea exacta, y se le *obliga* a mantener una conversación ante un desconocido y su grabadora.

Pese a ser de tanta importancia la actitud psicológica del entrevistador a la hora de presentarse ante el hablante y ser admitido por él, son pocas las referencias que en los manuales se hacen a este asunto. Y es un aspecto del que depende el éxito de la entrevista y la recogida de datos. El momento en el que el investigador se presenta ante su candidato es clave para su aceptación o rechazo.

Presentación del investigador

Las razones por las que un hablante acepta la colaboración o se niega a ella, o lo que es lo mismo, por las que acepta al entrevistador o le rechaza son tan numerosas como hablantes se incluyan en la investigación. Pero hay unas técnicas sencillas de aproximación a los sujetos encuestables que pueden ser de una valiosa ayuda para quien se enfrenta por vez primera con esta exigencia de la sociolingüística.

El investigador debe, por ello, estudiar muy bien el tipo de población a la que va a dirigirse y tratar de fundirse con ella en su aspecto externo y también lingüístico. Lo que se busca no es causar una impresión inolvidable en el hablante sino todo lo contrario, ser aceptado como un miembro más de su grupo social con naturalidad y rapidez. En este sentido, es preciso no revelar de manera inmediata al candidato la personalidad y objetivos del entrevistador, sino darle un tiempo para que se adapte a su presencia física, a su voz y a su forma de comunicarse. La aceptación o el rechazo es rápido y se advierte inmediatamente. Por eso, el investigador debe actuar también con rapidez para que el *principio de desigualdad* no eche a perder la entrevista.

El informante en el primer contacto establece su posición en la jerarquía. Si se siente en inferioridad de condiciones en comparación con el entrevistador desde el punto de vista social, profesional o personal *–más frecuente y más intenso en los estratos medio y bajo–* el peligro de perder la entrevista aumenta, y el riesgo de que aparezcan ultracorrecciones o el estilo no sea el buscado es muy elevado¹⁰. La

¹⁰ Algunos de nuestros candidatos, fundamentalmente mujeres, rechazaban colaborar en la investigación alegando como razones «yo hablo muy mal», o «luego se reirán de mí, de lo mal

conversación espontánea sólo se produce entre iguales, y es labor del sociolingüista *provocar* la igualdad, compensar el desequilibrio entre estratos y hacer que el hablante se sienta seguro en su presencia. En este sentido, sólo puede jugar con la igualdad personal. Comentar algunas anécdotas, mostrar interés por sus habilidades, reforzarle de manera oportuna y proporcionada y dejar que descubra que es él quien más sabe del tema del que se está hablando suele bastar para corregir esa desigualdad. La experiencia de nuestra investigación nos mostró que cuando se pedía a un hablante que mostrara el funcionamiento de algún aparato que utilizaba, o se le manifestaba el absoluto desconocimiento sobre su trabajo, se solidarizaba con él en las dificultades que se derivaban de él o se le reforzaba verbalmente, el hablante mostraba una satisfacción personal muy saludable, avanzaba mucho en la conversación y se hacía más espontánea y amigable. Además, este tipo de actitud con el hablante no sólo es *técnicamente adecuada* sino humanamente más agradable.

Por otra parte, la única actitud incorrecta no es sólo la del entrevistador prepotente, sino también la contraria. Es decir, que el investigador se presente ante su hablante solicitándole desesperadamente que acceda a colaborar con él. Funciona en contadas ocasiones y cuando ocurre la situación generada anula el valor de la entrevista en un alto grado de probabilidad. Si el informante accede por lástima va a falsear todos los datos que sean precisos para ayudar, y sus respuestas serán las que él sospecha que el investigador busca. Hechos que provocarían la adscripción incorrecta del hablante a un estrato. Se acepta y se atiende a quien emana tranquilidad y seguridad en sí mismo, aunque nunca avasalladora, y no al contrario. La sonrisa, el sentido del humor, el interés por el entorno son informaciones que el hablante recibe del investigador y que le predisponen favorablemente.

Por otro lado también hay que tener en cuenta la actitud del hablante. Aunque ningún análisis psicológico podría hacerse en el brevísimos tiempo de la entrevista y tampoco se cuente en la mayoría de los casos con conocimientos suficientes, sí que se deben tener en cuenta algunos signos que dan los hablantes. Las posturas corporales, los tonos de voz, la dirección de la vista, movimientos continuos de manos o piernas pueden poner al entrevistador en alerta sobre la timidez del sujeto al que se ha dirigido. Algo que puede resultar trascendental para el éxito de la entrevista. Si el investigador no consigue relajarle con su actitud tras unos momentos, tal vez deba plantearse intentarlo con otro candidato¹¹. Lo que no debe

que hablo...» Mucho de ellos consideraban que quien les hacía la entrevista se burlaría con sus colegas de su habla, e imaginaban un prestigio social y un mundo de relaciones y poder que no se correspondían en absoluto a la realidad.

¹¹ Un punto en el que siempre se insiste es en la corrección del entrevistador a la hora de dar por finalizada una entrevista. Se debe evitar en absoluto terminar la sesión de manera brusca que haga que el hablante se sienta «utilizado». Por delicadeza personal y por respeto al que ha sido nuestro colaborador la despedida debe ser también agradable y emplear en ella tanto tiempo como requiera la situación. En ocasiones, los hablantes entablan conversación gustosamente y no

hacerse en estos casos es saturar al hablante de preguntas que él contestará tan brevemente como pueda tratando de dejar el entrevistador el peso mayor de la conversación. Se busca una intervención lingüística del sujeto tan extensa como sea posible, no una secuencia de pregunta-respuesta.

La colaboración de los hablantes

Las reacciones de los hablantes ante el investigador y los motivos por lo que participa en la entrevista son múltiples e impredecibles. Identificación emocional con el sociolingüista¹², deseo de dar a conocer sus actividades, simpatía natural, atracción por lo nuevo y deseo de romper la rutina. Sin embargo, además de todos estos factores que predisponen a los hablantes a la colaboración, hay otros que pueden actuar como elementos disuasores, o cuyo efecto no se puede controlar de antemano. Son los que tienen que ver con los aspectos técnicos de la grabación.

Es evidente que se necesita una grabadora, *reportero*, como se la suele conocer, que registre la conversación, así como un micrófono unidireccional¹³ que permita recoger la voz del hablante aun en medios en los que hay ruidos de fondo, músicas y conversaciones diversas. Por lo general, los hablantes no aceptan de buen grado ser extraídos de los ambientes en los que se ha producido el contacto con el entrevistador. Esto le supone al sociolingüista enfrentarse y adaptarse a condiciones acústicas muy lejos del ideal de grabación. Los hablantes se localizan en asociaciones, peñas, *pubs*, establecimientos comerciales, gimnasios, parroquias y lugares diversos que hacen del micrófono unidireccional un recurso imprescindible¹⁴.

comprenden que el investigador considere suficiente una charla de diez minutos. la experiencia de un hablante, en este sentido, es decisiva a la hora de volver a colaborar en investigaciones de esta naturaleza que para los estudios de lingüística actual son fundamentales.

¹² Algunas informantes accedieron porque identificaban a quien les entrevistaba con un hijo estudiante y asociaban las dificultades de éste en sus trabajos académicos con las de aquel en el suyo. Y lo manifestó textualmente una informante. «te ayudo porque pienso en mi hijo, y me gustaría que si hiciera esto alguien le ayudara. Pero a mi no me gusta».

¹³ En la mayoría de los manuales sobre sociolingüística no se tratan estos aspectos. Sin embargo. la experiencia de mucho tiempo de investigación aconseja hacer algunas referencias al tema por suponer una parte fundamental de la metodología sociolingüística que tiene una repercusión directa en la obtención de los datos de la investigación.

