

Metodología para un estudio dialectal urbano: El caso del estudio del español de Bucaramanga*

Luz Mary Rincón
luzmary.rincon@tamut.edu

*Texas A&M University-Texarkana
College of Arts & Sciences and Education*

Resumen

Una constante preocupación de los estudiosos de la lengua es el desarrollo de técnicas confiables en la investigación del habla en contextos urbanos con el fin de ofrecer una teoría de la lengua como se observa en su uso social. Para esto hay que tener en cuenta el aporte de la sociolingüística en el estudio de la lengua en su contexto social y toda aquella gama de contribuciones que se han alejado de las tradicionales técnicas usadas en dialectología. En el presente artículo se da cuenta de la metodología utilizada en cada una de las etapas en el estudio del español de Bucaramanga con el fin de mostrar un modelo metodológico que pueda servir como punto de partida en la exploración de otras hablas urbanas en Colombia. La propuesta expuesta se presenta únicamente como una alternativa de trabajo aceptando que llegar a la técnica ideal para el estudio de fenómenos sociolingüísticos urbanos constituye un ideal que esta aún por resolverse. Este modelo metodológico se ha valido de ensayos anteriores y de propuestas con respecto a la recolección, selección y análisis de datos con el fin de llegar de mejor forma, no al sistema lingüístico ideal, sino al habla vernácula que se desea describir.

Palabras clave: sociolingüística, español de Colombia, español de Bucaramanga, métodos de estudio de lengua vernácula, metodología de recolección, selección y análisis de datos lingüísticos.

Introducción

Uno de los principales objetivos de la sociolingüística es el desarrollo de metodologías y técnicas confiables de recolección y análisis de datos que puedan ser aplicadas al estudio del habla en contextos urbanos. Utilizando como base el estudio sobre el español de Bucaramanga, este trabajo propone una alternativa metodológica que puede

* Este artículo es tomado de uno de los capítulos utilizados en mi tesis doctoral, ganadora del premio a la disertación del año, Ball State University 2004.

servir como punto de partida en el diseño de metodologías para el estudio de fenómenos lingüísticos en otros contextos urbanos colombianos. Con este estudio no se pretende mostrar un modelo que haya encontrado la solución a los problemas inherentes al estudio de hablas urbanas caracterizadas por su complejidad y dinamismo sino plantear una alternativa metodológica que puede ser aplicada al estudio de fenómenos lingüísticos en contextos urbanos colombianos.

El estudio sobre el español de Bucaramanga, con énfasis en el habla de la clase media, describe los principales rasgos fonéticos, morfosintácticos, léxicos y del discurso de esta variedad y su correlación con factores sociales. Los datos fueron recogidos de una muestra representativa de la clase media y analizados utilizando una propuesta de metodología sociolingüística aplicable al estudio específico de esta variedad. La investigación se caracteriza por un enfoque pancrónico (Silva-Corvalán, 2001) ya que se atiende a la variación del habla tanto en su evolución en el tiempo como en lo observado en el ahora. Desde el punto de vista sincrónico (Saussure, 1966) se hace una descripción del uso del habla por los hablantes distribuidos en grupos de edad, con especial referencia al habla del grupo más joven a partir de lo cual se hacen propuestas sobre la dirección futura de algunas variables. Desde el plano diacrónico (1966), se examinan los diversos cambios que ha tenido este dialecto desde la segunda mitad del siglo pasado para lo cual se toma como referencia el estudio del lingüista Luis Flórez (1965), bajo el título *El Español de Santander*. Flórez incluye a Bucaramanga aunque de una manera tangencial como parte de un amplio corpus que abarcaba las variedades habladas en un buen número de municipios del departamento. *El Español de Santander* es quizás el único estudio documentado para esta zona dialectal y no se reconocen estudios previos sobre el habla de Bucaramanga, uno de los centros urbanos más importantes del país, con lo cual este estudio además de revelar los principales rasgos de esta variedad, contribuye a enriquecer el campo de la dialectología urbana en Colombia.

Los pasos que siguió el trabajo se identifican con la mayoría de estudios sociolingüísticos cuyo objetivo es la descripción y explicación por medio de análisis

cuantitativo de hechos reales del habla de una comunidad. Estos se resumen así: observación de la comunidad y planteamiento de hipótesis del trabajo, delineamiento del perfil de la muestra y selección de los hablantes, recopilación de datos, análisis cualitativo y cuantitativo de datos, interpretación de los resultados, y propuestas de investigación futura. En el presente artículo se da cuenta de la metodología utilizada en el desarrollo de cada una de estas etapas y se pretende mostrar un modelo metodológico que puede servir de base en la investigación de hablas urbanas en el país. Con éste se busca igualmente impulsar la investigación sociolingüística urbana en Colombia, un campo en el cual hay un número muy escaso de contribuciones (ver Becerra, 1980; Montes et al., 1998; Otálora, 1997, 1998; Otálora & González, 1986).

1. Ubicación del contexto de estudio en el plano dialectal

El primer aspecto a tener en cuenta para delimitar las variables que se desean examinar desde el punto de vista sociolingüístico, es recoger información pertinente a partir de investigaciones anteriores y junto con observación empírica crear hipótesis sobre los rasgos que se desean incluir en el estudio y su variabilidad. En el caso de Bucaramanga no se encontraron estudios documentados previos, por lo que se debió entonces recurrir a su ubicación dentro del plano dialectal hispano en Colombia con el fin de ubicar aquellos rasgos que generalizaran a esta variedad. Para ello se utilizó la propuesta de clasificación bidialectal para Colombia que sugiere la división del español de Colombia en dos superdialectos, el costeño y el andino, para diferenciar, por un lado, las variedades regionales habladas en las zonas costeras y zonas bajas adyacentes y, por otro, las variedades regionales habladas en toda la zona Andina o tierras altas (Montes, 1982). La Tabla 1 muestra esta propuesta dentro de la cual se ubican dos superdialectos que incluye la subdivisión de la lengua como entidad abstracta, que aglutina dialectos que incluyen variedades que comparten rasgos lingüísticos y que a su vez se dividen en subdialectos que son variedades con características lingüísticas más afines. Criterios de clasificación similar habían sido utilizados ya con anterioridad por Flórez (1961) al sugerir dos grandes variedades, la costeña y las variedades habladas en el resto del país

Adaptado de “Probables áreas dialectales del español en Colombia” (Flórez, 1961). El mapa muestra las áreas dialectales propuestas por Flórez, dentro de las dos superzonas propuestas por Montes (1982).

