

## El habla de Mérida: un corpus de estudio

**Carmen Luisa Domínguez M.**

*Departamento de Lingüística*

*Universidad de Los Andes*

### Resumen

En este artículo se presentan los fundamentos y las variables que fueron consideradas para la constitución del *Corpus sociolingüístico de la ciudad de Mérida*. Esta muestra, recogida entre 1990 y 1994 representa un conjunto de datos para el análisis del habla de Mérida en los distintos niveles de especialización de nuestra ciencia; además, refleja en su configuración la incidencia de variables sociales y extralingüísticas que pueden considerarse pertinentes para la descripción lingüística como son: *edad* de los informantes (en el corpus se distinguen cuatro grupos etáreos), *sexo* y *nivel sociocultural* (se distinguen aquí cinco grupos). Además de las variables y los criterios para su selección, se informa aquí sobre la constitución misma del corpus, el proceso de selección de los informantes, la grabación y transcripción de las cintas y se anexa un modelo de transcripción ortográfica tras las especificaciones sobre las dificultades y decisiones a que dio lugar este trabajo.

### Introducción

Cada vez que un lingüista se propone el análisis sistemático y sistémico de su objeto de estudio, debe, necesariamente, proponerse también la búsqueda de los datos que sustenten sus intuiciones y permitan la verificación científica de sus hipótesis. Esta búsqueda de los datos, sin embargo, representa una tarea larga y detallista, que no puede ser tarea de uno solo, que requiere la existencia de la infraestructura adecuada y que, muchas veces, termina convirtiéndose en obstáculo insalvable en nuestro quehacer.

Para salvar este obstáculo, los centros de investigación en lingüística se han propuesto, cada vez más consistentemente, la recolección de muestras extensas que permitan la facilitación de las investigaciones en nuestra ciencia. Con esta intención, en 1990 se inicia la recolección del *Corpus sociolingüístico de la ciudad de Mérida* bajo la responsabilidad de Elsa Mora Gallardo quien, en 1992, es relevada por mí.

El *Corpus sociolingüístico de la ciudad de Mérida* (en adelante CSM) está constituido por cuarenta horas de grabación de entrevistas semielicitadas con ochenta hablantes merideños nativos (media hora cada uno) y es el resultado de cinco años de trabajo de campo para obtener una muestra del español hablado en Mérida que permita, a los investigadores interesados, la descripción de esta variedad del español venezolano.

Una muestra de este tipo es, seguramente, de interés para los estudiosos del lenguaje en sus aspectos fonético y fonológico pues, a pesar de no estar transcrito en caracteres especiales, puede, sin duda, servir de ayuda en la localización de las realizaciones que puedan interesarles y facilitar así el recurso obligatorio a la grabación original; los estudiosos de la

morfología, sintaxis, significación, realizaciones discursivas y pragmáticas encontrarán también en esta muestra un importante apoyo para sus investigaciones.

En los apartes siguientes explicaré cómo se recogió y transcribió este corpus para el estudio del habla de Mérida.

### **1. Ubicación de los informantes. Realización de las grabaciones.**

Tal como mencionaba arriba, el *CSM* está constituido por cuarenta horas de grabación de conversaciones con ochenta hablantes merideños. Para la selección de los hablantes entrevistados se consideró el que hubieran nacido en la ciudad Mérida o bien que, habiendo nacido en el Estado, hubieran vivido gran parte de su vida en esta ciudad.

Quizá el aspecto más arduo en la constitución de una muestra es el que tiene que ver con la localización de los hablantes adecuados para ser grabados, esto es, que representen los criterios de selección que se establecen de antemano y que, además, estén dispuestos a conversar durante media hora con una persona que, por lo general, es un perfecto desconocido. En efecto, se intentó que nuestros entrevistadores no tuvieran trato familiar con los hablantes entrevistados para evitar las conversaciones en las que la familiaridad de los interlocutores impidiera luego la recuperación de la referencia a los temas de la conversación. Así, casi siempre se trata de un vecino, o de un amigo no muy cercano, y, más frecuentemente, del amigo de un familiar o amigo.

Se instruyó a los entrevistadores para que, en lo posible, no interrumpieran al entrevistado cuando este parecía estar dispuesto a la conversación, sin embargo, como en todo diálogo, en la mayoría de los casos el entrevistador no puede evitar su intervención y, con mucha frecuencia, expresa su aquiescencia con un *mjm* o *ajá* que parece inevitable.

Para eludir la mención de nuestra ciencia ante un hablante que se intimidaría o hipercorregiría al oír hablar de lingüística, se les dijo que nuestras grabaciones se interesaban por las tradiciones, costumbres y valores regionales. Es por esto que, en la muestra, tenemos las más variadas versiones de la Paradura de Niño, las tradiciones de la Semana Santa, los cuentos de aparecidos y las costumbres del Páramo merideño; algunas veces el hablante prefirió contar sus propias experiencias en la ciudad.