¹⁴ Muchas entrevistas de nuestra investigación sobre el habla de Burgos se desarrollaron en lugares que tal vez un investigador que se enfrentara por primera vez a este tipo de trabajo en sociolingüística hubiera descartado por completo. Se grabó con éxito en salones de recreo, en pubs con música alta, en la cocina de un bar mientras la hablante preparaba tortillas de patata para los pinchos, con los ruidos lógicos de esta actividad, en asociaciones culturales con un número considerable de personas trabajando en una carroza... en una peña taurina con retransmisión de una corrida... etc., Hay que tener muy en cuenta que el hablante accede en un número revelante de ocasiones si no le supone un tiempo extra o salir del ambiente en el que se encuentra.

Ambos, micrófono y grabadora, como acabamos de señalar, tienen efectos en los hablantes que no se pueden prever. Algunos informantes se entusiasman ante la novedad de ser entrevistados y poder opinar sobre asuntos diversos. La presencia cotidiana de cámaras y medios de grabación distintos facilita la labor del sociolingüista en este sentido. En ocasiones, hasta el punto de que los hablantes cogían el micrófono y lo manejaban ellos mismos¹⁵. Sin embargo, a otros les intimida el sencillo equipo técnico del entrevistador. La forma de conseguir que el hablante se olvide del micrófono y la grabadora es que el propio investigador no les preste más atención que la necesaria y que no esté angustiosamente controlando si *el reportero* sigue funcionando o falta mucho o poco para que se acabe la cinta. Si el sociolingüista maneja con naturalidad sus medios técnicos, el hablante se sentirá más relajado.

Salvadas las dificultades técnicas, la realización de grabaciones en el mismo espacio en el que han sido encontrados los informantes es fundamental para que el hablante se comunique en el estilo más próximo al informal y surja una conversación espontánea y natural entre ambos. Extraer al hablante de su medio natural y llevarle a nuestro terreno puede, en el caso de que el hablante acepte, generar una situación de expectación o una impresión de examen que aleje su discurso del estilo informal que buscamos. Con todo, no se debe olvidar que la conciencia lingüística del hablante está siempre despierta. Un estilo absolutamente informal es una ilusión. Más ajustado sería hablar de estilo semiinformal, o con oscilaciones entre informal o formal. Precisamente, éste es uno de los capítulos que más polémica ha suscitado en lo que a instrumentos de investigación se refiere.

La aparición del estilo informal está vinculada a múltiples factores: tiempo de contacto, tema de conversación, actitud del entrevistador, personalidad del hablante, circunstancias contextuales, etc. Sobre lo primero, ha sido mucho lo que se ha escrito y de poca ayuda para el investigador desde el punto de vista meramente práctico. Milroy (1987:39) aseguraba que el estilo informal surgía a partir de la primera hora de la conversación. Afirmación que López Morales y Francisco Moreno rebatieron de una forma tan contundente como sencilla. Es cierto que a partir de un cierto tiempo, el hablante puede emplear el estilo muy espontáneo en su discurso, pero nada garantiza que eso ocurra ni que ocurra a partir de un tiempo establecido. Más importancia que este aspecto para la consecución de un estilo informal es la actitud de entrevistador y el tema que se suscite. Asunto a propósito del cual han surgido muy interesantes debates que sacan a la luz la importancia de los esquemas culturales de cada comunidad y la incidencia que éstos

¹⁵ Hecho éste que debe controlarse puesto que aunque empiezan dirigiéndose al micrófono terminan moviéndolo libremente y el micrófono, al seguir el movimiento de sus manos, recoge el sonido del punto al que lo dirige el hablante.

tienen al determinar la intensidad emocional de los temas suscitados en la entrevista. Labov (1983: 137) señalaba que el camino más corto a un estilo informal era la utilización de temas de fuerte tensión psíquica y emocional, como situaciones de peligro de muerte. Francisco Moreno ya manifestó sus reservas al respecto en la obras citadas y nosotros nos adherimos a su opinión. La implicación afectiva es la forma más rápida de lograr un estilo espontáneo, pero siendo esto verdad no lo es menos que en ocasiones la utilización de elementos temáticos excesivamente cargados de afectividad puede producir bloqueos en el hablante o el deseo de acabar la entrevista por considerar el tema propuesto como una invasión de la intimidad. No se debe olvidar que el investigador se encuentra ante un desconocido de cuya experiencia vital nada sabe y que es el mismo hablante el que permite el acceso al sociolingüista a su mundo afectivo. Será el informante el que marque de nuevo las pautas, el modo y ritmo de acercamiento.

La experiencia indica que el estilo más próximo al conversacional surge cuando el hablante se manifiesta a propósito de experiencias emocionales, positivas o negativas, pero no hay por qué forzar al hablante a ser demasiado «íntimo». El espacio en el que se contacta con el hablante, lugar de trabajo o de ocio, puede dar pie a conversaciones muy distendidas. Temas de actualidad que pueden ser ligeramente polémicos suelen bastar para activar un estilo informal. En este sentido, es importante que el investigador prepare de modo general las entrevistas informándose en los medios de comunicación o haciendo pequeñas investigaciones previas sobre la zona donde residen sus candidatos, fiestas del barrio, etc.. Es absolutamente necesario que el sociolingüista pueda seguir al hablante en la conversación, que pueda iniciar *temas de apertura*, sustituir los temas que se agotan por otros nuevos y relacionados con él, y aprovechar todas las posibilidades que se vayan presentando en la conversación. Técnicas a las que López Morales (1994: 78) da el nombre de *amplificadoras y relacionantes*, y que como el autor, con quien coincidimos plenamente decía, «podrán dar una idea precisa de cuánta atención habrá que prestar en todo momento a las palabras del sujeto».

Una atención que debe mantenerse en la entrevista alrededor de 10 o 15 minutos, tiempo suficiente para una muestra de actuación lingüística. Aunque haya autores y estudios que consideran el tiempo ideal de entrevista en 30 minutos, hay que tener en cuenta que el tiempo es uno de los aspectos que más influyen para que el hablante acepte o no la entrevista. Se aborda al sujeto en el trabajo o en su tiempo de descanso y la idea de una interrupción muy larga no favorece su colaboración. En el corpus sobre el habla de Burgos, hay ejemplos de todo tipo, de hablantes que al final de la entrevista empezaban a incomodarse por la duración de la conversación, y de informantes que pretendían seguir hablando –y que lo hicieron ya con el reportero apagado– porque les había agradado la situación y el tema de la charla. Pero en ambos casos y antes de aceptar, todos querían asegurarse de que su participación iba a ser breve.

Por otra parte, hay que tener presente al estructurar el tiempo que se va a requerir al hablante, si se va a manejar una única técnica de recogida de datos o varias. En nuestra investigación inicialmente preparamos tres: una muestra de actuación lingüística, un cuestionario que pretendía obtener datos para un posterior estudio comparativo sobre el empleo de las formas verbales personales en estilo formal *cuestionario* y en estilo informal *entrevista*; y, finalmente, un cuestionario de actitud lingüística. El volumen de los datos de la entrevista fue tal que pareció recomendable centrarse exclusivamente en el análisis de los datos que aportaba y dejar la información procedente de los cuestionarios para un momento posterior.

Aspectos técnicos de la grabación

En los casos en los que la investigación sociolingüística requiere el análisis de la actuación lingüística de un número considerable de sujetos y se trabaja con una limitación de tiempo y recursos económicos, es preferible no cometer errores en la grabación de las entrevistas ni en su posterior tratamiento previo al estudio en sí. Muchos de los aspectos que vamos a tratar en las siguientes líneas pueden causar extrañeza tanto por lo evidentes como por el hecho de incluirse en un texto de estas características. De hecho, son asuntos que no se tratan en los manuales al uso. Sin embargo, la importancia de no cometer errores en esta fase nos aconseja hacerlo.