Parte de la región norte de los departamentos de Santander y Norte de Santander se encuentran dentro de la subzona nortesantandereana dentro del dialecto Costeño Caribe (Montes, 1982). Igualmente, el cordón occidental del departamento a lo largo del Río Magdalena presenta rasgos característicos de habla de las zonas costeñas. El habla de Bucaramanga se ubica dentro del subdialecto Santandereano dentro de la zona Andina o de tierras altas. Aunque se reconoce otra variedad distinta, el subdialecto Cundiboyacense, aunque hay cierta distinción lexical, no se encuentran fenómenos fónicos o gramaticales que justifiquen esta separación, como lo sugiere Montes. Como se concluye en la investigación, aunque Bucaramanga está ubicada dentro de la zona subdialectal Santandereana, es el mayor centro urbano de la zona y como es de esperar, muchos de sus rasgos lingüísticos gramaticales y lexicales, junto con algunos fónicos la distinguen del resto de variedades de la zona caracterizada por rasgos propios de habla rural. Por otra parte, especialmente su sistema fónico ha sido influenciado por el habla costeña. La aspiración parcial de /s/ por los hablantes de edad media, por ejemplo, puede ser el resultado de la influencia del dialecto costeño que consistentemente aspira

el segmento, especialmente en posición final de sílaba. Esta influencia se debe no tanto a su ubicación geográfica próxima a zonas costeñas de influencia como el puerto petrolero de Barrancabermeja, sino al fenómeno migratorio de la región. En las últimas décadas, la ciudad se ha visto forzada a recibir cientos de familias desplazadas por el conflicto armado provenientes áreas costeras y zonas bajas del interior del país lo cual puede haber influenciado algunos de los rasgos del habla.

Según criterios geográficos, Bucaramanga se encuentra dentro de la región Andina con lo que se hipotetiza que un buen número de rasgos del español hablado en esta área pueden ser tomados de los rasgos generales descritos para las tierras altas. En un punto medio entre la costa Atlántica y la capital del país, Bucaramanga se encuentra localizada sobre una meseta en pleno corazón de la cordillera oriental de los Andes, con vista hacia el valle del Río de Oro por lo que se distinguen en la ciudad un sector conformado por el valle y otro por la meseta con suelos de una alta potencialidad erosiva. Hoy la meseta ha sido casi en su totalidad poblada al unirse con sus municipios vecinos de Girón, Floridablanca, y Piedecuesta, los cuales, junto con Bucaramanga, conforman el *Area Metropolitana de Bucaramanga*, un importante centro industrial, comercial y educacional con gran influencia no sólo en Santander sino sobre el resto del nororiente del país. Esta agrupación metropolitana ha obligado a un proceso de reordenamiento territorial adelantado por el Área Metropolitana de Bucaramanga.

Los criterios de ubicación en el plano dialectal que emanan en gran parte de ubicación geográfica se juntaron con factores de socio-político y socio-económico para fijar hipótesis sobre las variables que constituyeron el objeto de estudio del presente trabajo y sobre las etapas que se iban a tener en cuenta en el análisis de las variables, descritas más adelante. En lo que sigue, se hace una descripción de la incidencia de los factores socio-político y socio-económico en la variabilidad del español de Bucaramanga, con lo que se quiere mostrar la importancia del análisis de los fenómenos de la lengua más allá del contexto lingüístico, es decir, el estudio de la lengua dentro de la sociedad y en su contexto social más amplio.

2. La clase social como factor social invariable

El objetivo del estudio tenía en cuenta dos grandes consideraciones: una hacer un esbozo general sobre de este dialecto a partir de la delimitación dialectal, y otra describir su variabilidad en el ahora. Ya que estudios documentados datan de los 60s se debía entonces mirar la lengua en el plano diacrónico para lo cual se presentaron puntos de comparación con los rasgos propuestos por Flórez (1965) para el subdialecto santandereano. Para estudiar su variabilidad en el plano sincrónico se debía limitar el estudio a una muestra representativa de Bucaramanga por lo que se tomó como modelo el habla de la clase media lo que a su vez implicó llevar a cabo un estudio de estratificación social. La clase social constituyó de esta forma el factor social pre-determinado en el análisis de la correlación de las variables lingüísticas con factores extra-lingüísticos. En lo que sigue se hace una descripción de la metodología utilizada en la escogencia de la muestra y de las razones por las que se decidió escoger la clase media para obtener la muestra representativa del habla de Bucaramanga con el fin de analizar la variabilidad presente de la lengua y establecer una hipótesis sobre su proyección.

El principal criterio para la selección de la clase media como representante del habla de Bucaramanga es la caracterización tradicional de la ciudad como un centro de clase media (Secretaría de Desarrollo Municipal de Bucaramanga, 2001, comunicación personal). La clase media bumanguesa fue constituida durante los años 1920s cuando la ciudad empezaba a recuperarse de la depresión económica causada por la Guerra de los Mil días de finales del siglo anterior. Luego de esta crisis económica, Bucaramanga comienza su recuperación en forma paralela con el proceso de urbanización y la migración hacia la ciudad de campesinos provenientes de zonas rurales del departamento en busca de mejores oportunidades laborales y a mediados del siglo pasado por huir de la violencia partidista que azotó al país en esa época. Bucaramanga, como el resto de centros urbanos del país inicial un crecimiento evidente a partir de los años 50 (Cardona & Simmons, 1976). A partir de entonces y durante varias décadas

Bucaramanga gozó de un auge económico que posibilitó la movilidad de sectores socio-económicos más bajos hacia la clase media. Hoy por hoy, el desplazamiento hacia la ciudad motivado por el subdesarrollo de los campos y la inseguridad causada por el conflicto armado que ha tenido como foco zonas rurales del departamento y un amplio sector de la costa norte, junto con el receso económico del país han causado el aumento del desempleo y la movilización inversa desde los estratos superiores hacia sectores menos favorecidos, especialmente de miembros de la clase media hacia sectores socio-económicos más bajos. Por otra parte, es de tener en cuenta que aunque la migración hacia las ciudades promueve la movilidad social (1976), el fenómeno reciente ha motivado una movilidad inversa, hacia estratos de nivel inferior, posibilitando el debilitamiento de la clase media.