Ya en momentos anteriores nos hemos referido al uso del *reportero* y del micrófono unidireccional, uno no profesional de pequeño tamaño es suficiente. El sociolingüista deberá asegurarse de que su grabadora tiene incorporado el sistema V.O.R. que permite grabar incluso si el nivel de voz es bajo.

Tan importante como lo anterior es llevar en el equipo baterías y cintas de grabación suficientes. Aunque el proceso de recopilación de datos es lento, nunca se sabe cuántos candidatos puede encontrar el investigador en una jornada ni si esos informantes *perdidos* por una falta lamentable de previsión son especialmente importantes o difíciles de localizar por sus características sociolingüísticas.

Aunque López Morales (1994: 91) considera una labor innecesaria y excesiva grabar en la propia cinta la identificación del hablante, nosotros sí lo utilizamos en nuestra investigación sobre el habla de Burgos, y lo hicimos por el elevado número de sujetos encuestables y con el mismo propósito con el que se hacen copias de seguridad en el tratamiento informático de textos. No es necesario hacerlo delante del hablante, en un momento al finalizar la entrevista y en lugar apartado puede realizarse, o bien, en el transcurso normal de la entrevista. El sociolingüista puede dirigirse al hablante por su nombre e incluir con habilidad ciertos datos personales del informante que luego le sirvan de ayuda en este sentido.

De la misma manera debe llevar una matriz o plantilla en la que aparezcan de forma clara las entrevistas que tiene que realizar. Una ordenación posible es la división de todos los candidatos en zonas. Cada una de las zonas resultantes se

dividirá a su vez en hombres y mujeres *variable de sexo* y la variable de sexo desglosada, se dividirá en filas *variable de edad*; y en columnas *variable de nivel*. El número de entrevistas correspondientes a cada cuota puede aparecer marcado con un cuadro que se tachará a medida que se vayan completando las cuotas. Así, el investigador tendrá en todo momento, y de una manera visual, una idea clara de cómo se va desarrollando su trabajo y cuántos hablantes debe aún encontrar.

Es muy conveniente, además, que el investigador diseñe una ficha de grabación, en la que identifique a su hablante. Esta ficha debe constar de cuatro apartados. El primero debe recoger información que permita identificar fácilmente al hablante por su perfil sociolingüístico, hechas las operaciones necesarias. Por ejemplo: A/N1/G1/M, que corresponde a una mujer, de hasta 30 años, de nivel sociocultural bajo y que reside en la zona A. Un segundo segmento de la ficha debería recoger el número de cinta¹⁶ en el que aparece su grabación, <V, qué número de voz es que le corresponde, 8, *segunda voz de la grabación*, y en qué cara de la cinta aparece grabada (NB). No está de más que se dé un nombre, aunque sea falso, a la grabación para que el sociolingüista pueda localizarla rápidamente entre sus entrevistas cuando sea necesario.

Seguidamente, en la tercera sección de la ficha, deben aparecer también con el sistema convencional de formulario (*Varón 1&1 Mujer 0*) las variables que se manejan en la investigación: sexo, edad y nivel sociocultural, además de un apartado distinto para la profesión, los ingresos, y el grado de instrucción, con las 15 modalidades especificadas. No debe olvidarse que el nivel sociolingüístico se halla a través de una suma ponderada y deberá conocerse exactamente en qué punto de la escala de los distintos parámetros se encuentra el hablante. Y finalmente se debe dejar un apartado para añadir información suplementaria que sirva al sociolingüista para recordar al hablante por el tema de la conversación, el lugar en el que se realizó o cualquier circunstancia que le ayude a identificar al sujeto. Las primeras entrevistas se recuerdan con claridad, pero a medida que se va desarrollando el trabajo, las dificultades para acordarse de ellas se dejan notar. Tener presentes aproximadamente 160 conversaciones es un esfuerzo de memoria muy considerable. Un ejemplo de ello podría ser el siguiente: NN 3/G 11M, Cinta 1, voz 4. cara A, CBV, clave lugar.

¹⁶ Un orden escrupuloso a la hora de realizar esta fase de la investigación es de capital importancia. Las cintas de grabación deben estar numeradas. En las librerías están a la venta pegatinas con números impresos que resultan de excelente ayuda en esta labor. El investigador simplemente debe hacerse con ellas y pegar en la cinta de grabación el número que corresponda. Este procedimiento tan simple es de una ayuda extraordinaria cuando el sociolingüista tiene una cuota de 160 hablantes, más de 33 hora de grabación y más de 22 cintas grabadas, como ocurrió en nuestra investigación. Todas las precauciones para no perder grabaciones y poder localizar rápidamente son pocas.

cafetería, tema, viaje a París y sorpresa por los precios desorbitados de los bolsos de mano.

Corpus y transliteración

Una vez efectuadas las entrevistas se inicia la fase de preparación del material para el análisis cuantitativo y cualitativo. Y ésta comienza por la transliteración del texto oral y su conversión, por tanto, en texto escrito. En esta fase se puede comprobar también si ha habido algún error en la fase de grabación y subsanarlo, bien mediante nuevas grabaciones, bien adscribiendo al hablante al estrato correcto si se descubre que ocultó información al cumplimentar sus datos, bien seleccionando el material grabado, si, como dice López Morales, los problemas sólo afectan a una parte razonablemente pequeña de la grabación, o eliminándolo si así lo aconseja el rigor científico.

La transliteración es un proceso lento que suele suponer un trabajo de aproximadamente 1, 30 para grabaciones de 10 minutos. Debe efectuarse con auriculares y resulta más cómodo y rápido manejar los *retrocesos* y *avances* del aparato. Por lo general, estas grabadoras son lo suficientemente duras como para permitir una manipulación de estos dispositivos tan intensa como a la que se somete al aparato durante la transliteración.

Es conveniente, por lo que decimos, que la primera transliteración sea manuscrita, de tal forma que el sociolingüista active los avances y los retrocesos del aparato con una mano y escriba el texto con la otra. Hacerlo directamente en el ordenador, supone, en contra de lo que cabe pensar, una pérdida de tiempo. Y por otra parte, en investigaciones de esta naturaleza en la que el material es tan valioso conviene no confiárselo exclusivamente a *la técnica*. Una copia manuscrita es seguridad para el investigador. Pero hay una segunda razón.

Este tipo de transliteraciones se realizan según un protocolo establecido. Hay varios, pero el que nosotros escogimos y que generalmente se ha utilizado en los estudios más notables de sociolingüística es el Protocolo de Caracas¹⁷

Sin embargo, el necesario tratamiento informático del corpus exige adaptaciones o modificaciones de algunos puntos del Protocolo a las rutinas informáticas y a los objetivos de la investigación. Convertir el texto oral en escrito, fijarlo según el Protocolo y adaptarlo a las exigencias del programa informático simultáneamente es tal vez una labor de una complejidad innecesaria, cuya única

¹⁷ No vamos a especificar los quince puntos del Protocolo por cuestiones de extensión de este artículo. Al investigador que accede por primera vez a estos estudios de corte sociolingüístico no le será difícil encontrarlo en los títulos que hemos citado aquí, y más concretamente en López Morales (1994: 99-100).

ventaja -ahorro de tiempo- es muy discutible. Ahorro de tiempo, en cambio, sí es diseñar un documento de transliteración, igual para todos los hablantes, en el que figure su identificación: el *nombre del hablante*, *número de entrevista* (10) *código sociolingüístico*, A / N1/ G1 /M *número de cinta en la que se encuentra la grabación*, (5) *cara*, A *posición* (5). Estas tareas de identificación pueden parecer reiterativas pero son del mayor interés dado el número de hablantes que se maneja y los diferentes puntos de vista desde donde se estudia cada documento. Tener una identificación precisa es fundamental para el estudio.

Tratamiento informático del corpus. Creación del programa informático

Actualmente, trabajar en sociolingüística y en estudios lingüísticos en general sería impensable sin contar con la informática. El extraordinario volumen de datos que se manejan en estudios de esta naturaleza, su forma de aparición, en el contexto de una actuación lingüística de la que se deben entresacar, y los objetivos con que nace la investigación hacen prácticamente imposible estudios como los que se hacían hace treinta años. Sin embargo, la aplicación de la informática a los estudios lingüísticos, aunque los haya hecho posibles y más rigurosos, no los ha hecho más sencillos. Son muchos los retos a los que se tiene que enfrentar el sociolingüista

En la actualidad, los programas informáticos se desarrollan *a medida* y a medida de las necesidades del investigador de acuerdo con los objetivos de su estudio, lo que supone que será el propio sociolingüista el que diseñe el programa que mejor se ajuste a sus intenciones. Un problema si no está familiarizado con la programación. En ese caso, deberá contar con la ayuda de un experto en este campo y ser lo más preciso posible acerca de lo que pretende investigar y cómo necesita que se presenten los datos en pantalla además de las operaciones estadísticas que se deben incorporar al programa informático a medida. En resumen, el programador ejecutará las órdenes que el investigador le transmita. Las conversaciones previas son extremadamente útiles y conviene que ambas partes se aseguren respecto a lo que se necesita y se puede conseguir. Si es cierto que un sociolingüista no tiene por qué tener conocimientos tan específicos de informática, también es verdad que los conocimientos de lingüística del programador informático se limitarán casi con toda seguridad a lo que recuerde de su formación académica. Así, pues, el sociolingüista deberá ser preciso y asegurarse de que el técnico sepa perfectamente qué se espera de él.

Una de las necesidades primeras del programa a medida es un sistema de captura de datos. El programa debe estar preparado para extraer del texto, que constituye la actuación lingüística del sujeto, el elemento lingüístico que sea objeto de la investigación. En nuestro caso, sobre el habla de Burgos, las formas verbales personales. Es evidente que la codificación debe ser manual. Ningún programa por

avanzado que sea puede distinguir formas verbales de otras categorías gramaticales en un texto. Así pues, conviene elegir unos signos del teclado <>¹⁸ que *atrapen* el elemento lingüístico que se estudiará. En nuestro caso, fueron las formas verbales personales, como ya se ha dicho. Es de la mayor importancia que no haya errores en los códigos de captura, puesto que tendría consecuencias fatales al aplicar el programa. En este sentido, las revisiones son imprescindibles. Dado que las transliteraciones se efectúan con un procesador de textos normal, la opción *Buscar* (<) y (>), ejecutadas por separado, facilita considerablemente esta tarea, y descubrirá todas las parejas anómalas, <X +, <X <, <X*s*, etc.

Esta transliteración con el código de captura se introdujo en un único archivo en el que los subdocumentos, es decir, las intervenciones de los hablantes, aparecen delimitadas por su definición o *cabecera*. Aspecto éste al que hay que prestar la máxima atención para evitar confusiones que serían igualmente fatales para el estudio. El programa da por finalizado un subdocumento, intervención del hablante, cuando encuentra la siguiente cabecera. Una cabecera incorrecta, por tanto, sería interpretada no como un nuevo hablante, aunque en realidad sí lo sea, sino como intervención del hablante anterior, con la consecuente, y completa, invalidación del estudio.

Cada uno de los subdocumentos llevó la misma *cabecera* para que el programa informático reconociera el principio de un nuevo subdocumento. La cabecera fue VOZ con la numeración correspondiente en el corpus. En una segunda línea se especificaba el código sociolingüístico, siguiendo el mismo orden, Z/Nx/Gx/S" y en una tercera línea las iniciales del hablante seguido de su nombre entre asteriscos, *sss*. Igualmente entre asteriscos incluíamos información referida a verbos, especialmente si su uso era incorrecto. Hay que tener en cuenta que cuando el programa capturaba el verbo lo aislaba de su contexto y era imposible saber si había sido utilizado por el hablante correcta o incorrectamente si no se añadía un comentario que posteriormente se pudiera eliminar al editar el texto transliterado limpio. Al programa informático se le dotó, pues, de una rutina por la que borraba toda la información entre asteriscos.

El programa informático, al que dimos el nombre de ALDRA, era una serie de utilidades de captura de datos en documentos de texto cuyo uso iba dirigido al cómputo de verbos, y estaba formado por un grupo de programas necesarios para su ejecución. Los datos se introducían por medio de un archivo de texto, y mediante el sistema de codificación al que nos hemos referido recuperaba las palabras marcadas y las guardaba junto a su definición, que corresponde a la identificación del hablante.

¹⁸ La elección del signo es arbitraria, pero no inmotivada. Ténganse en cuenta los signos que se emplean en el Protocolo de Caracas para transmitir distintas informaciones sobre la intervención del hablante.

Por otra parte, se diseñó el programa de manera que sólo computara como palabras las aparecidas tras la marca HABL:, lo que eliminaba del cómputo las intervenciones del investigador. Y así mismo se consiguió que tampoco contara como palabras las incluidas entre corchetes [xxx] pero que no las borrara de la edición de texto limpio, puesto que corresponden a convenciones sobre reproducción de sonidos y otros comentarios del Protocolo de Caracas. Así mismo se adaptó el punto 3 de este citado Protocolo puesto que la exigencia de colocar tres puntos precedidos y seguidos por espacio en blanco para indicar la prolongación de una palabra por vacilación del hablante o para indicar el corte de una idea y comienzo de otra suponía que se computara como palabra esos tres puntos seguidos. Nosotros eliminamos el espacio en blanco de manera que el programa sólo reconociera una palabra, como realmente era, y no dos.

El programa informático adaptado para este tipo de estudios debe presentar una pantalla inicial que permita, a través de los sistemas habituales de iconos de función, acceder al archivo de texto de la transliteración, *cargar estudio*, y trabajar cada uno de los subdocumentos de acuerdo con los objetivos de la investigación. Teniendo en cuenta que cada subdocumento puede presentar un número muy elevado de elementos de análisis, es imprescindible que el programa permita abandonar la sesión de trabajo en cualquier momento conservando los cambios. El subdocumento analizado se guardará a través de un sistema de carpetas que permita guardar la información de manera independiente, y que al ser abiertas por el investigador le indiquen en qué VOZ se abandonó el estudio. *Se dejó el estudio en la voz 160.*

El manejo de un volumen de datos tan grande y la importancia de los análisis efectuados aconseja ser extremadamente prudente al guardar el trabajo. Es aconsejable disponer de un juego de carpetas, en nuestro caso fueron cuatro, que se empleen según los sistemas de seguridad habituales, por ejemplo, el de días impares. En caso de pérdida o error, no se pierde el trabajo completo hasta ese día, sino sólo el del día en el que se produjo el fallo.

Una vez solucionados los problemas derivados de estas operaciones, el sociolingüista deberá diseñar el núcleo del programa que será el que contenga los criterios de clasificación del elemento lingüístico que estudia. En nuestro caso, se diseñó una pantalla en la que la que aparecían tres ventanas: una referente al verbo capturado y las otras dos de identificación, a qué VOZ pertenecía y qué número de verbo hacia en la intervención del hablante. La identificación de la voz es imprescindible, y el número de verbo es importante puesto que actúa como factor de orientación para el sociolingüista. Si la media de verbos empleados por hablante es de unos 250, el investigador puede saber cuánto tiempo necesita para concluir ese estudio. En el resto del área de la pantalla se distribuyeron los patrones de estudio o aspectos de definición del verbo. Cada uno de los bloques se llamó *l/po* y contenía a la vista los subtipos.