La estratificación social es importante para entender el dinamismo social de las áreas de mayor concentración demográfica como es el caso de los municipios y distritos urbanos en el país. En Bucaramanga, la estratificación social está determinada tanto por la calidad de la vivienda como por las características residenciales del vecindario donde se encuentran ubicadas estas viviendas, la calidad de la construcción y las facilidades del sector (Secretaría de Desarrollo Municipal de Bucaramanga, comunicación personal, 2002). Estos criterios, más que los ingresos económicos de sus habitantes son tomados para definir el estrato socio-económico de las áreas urbanizadas. Para efectos de organización, planeamiento, y desarrollo urbano, Bucaramanga ha sido dividida en 6 estratos sociales: estrato 1 o bajo bajo, estrato 2 o bajo, estrato 3 o medio bajo, estrato 4 o medio medio, estrato 5 o medio alto, y estrato 6 o alto (Ciudad Digital, 2002; AMB, 2002; Secretaría de Desarrollo Municipal de Bucaramanga, 2002; Superintendencia de Servicios Públicos Domiciliarios, 2002), de los cuales el estrato 4 (objeto del presente trabajo) está asociado con la clase media (AMB, 2002: 209). Estudios anteriores ya habían seguido criterios de estratificación de las comunidades (ver por ejemplo Bentivoglio, 1980-1981; Bentivoglio & Rosenblat, 1979; Trudgill, 1974; Labov 1972, 1972a; y para un análisis más detallado de la estratificación social en comunidades hispanas Silva-Corvalán, 2001). Un estudio importante para Colombia la estratificación del español de Bogotá (Montes, 1998) y su incidencia en el estudio de una serie de

fenómenos lingüísticos, el cual sirve de modelo para estudios de este tipo en el contexto colombiano.

Tres barrios de clase media fueron escogidos para representar la comunidad de clase media participante en el estudio: *Barrio Provenza*, en el sureste, *Barrio San Alonso*, en el nororiente, y *Ciudadela Real de Minas*, en el centro oeste. Además de estar localizados en zonas estratégicamente ubicadas dentro de la ciudad, éstos representan tres tendencias diferentes de la clase media: el *Barrio Provenza*, un sector relativamente nuevo en su mayoría conformado por casas unifamiliares modernas, el *Barrio San Alonso*, un sector tradicional en su mayoría conformado por casas de construcción antigua con algunos sectores más modernizados por la reciente construcción de bloques de apartamentos, y *Ciudadela Real de Minas*, el sector más popular de clase media de la meseta. La muestra más representativa de la clase media la constituye la *Ciudadela Real de Minas* cuya urbanización es el resultado de una propuesta de planeación territorial que buscaba la optimización del espacio urbano por medio de la creación de ‘ciudades dentro de la ciudad’ (Amaya Pedraza, 2000). *Ciudadela Real de Minas* es un proyecto urbanístico de reconstrucción y reordenamiento del espacio público que ha dado como resultado el auge en la construcción de urbanizaciones y conjuntos residenciales pertenecientes al estrato 4 con alternativas adicionales de comercio y servicios comunitarios y educativos. Uno de los requisitos para la escogencia de la muestra fue su pertenencia al grupo social de la clase media, con lo cual se impuso el criterio de residencia en alguno de estos tres sectores por un periodo no inferior a 5 años. Este criterio de igual forma excluía la posibilidad de movilidad reciente entre estratos de la muestra participante y por lo tanto mayor representatividad del habla de este grupo social.

Como se concluye de la descripción presentada en los apartes 1 y 2 de este trabajo, se parte de un nivel macro hasta llegar a contextos de estudios más específicos con el fin de delimitar la muestra y a su vez con el fin de que existan criterios suficientes para sustentar su representatividad. Con estas consideraciones se pasó a la siguiente etapa: la selección de la muestra.

3. Perfil de la Muestra y Recopilación de Datos

El siguiente paso consistió en el delineamiento de los criterios para seleccionar la muestra que iba a ser invitada a participar: número de participantes, edad, nivel educativo, agrupación por género, origen, y el criterio de residencia, como se describió en el párrafo anterior. Con esto se pretendía formar un grupo de participantes que fueran representativos de la clase media y a su vez de la población bumanguesa. Una vez establecidos estos los demográficos se debía determinar la técnica más apropiada para la selección de la muestra. La metodología del *survey* o muestreo al azar (Silva-Corvalán, 2001) ha mostrado ser la más eficiente en la obtención de una muestra representativa y la técnica más usada por estudios representativos de este tipo (Labov, 1966; Cedergren 1973; Molina Matos, 1998; Samper Padilla, 1990). Por medio de esta técnica se escoge uno de cada X número de individuos hasta completar el número requerido. La otra técnica comúnmente utilizada es la de muestreo intencionado que es un sistema estrictamente estratificado en el que se establecen de antemano las categorías sociales y el número total de participantes. Con esto se busca que haya un número equilibrado de participantes distribuidos por celda (Silva-Corvalán, 2001), en donde ‘celda’ se refiere a las divisiones obtenidas en el plano horizontal y vertical, por ejemplo, de hablantes por en alguno de los criterios sociales y las variantes. Ya que el estudio tenía un sólo investigador principal, se optó por esta técnica con el fin de descartar de inmediato los individuos que no llenaban los requisitos preestablecidos e ir completando las celdas con participantes igualmente distribuidos según los criterios de género de hablante, edad --con mayor representatividad de los hablantes jóvenes, seguidos por el grupo de edad media y menor número de participantes adultos, y educación, excluyendo la clase social ya que, como se mencionó en el aparte anterior, era el factor pre-establecido. En lo que sigue de esta sección se muestran detalladamente los pasos en esta rigurosa selección.