Es decir, Tipo, *forma personal forma no personl forma perifrástica*. Cada uno de los subtipos debe contar con un botón que se pueda pinchar según la opción que corresponda. Una precaución importantes es que los patrones de estudio estén contenidos en una pantalla única, y otra, que dado el elevadísimo número de veces que se deberán pinchar las distintas opciones, el sistema cuente con una clasificación por defecto, es decir, la que corresponda a la opción mayoritaria por frecuencia, si es que puede conocerse de antemano. En nuestro caso, después de unos cientos de clasificaciones se pudo saber que la mayoría eran formas personales, irregulares, de indicativo, en presente, de forma simple, aspecto neutro y en estructura enunciativa afirmativa. Resulta rentable hacer correcciones y depuraciones sucesivas del sistema en este sentido. La clasificación por defecto ayuda mucho al investigador cuando se tiene que hacer, como nosotros hicimos, más de dos millones de pulsaciones con el ratón y sólo en esta fase.

Vinculado a este hecho, también es importante que la pantalla cuente con un botón, *Anterior*, que permita ir al verbo anterior en caso de error o inseguridad en la clasificación. De la misma manera, y aunque es obvio, debe aparecer igualmente un icono que dé por válida la operación y guarde la clasificación en la memoria. Relacionado con esto, no está de más diseñar un sistema de corrección de datos, incluso después de haber terminado la operación de clasificación. Puesto que nuestro estudio suponía el manejo y la definición de un número elevadísimo de verbos, algo más de 37.000, nosotros sí lo hicimos, y le dimos el nombre de Aldra-Alcor. En nuestro caso, se pudo llevar a la práctica puesto que contábamos con cadenas incompatibles de definición para el verbo. Por poner unos ejemplos obvios, todo *imperfecto es imperfectivo*, *ninguna forma compuesta es aspectualmente neutra o imperfectiva etc*. Lamentablemente no es posible en el estudio de todos los elementos lingüísticos. Dependerá del investigador y de los objetivos de la investigación desarrollar o no este sistema de corrección, extraordinariamente útil en cualquier caso.

ALMUL

El programa completo debe contar además con otro que contenga rutinas para discriminar el elemento lingüístico seleccionado para la investigación por variables y parámetros de búsqueda de tipos o subtipos. Es decir, debe permitir estudiar la variable lingüística en relación con la variable de sexo, edad, nivel y zona de residencia. Así, el programa debe estar en condiciones de responder a necesidades de la investigación del tipo, *estudio del subjuntivo en varones, V, del grupo de edad más joven, G1, del nivel sociolingüístico superior, N3, de la zona centro norte, A*. Para poder ponerlo en relación con grupos distintos, bien de mujeres, bien de zonas distintas, o bien de edades diferentes.

A este efecto, la pantalla deberá diseñarse de tal modo que permita seleccionar cada una de las opciones de las variables y cada uno de los subtipos de clasificación de la variable lingüística. Es decir, que permita estudios tan pormenorizados como el estudio del futuro imperfecto en el A/N1/G1/M, o bien en un único hablante. Y deberá permitir también diseñar desde ahí la forma en la que se quiere ver presentada la tabla correspondiente: qué variable debe aparecer en filas y qué variable en columnas.

Por lo que acabamos de decir, se deduce que el programa informático que gestione estos datos debe contar también con un pequeño aparato estadístico relativo fundamentalmente a los conceptos de estadística que usa la sociolingüística: frecuencia absoluta, frecuencia relativa, varianza, desviación típica y prueba estadística de Ji-cuadrado.

En primer lugar el programa debe indicar qué número de elementos lingüísticos seleccionados, para nosotros verbos, ha contabilizado y qué porcentaje supone respecto a las palabras que componen la emisión del hablante. Esta operación debe hacerla también para cada variable y subtipo de variable. Si se pretende estudiar, por ejemplo, la incidencia del llamado condicional en los hablantes según su nivel sociolingüístico, deberá aparecer la variable nivel en filas, N1, N2, N3, Y a continuación el número de verbos que cumplen esa condición, 247, 37 y 21 respectivamente, además de la frecuencia relativa que suponen en comparación con los verbos empleados. Ambas informaciones son cruciales para poder determinar, mediante pruebas estadísticas, la relevancia del dato obtenido.

Y, por otro lado, estar en condiciones de generar tablas con información estadística. Es fundamental contar con una tabla en la que quede registrada la frecuencia absoluta de un elemento lingüístico, verbos, en todos y cada uno de los hablantes. Así contamos con una tabla, a la que llamamos, *Tonkens.men*, que facilitaba esta información. A partir de ella, hacia el estudio de la varianza y desviación típica.

Debe diseñarse teniendo en cuenta las listas de selección, que corresponderán a las variables desplegadas en sus correspondientes opciones, y un sistema de selección que active los tipos y subtipos desde los que se puede abordar el estudio de la variable lingüística elegida. En nuestro caso, cada uno de los aspectos morfológicos, *tiempo, modo, uso, conjugación etc.*, desde los que se puede estudiar el verbo. Este sistema permite hacer una selección por cada una de las listas, eligiendo la variable sin especificaciones de modalidad o con una de ellas, (variable de sexo, H, hombres). De esta manera se pueden hacer estudios de lo más general, sin especificación de variables (estudio del subjuntivo en todo Burgos,) a lo más particular (estudio del subjuntivo en varones, H, del tramo generacional más joven, G1, de nivel sociolingüístico superior, N3, en la zona centro de Burgos, A).

El sistema debe también contar con la posibilidad de generar tablas de datos de dos tipos. Una de ellas que analice la frecuencia absoluta del elemento lingüístico

en todos y cada uno de los hablantes, bien considerados en general, bien seleccionados por variables, y que calcule a partir de esos datos la media, la moda, la varianza y la desviación típica. Le dimos el nombre de T onkens.men. Y otra tabla, a la que llamamos Tabla ALMUL, que permita hallar las frecuencias absolutas y relativas de un elemento lingüístico en dos variables combinadas y presentadas en filas y columnas. Por ejemplo, incidencia del llamado imperativo según nivel, N1, N2, N3 en filas, y según la edad, G1, G2, G3 en columnas. Esta presentación nos permite interpretar los datos en filas, es decir, averiguar si existen diferencias entre los miembros de un mismo tramo generacional según su nivel sociocultural, o si la edad supone comportamientos lingüísticos significativos en el mismo nivel sociolingüístico y todas los posibles análisis que pueden realizarse a partir de los datos ofrecidos por esta tabla, que son muchos. Esta tabla resulta fundamental para poder realizar a partir de ella la prueba Ji-cuadrado, básica en sociolingüística y poder afirmar que el comportamiento lingüístico estudiado es relevante.

Como se puede comprobar, gran parte del programa informático está diseñado para realizar las funciones matemáticas necesarias para el desarrollo de pruebas estadísticas. En algunos casos, ofrece los resultados de dichas pruebas, como en el caso de media, moda, varianza y desviación, y en otros presenta los datos necesarios para la prueba pertinente, en nuestro caso, Ji-cuadrado.

ESTADÍSTICA DEL CORPUS

La importancia de la estadística en la lingüística y la sociolingüística, es, ciertamente hoy, indiscutible. Y lo es porque saca a la luz una información sobre los fenómenos lingüísticos que no son inmediatamente perceptibles. Es imposible conocer la incidencia de un modo o un tiempo verbal, o cualquiera que sea el elemento lingüístico escogido para su estudio, en el discurso de los hablantes, y por lo mismo, es temerario e irracional afirmar si se emplea con mucha o escasa intensidad y menos aún si su presencia es significativa y está asociada a algún factor externo. Sin embargo, la importancia de la estadística para la sociolingüística no es el mero registro numérico de datos, *estadística descriptiva*, sino la posibilidad de aplicar las conclusiones que se establecen a partir de ellos a conjuntos más amplios de datos: *estadística inferencia*. Esa es la verdadera utilidad de la estadística en la sociolingüística, una importancia que queda clara en palabras de Francisco Moreno, «se ha podido afrontar el estudio de habla de centenares de personas partiendo de los datos recogidos en sólo unos pocos, sin que el principio de la representatividad quedara gravemente en entredicho».