Con el fin de determinar el número de participantes en el estudio se utilizó la fórmula propuesta por Labov (1966) quien sugiere que el 0.025% del total de habitantes de un contexto urbano estudiado es suficiente para obtener representatividad. Igualmente se usaron datos estadísticos del censo de 1993 y las proyecciones poblacionales para el 2002, año donde se inició la recolección de los datos. Según el censo de 1993 realizado en 24 de los 45 barrios de estrato 4 en la ciudad la población mayor de 18 años de clase media en la ciudad era de 99,438 habitantes (Ciudad Digital Bucaramanga, 2002—datos tomados de Departamento Administrativo Nacional de Estadísticas DANE, 1993). Estas estadísticas corresponden al 53% del total de habitantes adultos de clase media en la ciudad para el año 1993. Con el fin de obtener un número aproximado de población actual de clase media, se utilizaron estos datos junto con los patrones de crecimiento establecidos por el AMB (2002). De aquí, se calcula en 239,975 el número promedio de la población de este grupo social para el año 2002. Aplicando la fórmula de Labov (1966) la selección de 60 hablantes representaría un número significativo para el estudio. Sin embargo, con el fin de incrementar la distribución de hablantes según las variables sociales con las cuales se correlacionaron las variables lingüísticas, éste número fue aumentado a 70--35 hombres y 35 mujeres, i.e., 0.029%. La muestra incluyó un número de participantes proporcional al total de individuos de la población por categoría de clase social, no al número total de individuos según su distribución por sexo, edad, ocupación o nivel de estudio como en el caso de muchos otros estudios (véase Samper Padilla, 1990 y otros en Silva-Corvalán, 2001). Además como se muestra enseguida, se trató de mantener un número equilibrado de individuos por cada categoría social establecida para el estudio: edad y sexo. Según Lavandera (1975), en concordancia con la presente investigación, esta es una técnica apropiada cuando el objetivo del investigador es establecer la relación entre los hechos lingüísticos y las características sociales de la población, más que encontrar la frecuencia en el uso de una variable determinada.

El requerimiento de origen igualmente limitaba la selección de participantes a aquellos que fueran nativos de Bucaramanga o nativos de tierras altas de Santander y residentes en Bucaramanga desde la niñez. Sólo 1 hablante en el grupo de los jóvenes y

3 en el grupo de edad media no eran originarios de Bucaramanga, en contraste con los hablantes mayores quienes más de la mitad son originarios de pueblos de la región montañosa del departamento como *Lebrija, San Gil, San Andres, Málaga, Barichara, Landázuri, Rionegro, y Vélez*, lo cual a su vez es reflejo de la masiva inmigración de las zonas rurales del departamento hacia Bucaramanga antes de la primera mitad del siglo pasado. El estudio hace énfasis en las tierras altas como una opción de origen en el departamento, ya que el departamento de Santander tiene zonas hacia el noroeste a lo largo del Río Magdalena cuya habla tiene rasgos propios del español de zonas bajas. Otros departamentos que limitan con Santander no fueron considerados para el estudio puesto que corresponden a otras zonas dialectales.

El estudio no incluyó hablantes menores, i.e. por debajo de los 18 años, por las restricciones impuestas en la toma de datos a este tipo de participantes. Los participantes fueron divididos en tres grupos generacionales de acuerdo con los siguientes criterios:

- Hablantes jóvenes—18 a 34 años de edad
- Hablantes de edad media—35 a 54 años de edad
- Hablantes mayores—55 años de edad o más

En el trabajo se utiliza el término ‘mayor’ o ‘adulto’ para designar aquellos hablantes (ya sea participantes en el estudio o bumangueses en general) de 55 años de edad o más. En la Tabla 3 se observa el número de hablantes según las variables de edad y género, las cuales fueron los factores extra-lingüísticos que más incidieron en la variabilidad del habla de este grupo social. El más alto porcentaje de hablantes estuvo conformado por el grupo de jóvenes con lo cual se pretendió establecer de forma más objetiva la dirección de algunas de las variantes analizadas en la investigación.

Tabla 3. Distribución de los Hablantes de acuerdo con las Variables Edad y Género

Grupo de Edad	Hombres	Mujeres	Porcentaje según Edad
Hablantes Jóvenes	17	14	44.2%

Hablantes de Edad Media	12	12	34.2%
Hablantes Mayores	6	9	21.4%

Como lo muestra la tabla, el grupo de hablantes jóvenes representó el mayor número de participantes en la muestra, 44.2%, seguido por el grupo de edad media.

En el diseño de la propuesta de estudio se contempló la posibilidad de incluir el factor educación con el fin de determinar si el nivel de escolaridad incidía en el uso de las variables. Sin embargo, al comparar los niveles de educación de la muestra seleccionada se pudo determinar que la educación estaba directamente correlacionada con la edad ya que en un alto porcentaje los hablantes jóvenes tenían un grado universitario o estaban asistiendo a centros de educación superior; la mayoría de los hablantes de edad media había terminado una carrera universitaria o estaban realizando estudios de postgrado, mientras que la mayoría de hablantes en el grupo de los adultos habían alcanzado únicamente nivel de educación primaria y en pocos casos secundaria. Es importante anotar que en el pasado, completar la educación secundaria era esencial para el hijo de clase media, especialmente de clase media alta; en la actualidad tener un grado universitario es cada vez más indispensable, con lo cual se aumentan las posibilidades en el mercado laboral. Para el miembro de la clase media, un buen nivel de educación es indispensable para aumentar las opciones ocupacionales. Por lo que la edad estaba directamente correlacionada con la educación, ésta no se tuvo en cuenta.

Una vez se obtuvo la delimitación del perfil de la muestra demográfica, se procedió a delinear los criterios para la recopilación del corpus lo que incluía el diseño de una técnica apropiada de acercamiento a los participantes y de una técnica de recopilación de datos que combinara la observación participante y la entrevista individual con el fin de obtener datos de habla formal juntos con métodos apropiados para recoger datos de habla vernácula, sin ajuste a la presencia del observador. Estas consideraciones se presentan a continuación.

4. Pasos en la Recopilación del Corpus

El objeto de estudio de la sociolingüística es el habla en su contexto social real y el análisis de datos reales de habla iniciando por una interpretación de los hechos observados por medio de un análisis cualitativo seguido del análisis cuantitativo. El estudio del español de Bucaramanga se nutre de todos estos criterios con lo cual se plantea una metodología empírica con un componente importante de análisis de datos observados teniendo en cuenta generalizaciones lingüísticas a partir de su ubicación en el plano dialectal.

Para lograr un acercamiento a los participantes, se hizo una identificación en sondeos previos de aquellas personas que estuvieran interesadas en participar en el estudio. Con estos primeros contactos se empezó a diseñar una red cerrada en cada uno de los sectores seleccionados que incluyó los hablantes inicialmente escogidos, vecinos, amigos y familiares de éstos. Otros hablantes incluyeron dueños de tiendas, misceláneas y salas de belleza de los tres sectores seleccionados. Estos negocios representan los sitios más comunes de reunión de gentes del sector y por lo tanto constituyen una fuente de información importante sobre el habla de la comunidad.

Con todos los hablantes se hicieron al menos dos encuentros. En el primero se establecía si los participantes llenaban los requisitos propuestos por el estudio y se hablaba de las generalidades del trabajo y de la forma como iban a participar, informándoles que más que de un estudio de la lengua se trataba de un estudio de las costumbres de la gente con el fin de que éstos no se formularan juicios sobre la forma como debían hablar y evitar recoger datos no reales de habla. En esta visita se les hacía firmar una nota de consentimiento se les informaba sobre la confiabilidad y anonimato. En este sentido se les preguntaba si en caso de publicación de las entrevistas, los nombres usados debían ser eliminados o cambiados, a lo cual la mayoría de los participantes no indicó que se podían usar datos reales. Igualmente se recogían datos demográficos en una ficha individual. Con anterioridad a la iniciación del estudio fue necesario obtener de parte de un comité de revisión institucional de Indiana University

y de Ball State University, la aprobación de la participación de sujetos humanos en la investigación con el fin de asegurarse que la investigación no fuera causal de daño físico o psicológico a los participantes. Murray & Ross-Murray (1966) presentan un estudio sobre los aspectos legales y éticos relacionados con grabaciones anónimas hechas a sujetos humanos. En el segundo encuentro se programaba la entrevista grabada y el desarrollo de un cuestionario lingüístico. Algunos participantes aceptaron su participación en conversaciones de grupo con otros familiares o amigos en situaciones informales. Las entrevistas, la conversación grupal y el cuestionario lingüístico se complementaron ya que mientras éste arrojó datos de habla comparable, esquematizada y formal, las entrevistas y conversaciones grupales arrojaron datos de habla con varios niveles de informalidad.

La preocupación central durante la recogida de datos fue la obtención de datos de habla natural, lo más cercano posible al habla vernácula espontánea con presentación de datos de habla cotidiana. Con el fin de obtener datos de lengua vernácula se buscaron temas que fueran parte del habla diaria de los participantes. Aunque se iniciaba como una entrevista formal se trató de que se lograra llegar al nivel informal de conversación por lo que el entrevistador trataba de acomodar su habla al estilo del participante. El corpus fue recogido en el 2002, periodo que coincidió con las elecciones presidenciales y un alto nivel de inestabilidad política, económica y de desempleo. Estos representaban temas del habla cotidiana de la gente en ese momento por lo que a su vez facilitaba la obtención de datos. Anterior a la toma real del muestreo estos temas ya se habían piloteado con 15 hablantes de diversos estratos y edades mostrando ser aptos para mantener a los participantes interesados en la conversación. En ese momento eran igualmente apropiados para recoger datos de la clase media ya que éste ha sido uno de los grupos más afectados por la situación económica. La conversación iniciaba de manera formal preguntando al participante sobre sus estudios, rutina diaria, ocupación y la manera como la situación económica actual había incidido en su cambio de costumbres. Además se elicitaban narraciones de experiencias personales, ciertas experiencias de peligro a causa de la situación socio-política del

momento, y las expectativas del cambio de gobierno con el fin de lograr más tiempo de intervención por parte del entrevistado.

El cuestionario lingüístico consistió en una parte fonética grabada, y una sección morfosintáctica y léxica con respuestas tomadas por escrito. La sección fonética consistió en la lectura de una lista de palabras y frases que contenían los segmentos que según hipótesis planteada con anterioridad presentaban una alta variabilidad en el habla de Bucaramanga. El cuestionario contenía igualmente variables morfosintácticas que dada la hipótesis planteada inicialmente tenían una alta variabilidad. Sin embargo, a partir de la observación de datos de habla real se pudo establecer que algunos de los datos no ameritaban su análisis más allá de un nivel puramente cualitativo. El siguiente es un ejemplo de preguntas en la parte fonética del cuestionario seguido por un ejemplo de preguntas usadas para el tratamiento del pronombre direccional en la parte morfosintáctica:

1). Alternancia /j/~/ç/

(a) Pregunta: ¿Cómo se llama esa ave que pone huevos que nos comemos? Palabra a elicitar: gallina.

(b) anteayer: ¿Cómo se llama al día antes de ayer?

(c) colmillos: ¿Cómo se llaman estos dientes (señalándolos)?

(d) calle: ¿Por dónde circulan los carros?