Abordar el tema de la estadística en sociolingüística en un artículo forzosamente limitado en extensión como es éste es prácticamente imposible. Por ello nos referiremos exclusivamente a las cuestiones más elementales con las que

debe enfrentarse el investigador al realizar un estudio de este tipo y que serán, simplemente, la introducción a un estudio posterior en bibliografía especializada.

Un concepto básico en la estadística, y que constituye la piedra angular sobre la que se sustenta la sociolingüística, es la *frecuencia*. Es decir, la repetición de un elemento lingüístico en una unidad limitada. Frecuencia que puede expresarse matemáticamente de dos maneras: frecuencia absoluta -número exacto de ocurrencias- y frecuencia relativa -porcentaje que ese número de ocurrencias supone con relación a un total. La sociolingüística maneja ambas y, de la combinación de las informaciones distintas que aportan, extrae sus conclusiones. Labor del sociolingüista es interpretar los valores expresados por una y otra con cautela. Máxime cuando en los estudios de habla de una comunidad lingüística se trabaja con estratos de volúmenes distintos, en los que se pueden registrar frecuencias absolutas muy dispares, que, sin embargo, representen frecuencias relativas apenas diferentes¹⁹.

Sin embargo, la información aportada por la frecuencia no es suficiente en un estudio de sociolingüística. Importa saber cuántas veces se repite por término medio un elemento lingüístico en hablantes de una determinada caracterización sociolingüística, pero aún más comprobar si estos factores demoesociales se asocian o no al empleo medio más o menos intenso del elemento lingüístico observado. Es decir, importa la media, y hasta tal punto que López Morales indica que es dentro de la estadística descriptiva el único parámetro de posición que verdaderamente se usa en sociolingüística²⁰. Y que importa y adquiere relevancia cuando se pone en comparación con otros promedios, porque, y en esto coincidimos plenamente con el autor antes citado, la media se emplea para establecer comparaciones. En la comparación es donde se encuentra o no la relevancia del dato²¹. Todo ello sin perder

¹⁹ En el análisis del pretérito perfecto compuesto según el nivel sociocultural del hablante y en las enunciativas afirmativas, comprobamos que en el estrato sociocultural inferior, N1, se registraron 925 ocurrencias de este tiempo verbal, mientras que en el N2, aparecieron 184, y en nivel superior, N3, 79. El análisis de los datos tal y como aparecen podría hacer sospechar al sociolingüista que a medida que aumenta el nivel del hablante descende drásticamente el uso de esta forma compuesta. Nada más lejos de la realidad. Las frecuencias relativas, y la posterior prueba estadística aplicada, relevaron algo completamente diferente: La frecuencia relativa en el N1 era de un 77,47%, en el N2, de un 78,63%, y en el N3 de un 83,15%. Luego el análisis correcto reveló que se estaba ante un fenómeno general en el que el nivel no actuaba como elemento discriminador dado que 8 de cada 10 de estos tiempos verbales se incluían en estructuras oracionales enunciativas afirmativas en todos los casos.

²⁰ Véase López Morales (1994: 163).

²¹ Llegar al conocimiento de que el futuro imperfecto de indicativo se emplea en una media de 3 ocasiones por hablante puede no ser muy expresivo si nos falta una media de referencia. En comparación con el presente es cuando alcanza relevancia, y así se puede afirmar que la localización de los hechos en el futuro es evitada intensamente por los hablantes. La interpretación sociolingüística posterior dará cuenta de estas diferencias. [Por la abstracción que

de vista que la media puede surgir tanto de valores homogéneos como de valores extremos. Una media de 5 puede obtenerse a partir de $5+5+5=15/3=5$, o bien a partir de $10+5+0=15/3=5$. Este hecho tiene para la sociolingüística una gran importancia puesto que permite hablar de comportamientos lingüísticos homogéneos o heterogéneos, y hacer las interpretaciones pertinentes. Para poder averiguar el comportamiento de un estrato con relación a un elemento lingüístico el sociolingüista cuenta con otro parámetro de posición, la moda, es decir, la media que más intensamente se repite.

El cálculo a veces exige un recuento manual con una matriz al estilo de las empleadas por López Morales en las que se distribuyen los datos en intervalos o clases, de ahí su nombre, *marcas de clase*. Las marcas de recuento suelen hacerse de cinco en cinco para facilitar el cómputo final, (HHH) y los intervalos pueden establecerse de 10 en 10. (0-10, 11-21, 22- 32---etc. Este dato es trascendente para el estudio cuando hay determinados elementos que son usados por la mayoría de la población con una frecuencia relativamente baja, y otros, que son usados por muy pocos hablantes pero con una frecuencia considerablemente alta con relación al valor medio. Es necesario, por tanto, conocer *las colas de distribución* y poder estimar, si esos valores medios en cada uno de los hablantes que conforman el estrato estudiado, y dispares con relación a la media del estrato, son relevantes y merecen o no interpretación sociolingüística adicional.

La respuesta a esta cuestión es la *varianza*, un parámetro de dispersión, cuyo mecanismo estadístico analiza Francisco Moreno en su obra sobre metodología sociolingüística con tanta claridad y de forma tan sencilla que no podemos por menos citarla e introducirla aquí. El ejemplo que el propone para ilustrar la teoría y práctica sobre la varianza y la desviación típica se extrae del estudio de la *-s* en posición implosiva de sílaba y palabra que aparecen al leer un texto. Este análisis se efectúa en hombres y mujeres. En la primera de las columnas que emplea para su exposición se anotan las frecuencias absolutas registradas en los hombres y en la segunda en las mujeres, así como la suma total de las aspiraciones ($\sum x$) y la media (Cuadro 1).

A través de estos datos, el autor explica los distintos pasos con los que se llega al establecimiento de la fórmula de la varianza. El cálculo de la varianza, en un ejemplo como el que estamos tratando, consta principalmente de tres etapas:

1°. -Ver la diferencia (*d*) de cada frecuencia respecto de la media, para lo cual se realiza sencillamente una resta ($x_i - X$). Una vez que se han restado todas las frecuencias de la muestra, para comprobar que los cálculos están bien hechos, se suman los resultados entre sí $\sum d$, Siempre debe obtenerse 0.

supone y por ir, por tanto, en contra de la ley de comodidad lingüística, por cuestiones explicadas por la pragmática del discurso y porque existen tiempos verbales más cómodos para el hablante que pueden absorber la expresión de ese concepto].

2°.-Calcular el cuadrado de la diferencia obtenida en cada resta d^2 y sumar los resultados $\sum d^2$.

3°.-Dividir $\sum d^2$ entre el número de x analizadas (n ; en este caso diez) menos uno. Se obtendrá así la varianza.

$$v = \frac{\sum d^2}{n-1}$$

Finalmente, el cálculo de la desviación típica sólo requiere aplicar la fórmula

| Hombres | | | Mujeres | | |
|------------|-------------|----------------|-------------|------------|----------------|
| d | x | d ² | x | d | d ² |
| -2 | 3 | 4 | 1 | -4 | 16 |
| 1 | 6 | 1 | 18 | 13 | 169 |
| 7 | 12 | 49 | 22 | 17 | 289 |
| -1 | 4 | 1 | 1 | -4 | 16 |
| 3 | 8 | 9 | 2 | -3 | 9 |
| 2 | 7 | 4 | 1 | -4 | 16 |
| 1 | 6 | 1 | 2 | -3 | 9 |
| -3 | 2 | 9 | 1 | -4 | 16 |
| -4 | 1 | 16 | 1 | -4 | 16 |
| $\sum d=0$ | $\sum x=50$ | $\sum d^2=110$ | $\sum x=50$ | $\sum d=0$ | $\sum d^2=572$ |
| | $x=5$ | | | $X=5$ | |

Cuadro 1.-Ejemplos propuestos por F. Moreno para la obtención de varianzas.