2). Pronombre direccional de segunda persona

(a) ¿Cómo se dirige usted a sus amigos más íntimos, de ‘tú’, ‘usted’, o qué otra forma usa?

(b) ¿Cómo se dirige usted a su novia(o)/su esposo(a), de ‘tú’, ‘usted’, o qué otra forma usa?

(c) ¿Cómo se dirige usted a sus conocidos o compañeros de estudio/trabajo de ‘tú’, ‘usted’, o qué otra forma usa?

Con los datos recogidos se inicia el análisis de los rasgos de la fonología y morfosintaxis. Para el estudio de los rasgos lexicales, se incluyeron datos adicionales recogidos en el 2006 para usar datos más actualizados ya que este sistema presenta

mayor variabilidad. Tanto las entrevistas como las conversaciones fueron grabadas por medio de un sistema digital lo cual facilitó la computación de los datos. Un criterio importante de tener en cuenta en la recopilación de datos y que ha sido la preocupación de la sociolingüística (Lastra, 1992; Silva-Corvalán, 2001) es lo que Labov ha denominado “la paradoja del observador” (1972a) en la cual el investigador se ve enfrentado a recoger datos siguiendo una observación sistemática del comportamiento lingüístico del hablante cuando éste no se siente observado de forma sistemática, especialmente cuando no se puede evitar la intrusión del investigador observador. Como el evitar la intrusión del observador es algo que no se puede evitar pues en cierta forma, el investigador está encargado de mantener el flujo de la conversación, se han hecho varias sugerencias para que el entrevistador sea parte activa en la charla. Por ejemplo Silva Corvalán (2001) propone que el investigador no se limite sólo a hacer preguntas sino que también responda, haga comentarios, intercambie ideas, llene con marcadores las pausas en la conversación, y hable sobre sí mismo. Debe crear una atmósfera íntima y personal que estimule a hablar al individuo estudiado. Estos criterios fueron tenidos en cuenta con el fin de obtener datos más próximos al habla vernácula durante las entrevistas. Sin embargo este método causó la alta intervención del encuestador en algunas entrevistas.

Con los datos en mano se pasa a la siguiente etapa, la cual es la fase más larga y que amerita un trabajo más cuidadoso, conocimiento del sistema lingüístico y capacidad de interpretación: el análisis de los contextos lingüísticos y su relación con factores sociales, puntos referidos a continuación.

5. Análisis de datos e interpretación de los resultados del análisis

La investigación inicia con una hipótesis específica tomando como punto de partida algunos de los datos aportados por el trabajo de Flórez (1965) sobre el español de Santander, área dialectal donde se encuentra el contexto de estudio. Al realizar un análisis comparativo entre ambos dialectos se pudo determinar que había un buen

número de rasgos que distinguían ambas variedades. No se debe dejar de lado el hecho de que hay una distancia temporal de más de cuarenta años entre ambos estudios, lo cual contribuye a establecer la dirección que han tomado algunas variables en esta región dialectal. Hay que tener en cuenta igualmente que, aunque Bucaramanga había sido uno de los lugares incluidos en el estudio, los datos recogidos en Bucaramanga se sumaron al resto de datos del departamento con lo que logró juntar datos de una variedad casi eminentemente rural, a diferencia de los datos que se recopilaron en el presente estudio. En la interpretación hay que tener en cuenta igualmente que el dinamismo social y cultural inherentes a un contexto urbano aumentan la variabilidad de la lengua, especialmente en su sistema lexical y semántico. En seguida se describen algunos rasgos que muestran el cambio lingüístico y el estado actual de algunas variables en el habla de Bucaramanga como se pudo concluir en los resultados de la interpretación del análisis.

Mientras que según Flórez (1965) en el campo fonético, en habla de Santander hay una distinción marcada entre la palatal lateral /l/, ortográficamente representada con *ll*, y la palatal fricativa /y/, ortográficamente representada con *y*, el español de la Bucaramanga de hoy se caracteriza por la alternancia entre [l]~[y] a favor del *yeísmo*, quizás con menos frecuencia entre el grupo de los hablantes adultos. Por otra parte, mientras los rasgos aspirados no son comunes en las zonas altas del departamento, en Bucaramanga se da la aspiración de /s/ en algunos contextos fonológicos. En el aspecto morfosintáctico, el español de rural de Santander favorece el uso de la forma plural de *haber* seguido por un sintagma nominal como en “*en la puerta habían dos personas*” (Flórez, 1965: 71), mientras que en Bucaramanga ésta es una forma considerada no apropiada y los hablantes tienen a mantener normas prescriptivistas. Igualmente se pudo establecer que los bumangueses son conscientes de la forma como hablan y de lo que es considerado estándar o vulgar y tienden a seguir reglas conscientes en su uso de la lengua.

A medida que se van recopilando datos, el estudio puede tomar una dirección diferente a la planteada inicialmente por lo que se debe volver a considerar las hipótesis establecidas inicialmente. En el caso específico de esta investigación al haber hecho una serie de puntos comparativos iniciales, algunas variables fueron abandonadas totalmente del análisis. Un caso es la distinción del número y género de ciertos nombres de uso popular luego de haber notado su estabilidad en Bucaramanga, o el caso del análisis de *vos* en el estudio de los pronombres direccionales ya que no hubo datos que corroboraran su uso. Se decidió sin embargo, examinar más detenidamente los factores que han incidido en el uso de *su merded*, forma que se creía abandonada totalmente, pero que aún es usada por una porción, aunque mínima, del grupo de hablantes jóvenes. A medida que se iban recogiendo datos se pudo observar el interesante nivel de variabilidad en el uso de marcadores del discurso, especialmente en el uso de *pues*, *o sea* y *entonces*. Al avanzar más en la investigación se encontró una interesante variabilidad en el uso de la expresión de condicionalidad y de la expresión de tiempo futuro. Por ejemplo, se notó el bajo uso del futuro del indicativo *-re* (como en *iré al cine mañana*) en la expresión del tiempo futuro y del imperfecto del indicativo *-ba* (como en *si tuviera dinero compraba...*) en la cláusula independiente en la expresión de condicionalidad, variantes que son usadas con alta frecuencia por los santandereanos entrevistados por Flórez (1965). Es por esto que se decidió incluir una sección más en el cuestionario lingüístico para tener más datos a parte de los recogidos en el habla vernácula y con lo que se pudiera realizar una comparación sistemática.

Luego de recopilar el corpus con el que se desea inicial el análisis se debe tener claridad sobre la forma como estos datos van a ser utilizados para lo cual el investigador debe volver a la hipótesis inicialmente planteada. La mayoría de los estudios hacen un planteamiento inicial de un solo rasgo y en su mayoría en el aspecto fonológico. Este es el mejor planteamiento por razones de costo y tiempo. En el caso del presente estudio se decidió tomar una ruta más amplia ya que se partió del hecho de que no había estudios reconocidos documentados previos de este dialecto. Por esto se plantearon las siguientes etapas para el análisis de los datos:

- 1). Realizar una descripción de los principales rasgos del sistema fonológico, morfosintático, lexico-semántico y del discurso y a la vez incluir un análisis comparativo entre el español de Santander y el español de Bucaramanga.
- 2). Realizar el análisis cualitativo de un número determinado de rasgos lingüísticos en donde se había observado mayor variabilidad. Este análisis incluía el examen de la variable dentro de su contexto lingüístico y la frecuencia de la alternancia entre las variantes posibles. Esta frecuencia se determinaba a partir de la prueba por porcentajes con la que se establecía si había criterio suficiente para hacer un examen más detallado con respecto a los factores sociales que pudiesen incidir en su uso. Por ejemplo, la Tabla 4 muestra el total de casos de la porcentaje de frecuencia de la alternancia *había~habían+N plural* con un conteo de 59 casos únicamente, por lo que no se consideró número apropiado para llevar a cabo un análisis sociolingüístico más exhaustivo.

*Tabla 4. Frecuencia de la alternancia *había~habían+N plural**

<i>había</i>	<i>habían</i>	Total de casos
30%	69%	59

Nota: Los datos de la Tabla 4 son tomados del cuestionario lingüístico.

- 3). Definir algunos de los rasgos descritos en 2). con el fin de realizar un examen detallado de los factores sociales que pudiesen haber incidido en su variabilidad. En este caso se hizo un examen más exhaustivo por medio del análisis estadístico y la correlación con factores sociales de género y edad. Por su complejidad, esta parte del análisis se limitó a las siguientes variables:

- (a) alternancia [y]~[ll]
- (b) /s/ en posición inicial de sílaba, final de palabra seguido por vocal, y al final de sílaba seguido por consonante
- (c) pronombres direccionales
- (d) expresión de tiempo futuro y de condicionalidad
- (e) marcadores del discurso *pues, entonces, o sea*

La prueba estadística Chi Square, χ^2 , fue utilizada para determinar la validez entre la correlación de variabilidad de algunos rasgos y los factores sociales tenidos en cuenta. Este es un *software* que de forma eficiente determina los niveles de validez entre la correlación de las variantes de una variable determinada y el factor social con el que se desea correlacionar la variable. Un valor igual o inferior a 00.1 ($p \leq .001$) arrojado por la prueba muestra que la probabilidad de un resultado positivo falso es igual o inferior a .0001 lo que indica un mínimo margen de error y mayor grado de validez en la correlación. La Tabla 5 es un ejemplo de la validez en la correlación entre el uso de los marcadores *pues*, *entonces*, *o sea* y su correlación con el factor edad.

Tabla 5. Casos de *pues*, *entonces*, *o sea* en el corpus analizado por grupo de edad

Hablantes por grupo de edad	<i>pues</i>	<i>o sea</i>	<i>entonces</i>
Jóvenes	501	148	170
Edad media	321	65	211
Adultos	72	10	136

Nota: la relación numérica corresponde al número de casos de cada variante contados en el corpus

$p \leq .001$

Cuando el margen de error era mínimo como en este caso, el análisis incluía igualmente el promedio de uso de las variantes por hablante dentro del corpus recogido como en el ejemplo de la Tabla 6. El promedio se determinaba a partir de la totalidad de casos por celda dividido por número de hablantes –en cada grupo de edad en este caso particular. Obtener el promedio es una forma de mostrar la alta frecuencia de una variable frente a otra. Como se indica en la Tabla 6, se muestra por ejemplo que *pues* es el marcador más frecuentemente usado por el grupo de los hablantes jóvenes, mientras que los adultos prefieren *entonces* para marcar su discurso.

Tabla 6. Promedio por hablante en el uso de *pues*, *entonces*, *o sea*

Hablantes por grupo de edad	<i>pues</i>	<i>o sea</i>	<i>entonces</i>
Jóvenes	16.16	4.77	5.48
Edad media	13.37	2.70	8.79
Adultos	4.8	0.66	9.06

Al comparar el uso de los tres marcadores se observa una correlación directa entre la frecuencia y la edad, con aumento en la frecuencia del uso de *pues* y *o sea* una disminución en el uso de *entonces* por parte del grupo de los jóvenes. Estos datos cuantitativos muestran igualmente la dirección de las variantes a través del grupo de edad con lo que igualmente se pueden plantear hipótesis sobre su direccionalidad futura. Los resultados pueden ser una indicación de que *pues* y en menor grado *o sea* pueden estar desplazando a *entonces*, a la vez que están siendo más favorecidos entre la población más joven. La preferencia por el uso de *o sea* por los jóvenes en contraste con los hablantes mayores es obvia. De hecho, *o sea*, junto a *¿bueno?* son expresiones que a menudo identifica la generación joven en Bucaramanga. Su pronunciación es bastante notoria con una entonación más elevada y tensa en la sílaba [se] con decrecimiento de entonación bastante alto en [a].