Tanto la varianza como la desviación típica tiene como objeto cuantificar la variación de los datos y estimar cuánto se desvían los elementos computados con relación a la media. Vinculada a estos datos existe una prueba, el coeficiente de variación, que dirá si el comportamiento de un estrato es homogéneo o heterogéneo. La fórmula de esta prueba, que pese a su rentabilidad en determinados casos, apenas es citada en los manuales de esta disciplina es:

$$V = \frac{\sigma}{x}$$

El valor a partir del cual los resultados respecto a la dispersión son significativos es de $v = 1.20$, según las estimaciones de los estadísticos para un estudio como el que nosotros presentamos.

Pruebas estadísticas. T-Student

Como señalaba Francisco Moreno, «el segundo fin del análisis estadístico es el de hacer estimaciones de significación y de fiabilidad» (1990: 137). Y no podrían hacerse si no se contara con pruebas estadísticas que determinaran si el uso medio de un elemento lingüístico en los distintos estados de una variable, siendo diferente es, a la vez, significativo. En sociolingüística, hay dos pruebas intensamente empleadas con este fin: T-student y χ^2 .

A través de T-Student se puede comprobar si el comportamiento de hombres o mujeres -caso de la variable de sexo- o de hablantes de distinto nivel de edad, es significativamente distinto con relación al empleo de un elemento lingüístico. La principal dificultad que puede encontrar el sociolingüista al decidir si la incluye en su repertorio de pruebas estadísticas es que, y así aparece en todos los manuales, sólo puede manejar dos muestras, por ejemplo, hombres y mujeres. El inconveniente al que nos referimos ya se habrá percibido: las variables manejadas por la sociolingüística no son dicotómicas salvo en el caso de la variable biológica de sexo. Los manuales de estadística recomiendan para estos casos el empleo de ANOVA simple. Sin embargo, aunque todo lo dicho es verdad, debe tenerse en cuenta que aunque los estados de las variables que estudia el investigador no son dicotómicos, el análisis que hace sí podría considerarse así. Es decir, cuando se estudia el comportamiento de los hablantes de distintos grupos de edad respecto del futuro imperfecto, por ejemplo, no se estudia simultáneamente en los tres estados, sino en dos de ellos. Lo que no significa que se esté realizando el análisis por pares, circunstancia que anularía la validez de la prueba (Woods, Fletcher y Hughes 1986), sino un análisis que contempla dos estados, entre más, de una variable como si fueran estados únicos, ya que se pretende el descubrimiento de pautas de comportamiento lingüístico diferente entre dos únicos grupos de hablantes. Se compara el comportamiento de los grupos de edad extremos, G1-G3, o entre un grupo de edad y el tramo generacional inmediatamente anterior o posterior. La única precaución, en este sentido, que debe considerarse es la especificación de esta circunstancia. Por lo demás, el método resulta válido y de excelente rendimiento en sociolingüística. La fórmula de esta prueba es la siguiente:

$$t = \frac{x_1 - x_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$$

y requiere el conocimiento de las medias de cada una de las dos muestras, el valor de las desviaciones y el tamaño de las muestras. La operación matemática arrojará un valor. Para poder comprobar la significación de ese valor, el sociolingüista cuenta con *tablas de valores* que aparecen en cualquier tratado de estadística y para cuyo uso necesita conocer dos datos más: los grados de libertad con los que trabaja, *g, d, l*,

y el nivel de significación del test o número α . El primero corresponde a una sencilla fórmula ($n_1 + n_2 - 2$). Y el segundo, nivel de significación del test, hace referencia al porcentaje de posibilidades de que la conclusión a la que llegamos sea un error²². Se suelen manejar varios números α . = 0.05, α = 0.01 y α = 0.001. En nuestra investigación utilizamos un α = 0.001, lo que suponía que las posibilidades de cometer un error eran 1 de cada 100, en el caso que venimos poniendo como ejemplo, comprobamos que el valor $T=2.94$, cuando el valor mínimo establecido era de $T=2.57$. Por tanto, las diferencias encontradas en cuanto al uso del futuro imperfecto entre los grupos G1y G2, no se debían al azar, sino que eran significativas y ponían, por tanto, en marcha el mecanismo de interpretación sociolingüística.

χ^2 o Test de Pearson

La segunda prueba a la que nos referíamos era χ^2 o *Test de Pearson*. A la hora de hablar de su formulación teórica resultan fundamentales las indicaciones de Francisco Moreno, una vez más. Por esta razón referiremos a ellas, realizando cuando lo estimemos oportuno citas textuales.

Se trata de una prueba no paramétrica, de gran valor y utilidad en estadística, que tiene como función comprobar si la distribución de dos variables es independiente o dependiente. Su fórmula se construye manejando valores observados (o) y valores teóricamente esperados (e), y es la siguiente:

$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

La prueba de χ^2 trabaja a partir de frecuencias absolutas, de ahí la importancia de registrar tanto las relativas como éstas, y consta de los siguientes pasos:

- 1.-Cálculo de los valores esperados.
- 2.-Cálculo de los grados de libertad

²² Toda investigación se inicia con una hipótesis o con una serie de ellas. En trabajos sobre comunidades de habla, se parte de la hipótesis de que el entramado social, económico y cultural de los hablantes tiene un impacto en su patrón de habla. Desde un punto de vista estadístico, y por lo tanto, más técnico, asociadas a las hipótesis se establecen dos tipos de errores posibles. Puesto que las hipótesis solo pueden ser verdaderas o falsas, los errores también pueden ser dos, dando lugar a cuatro situaciones diferentes. Que siendo la hipótesis verdadera la reconozcamos como tal, que siendo la hipótesis falsa la reconozcamos como tal, que siendo la hipótesis verdadera la rechazamos y la demos por falsa, que siendo la hipótesis falsa, la aceptemos y la demos como verdadera. A la situación tercera se la conoce como error del tipo I y a la cuarta, como error del tipo II. Las posibilidades de rechazar una hipótesis verdadera dependen del intervalo de confianza que se ha establecido, del 95% o del 99%. Las de cometer el error del tipo II, no se pueden conocer de antemano dado que dependen de un conjunto variable de factores, tal y como señala Woods.

3.-Comprobación de los resultados en las tablas de distribución correspondiente.

La utilidad de esta prueba para estudios de actuación lingüística es extraordinaria porque permite responder a preguntas del tipo “cuando el hablante emplea el futuro imperfecto, ¿lo hace significativamente en una de las dos conjugaciones, regular e irregular o elige una significativamente en función de su edad; nivel sociolingüístico, sexo..?”

Respecto a la realización de la prueba, cabe hacer dos matizaciones. En primer lugar, es de importancia la correcta colocación de las variables en filas y columnas. Las primeras serán para variables lingüísticas y las segundas para las sociolingüísticas. Esta precaución procede del hecho de que esta prueba se aplica a estratos de volúmenes diferentes. La colocación inversa invalidaría la prueba. Y, por otro lado, aunque nos estamos refiriendo a filas y columnas, no existe impedimento para hacer la prueba sobre una columna única²³, que correspondería a la primera pregunta que hemos empleado como ejemplo: preferencia general, de todos los hablantes sin especificación sociolingüística, por una u otra conjugación al usar el futuro imperfecto.

Pese a su gran rendimiento, existen circunstancias en las que el empleo de χ^2 no es posible: Sólo se puede emplear con frecuencias absolutas, de ahí la importancia de que el programa diseñado para la investigación emita tablas de datos con los dos tipos, relativas y absolutas; no puede emplearse cuando el valor esperado es inferior a 5. $F(e);=5$; y, finalmente, es fundamental el cálculo correcto de los grados de libertad ya que en caso contrario la prueba se invalida.

En nuestra investigación, aplicando χ^2 sobre una columna única comprobamos que la preferencia por la conjugación irregular era extraordinariamente alta, 8 de cada 10, $\chi^2 =54.45$, y, aplicando la prueba de modo convencional, se vio que la edad del hablante no afectaba a esta elección en ningún caso.

MIRIAM LARROSA BARBERO

²³ En ese caso, el grado de libertad sería 1. Sobre este punto, precisamente, solicitamos información a los estadísticos puesto que en nuestro caso podríamos tener necesidad, como así se comprobó en la práctica, de conocer la distribución de dos tipos respecto a una variable. Es decir, simplemente saber si en términos generales los hablantes, todos, cuando emplean este tiempo lo hacen significativamente una de las dos opciones que se le presentan. Este es un primer paso para establecer las comparaciones pertinentes cuando esta pregunta se realiza sobre el conjunto de los hablantes estructurados sociolingüísticamente a través de variables. Puede que, en efecto, la tendencia significativa sea el uso de la conjugación irregular, pero puede ocurrir que en alguna variable, o en algún estado concreto de la variable, la preferencia sea más acusada o se neutralice. Es, como puede entenderse, fundamental comprobar la reacción de los distintos estratos con relación a una pauta general de comportamiento lingüístico.

BIBLIOGRAFÍA

SOCIOLINGÜÍSTICA

- ABAD NEBOT, F. "Diatopía y diastratía lingüísticas" en *Lecturas de Sociolingüística*, de Manuel Alvar, págs. 125-139.
- ALVAR, M. 1986. *Hombre, etnia, estado*. Madrid, Gredos.
- ALVAR, M. 1983. *La lengua como libertad*. Madrid, Ediciones Cultura Hispánica del Instituto de Cooperación Iberoamericana.
- ALVAR, M. 1977. *Lecturas de Sociolingüística*. Madrid, Edaf.
- ARDERNER, E., et al. 1976. *Antropología social y lenguaje*, Buenos Aires, Paidós.
- BEINHAUER, W. 1978. *El español coloquial*. 3a edición, Madrid, Gredos.
- BENTIVOGLIO, P. y M. SEDANO. 1993. "Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana"; *Boletín de Lingüística*.
- BERUTTO, G., La sociolingüística, México, Nueva Imagen, 1979.
- BICKMAN, L., "Recogida de datos. I Métodos de observación", en C. Selltiz et al., 345-398
- BLANCHE-BENVENISTE, C. y C. JEANJEAU. 1958. *Le français parlé. Transcription et édition*. Paris, Institut National de la Langue Française.
- BOISSEVAIN, J. 1974. *Friends of Friends: Networks, Manipulators and Coalitions*, Oxford, Blackwell.
- BORREGO, J. 1981. Sociolingüística rural Salamanca, Universidad de Salamanca.
- BRIGHT, W. 1964. "Social dialect and Linguistic History", en D. Hymes (ed.), 469-472.
- BRIGHT, W. 1966. *Sociolinguistics*, The Hague, Mouton.
- CAMERON, D. 1985. *Feminism and Linguistic Theory*; Hampshire, MacMillan
- CEDERGEN, H. J., *Interplay of social and linguistics factors in Panama*, Ithaca, Cornell University.
- FASOLD, R., *The sociolinguistics of society. Introduction to Sociolinguistics*, vol. 1, Oxford: Basil Blackwell.
- FISHMAN, J. 1979. *Sociología*. Madrid, Cátedra.
- FONTANELLA DE WEINBERG, M.B. "La s postapical bonaerense", *Thesaurus. Boletín del Instituto Caro y Cuervo*, 394-400.
- GUMPERZ, J. y BENNETT, A. 1981 *Lenguaje y cultura*. Madrid, Anagrama.
- GUMPERZ, J. 1971. *Language in social groups*. Sanford, University Press
- HUDSON, R. A. 1981. *La sociolingüística*. Barcelona, Anagrama.
- LABOV, W. 1983. *The social stratification of English in New York City*. Washington, D. C: Center for Applied Linguistics.
- LABOV, W. 1966. *Modelos sociolingüísticos*. Madrid, Cátedra.
- LÓPEZ MORALES, H. 1983. "En torno a la /s/ final dominicana. Cuestiones teóricas". *Voz y Letras*, 1. 129-137.
- LÓPEZ MORALES, H. 1994. *Métodos de investigación lingüística*. Salamanca, Ediciones Colegio de España.
- LÓPEZ MORALES, H. 1989. *Sociolingüística*. Madrid, Gredos.

- LÓPEZ MORALES, H. 1978. "Dialectos sociales en San Juan: Índices de conciencia lingüística", *Boletín de la Academia Puertorriqueña de la Lengua Española*, 6. 5-25.
- LÓPEZ MORALES, H. 1979. "Disponibilidad léxica y estratificación socioeconómica", en *Dialectología y sociolingüística. Temas puertorriqueños*, 173-181.
- LÓPEZ MORALES, H. 1983. *Estratificación social del español de San Juan de Puerto Rico*. México: Universidad Nacional Autónoma de México.
- MACAULAY, R. K.S. 1977. *Language, Social Class and Education; a Glasgow Study*, Edimburg, EUP.
- MARTÍNEZ MARÍN, F. M. 1983. *Fonética y sociolingüística en la ciudad de Burgos*. Madrid, Consejo Superior de Investigaciones Científicas.
- MILROY, L., 1980. *Language and social networks*. Oxford, Blackwell.
- MILROY, L., 1987. *Observing and analyzing natural language. A critical account of sociolinguistics method*. Oxford, Blackwell.
- MORENO, F. 1986. "Sociolingüística de los tratamientos. Estudio sobre una comunidad rural". *Anuario de las Letras*. XXIV. 87-120.
- MORENO, F. 1984. *Estudio sobre el habla de Quintanar de la Orden mediante ordenadores*. Tesis doctoral, Madrid, Universidad Complutense.
- MORENO, F. 1990. *Estudios sobre variación lingüística*. Alcalá de Henares, Universidad de Alcalá.
- Moreno, F. 1990. *Metodología sociolingüística*. Madrid, Gredos.
- MORENO, F. 1988. *Sociolingüística en EE UU (1975-1985)*. Málaga, Ágora.
- SAMPER, J.A. 1988. *Estudio sociolingüístico del español de las Palmas de Gran Canaria*. Las Palmas de Gran Canaria: La Caja de Ahorros de Canarias.
- SANKOFF, D. 1980. *Linguistics variation: Models and methods*, New York, 1978.
- SANKOFF, D. 1980. *The Social Life of Language*, Philadelphia, PUP.
- SANKOFF, D. y S. LEBERGE. "The linguistic market and the statistical explanation of variability". En D. Sankoff, *Linguistics variation. Models and methods*, New York. Academic Press, 239-259.
- SHUY, R.W. 1973. *Sociolinguistics. Current Trends and Prospects*, Washington, Georgetown, U. P.
- SILVA-CORVALÁN, C. 1989. *Sociolingüística. Teoría y análisis*. Madrid, Alhambra Universidad.
- TURNER, J.H., 1984. *Societal stratification. A theoretical Analysis*, New York, Columbia, U. P.
- URIBE VILLEGAS, O. 1974. *La sociolingüística actual*. México, Universidad de México.
- WILLIAMS, L. 1987. *Aspectos sociolingüísticos del habla de Valladolid* Valladolid, Universidad de Valladolid-Universidad de Exeter.
- WOLF, M., 1982. *Sociologías de la vida cotidiana*, Madrid, Cátedra.

ESTADÍSTICA GENERAL Y ESTADÍSTICA LINGÜÍSTICA

- BUTLER, C. 1985. *Statistics in Linguistics*. Oxford, Blackwell.
- MULLER, Ch. 1973. *Estadística lingüística*. Madrid, Gredos.
- TURNER, J. C. 1986. *Matemática moderna aplicada. Probabilidades, estadística e investigación operativa*. Madrid, Alianza Universidad.
- WOODS. A., FLETCHER, P. Y HUGHES. A. 1986. *Statistics in Language Studies*. Cambridge, CUP.