Como se aprecia en esta última anotación, el análisis estadístico va seguido de una explicación de los resultados que puede ir acompañados de hipótesis sobre la direccionalidad de una variable, especialmente si se tiene un número adecuado de hablantes jóvenes dentro de la muestra.

Con este sucinto análisis sobre la metodología utilizada en el estudio del español de Bucaramanga se pretendió mostrar una serie de técnicas que pueden ser aplicadas en diseño de investigación de otros dialectos urbanos en el país. Se busca con esto continuar enriqueciendo el campo de la sociolingüística en Colombia, y continuar develando los rasgos de las distintas variedades urbanas en el país, que como en el caso

de Bucaramanga son desconocidas o de las cuales únicamente se han producido trabajos con técnicas tradicionales de dialectología que no ofrecen una teoría sobre la lengua como es, en su contexto social.

Referencias

- Amaya, M. P. (2000). *Así es Bucaramanga. This is Bucaramanga*. Colombia: Ediciones Gamma.
- Area Metropolitana de Bucaramanga--AMB. (2002). *Plan de Ordenamiento Territorial, Componente Físico Territorial*. Santafé de Bogotá.
- Becerra, S. (1980). 'Consonantes implosivas en el español urbano de Cartagena de Indias (Colombia): implicaciones sociolingüísticas'. En G. A. Scavnyky (Ed.), *Dialectología Hispanoamericana*, 110-112.
- Bentivoglio, P. (1980-1981). *Why canto and not yo canto?* Dissertation: Thesis (M.A.)-University of California
- _____. & A., Rosenblat. (1979). *Habla culta de Caracas*. Caracas : Ediciones de la Facultad de Humanidades y Educación, Instituto de Filología "Andrés Bello," Universidad Central de Venezuela
- Cardona, R. & A. B. Simmons. *Colombia: distribución especial de la población*. 1976. Bogota: Corporación Centro Regional de la Población.
- Cedergren, H. (1973). *The Interplay of social and linguistic factors in Panama*. Thaca, New York: Cornell University.
- Ciudad Digital. (2002). Bucaramanga.
- DANE--Departamento Administrativo Nacional de Estadísticas. (1983). Santafé de Bogotá.
- Flórez, L. (1961). 'El atlas lingüístico-etnográfico de Colombia (ALEC)': Nota Informativa. En *BICC*, 26, 77-125.
- _____. (1965). *El Español Hablado en el Departamento de Santander*. Bogotá: Instituto Caro y Cuervo.
- Labov, W. (1966). *The social stratification of New York City*. Washington D.C: Center for Applied Linguistics.
- _____. (1972). *Language in the inner city: studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.
- _____. (1972a). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- _____. (1984). 'Field Methods of the Project on Linguistic Change and Variation'. En *Language and Use: Readings in sociolinguistics*. Englewood Cliffs, NJ: Prentice Hall.
- Lastra de Suárez, Y. (1992). *Sociolingüística para Hispanoamericanos: una introducción*. Mexico. D.F.: El Colegio de Mexico, Centro de Estudios Lingüísticos y Literarios.
- Lavandera, B. (1975). *Linguistic structure and sociolinguistic conditioning in the use of verbal ending in "si" -clauses*. Proefschrift Philadelphia: University of Pennsylvania, Xerox University Microfilms.

- Molina Martos, I. (1998). *La Fonética de Toledo, Contexto Geográfico y Social*. Alcalá de Henares: Universidad de Alcalá, Servicio de Publicaciones.
- Montes, J. J. (1982). 'El español de Colombia: Propuesta de clasificación dialectal'. En *Thesaurus*, 37, 1, Jan-Apr, 23-27.
- _____. et al. (1998). *El español hablado en Bogotá: Análisis previo de su estratificación social*. Santafé de Bogotá: Instituto Caro y Cuervo.
- Murray, T. E. & Ross-Murray, C. (1996) *Under cover of law: more on the legality of surreptitious recording*. Tuscaloosa: University of Alabama Press.
- Otalora, H. (1997). *Léxico del habla culta de Santafé de Bogotá*. Presentation by J. J. Montes. Bogotá: Publicaciones del Instituto Caro y Cuervo.
- _____. (1998). 'Estudio del habla culta de Bogotá'. En *Litterae*, 7, Apr., 210-217.
- _____. & A. Gonzalez. (1986). *El habla de la ciudad de Bogotá: Materiales para su estudio*. Bogotá: Instituto Caro y Cuervo.
- Samper Padilla, J. A. (1990). *Estudio sociolingüístico del español de Las Palmas de Gran Canaria*. Santa Cruz de Tenerife: La Caja de Canarias.
- Saussure, F de et al. (1966). *Course in General Linguistics*. New York: McGraw-Hill Book Co.
- Silva-Corvalán, C. (2001). *Sociolingüística y Pragmática del español*. Georgetown University Press. Washington D.C.
- Secretaría de Desarrollo Municipal de Bucaramanga. (2002). Bucaramanga.
- Superintendencia de Servicios Públicos Domiciliarios. (2002). Santafé de Bogotá.
- Trudgill, P. (1974) *Sociolinguistics: an Introduction*. Harmondsworth : Penguin,

Abstract

The methodological model described here is based on developments in sociolinguistic research which have resulted from a constant concern for the design and implementation of techniques that reflect the language in its social context, departing from the dialectological tradition to the analysis of language. The present monograph is a description of the methodology used in the study of the main phonological, morph-syntactic, lexical and discourse features of the Spanish of Bucaramanga (Colombia). The methodological approach presented here is based on the experience of previous studies in the collection, selection, and analysis of data and methodological improvement from these experiences to better converge on the linguistic system that is to be described: the vernacular, the language in use and not an idealized system. It is a model that may be applicable to the study of other varieties of Spanish in Colombia, as an alternative accepting the view that sociolinguistics has not yet arrived at the ideal method in the study of sociolinguistic phenomena in urban contexts.

Key words: sociolinguistics, Colombian Spanish, Bucaramanga Spanish, Study methods of vernacular language, data-collecting methodology, selection and analysis of linguistic data.