En algunos casos hubo que volver a grabar, y esto por dos razones, a) las razones relativas al hablante: pues este no se adecuaba a los criterios de selección establecidos, o porque, en el transcurso de la grabación, el entrevistador se percataba de un serio problema de inteligibilidad del hablante (por mala pronunciación, defectos o carencia en la dentadura, etc.), en este caso se buscaba a otra persona; y b) las razones relativas a la grabación: estas resultaban inaudibles por exceso de ruido en el entorno, en ese caso se volvía a contactar a la misma persona y se trataba de grabar de nuevo en mejores condiciones (lo cual pudo ser posible en la mayoría de estos casos).

### **2. Las variables extralingüísticas. El cálculo de los índices**

Una vez seleccionados, los hablantes fueron agrupados de acuerdo con las variables que se indican a continuación:

- i) Sexo: 40 hombres y 40 mujeres.
- ii) Edad: cuatro grupos generacionales, a saber:
  - A: 14 a 29 años: 20 hablantes;
  - B: 30 a 45 años: 20 hablantes;
  - C: 46 a 60 años: 20 hablantes;
  - D: 61 años o más: 20 hablantes.
- iii) Nivel socioeconómico: los hablantes fueron distribuidos en cinco grupos (16 hablantes en cada uno), de acuerdo con el nivel socio-económico. Estos grupos se identifican así:
  - 1: alto;
  - 2: medio-alto;
  - 3: medio;
  - 4: medio-bajo;
  - 5: bajo.

Siguiendo estos criterios se fueron llenando cuarenta casillas, con dos hablantes en cada una.

Para identificar las variables extralingüísticas según las cuales se agrupó a los hablantes en niveles socioeconómicos, se siguieron los siguientes parámetros:

Variables socioeconómicas (aplicadas a la ocupación del hablante, la de su padre y la de su madre)

- 1 Altos funcionarios del gobierno; altos oficiales del ejército; empresarios privados; hacendados; altos ejecutivos (sectores público y privado); autoridades universitarias.
- 3 Profesionales universitarios de libre ejercicio; gerentes medios del sector público y privado; oficiales de rango medio; industriales y productores medios; docentes universitarios; artistas reconocidos.
- 5 Profesionales universitarios no liberales; profesores de educación media y básica; oficiales de rango bajo; pequeños empresarios y productores; técnicos superiores; secretarías ejecutivas; supervisores; enfermeras graduadas; miembros de la farándula.
- 7 Pequeños comerciantes; secretarías y oficinistas; obreros especializados; artesanos; mecánicos; vendedores; cobradores; ayudantes técnicos; policías y agentes de tránsito; deportistas profesionales; regulares de las fuerzas armadas.
- 9 Buhoneros y vendedores ambulantes; obreros no especializados; servicio doméstico; mesoneros, bedeles y vigilantes.

Estudios

- 1 Doctorado
- 2 Maestría
- 3 Pregrado universitario completo
- 4 Pregrado incompleto / Técnico superior
- 5 Secundaria completa / Carreras técnicas
- 6 Secundaria incompleta / Cursos de capacitación
- 7 Primaria completa
- 8 Primaria incompleta

9 Analfabeta

Condiciones de alojamiento

- 1 Casa o apartamento lujoso o espacioso
- 3 Casa o apartamento menos lujoso o espacioso
- 5 Casa o apartamento sin lujo
- 7 Casa o apartamento modesto
- 9 Vivienda sin comodidades sanitarias y de difícil acceso

Ingreso total familiar (julio 1993).

- 1 305.000 o más
- 3 130.000 a 305.000
- 5 35.000 a 130.000
- 7 12.000 a 35.000
- 9 0 a 12.000

Ingreso familiar promedio (julio 1993).

- 1 75.000 o más
- 3 28.000 a 75.000
- 5 6.800 a 28.000
- 7 4.000 a 6.800
- 9 0 a 4.000

Para calcular los índices que se asignaban a cada hablante y, con ello, el nivel socioeconómico en el cual se agruparía a este, se procedía como sigue:

a) se asignaban los puntos correspondientes (columna de la izquierda) de acuerdo con los datos del hablante que coincidían con los datos extralingüísticos asociados a cada punto (columna de la derecha), y ello en cada una de las variables; b) se multiplicaban estos puntos por los pesos de ponderación asignados a cada variable, estos son:

- Ocupación del hablante x 0.12
- Ocupación del padre x 0.12
- Ocupación de la madre x 0.14
- Estudios x 0.10
- Vivienda x 0.15
- Ingreso total x 0.19
- Ingreso promedio x 0.18;

c) los resultados de las multiplicaciones anteriores se sumaban y el resultado correspondía al índice socioeconómico del hablante. Este índice, a su vez, se cotejaba con la escala que se presenta enseguida, para obtener entonces el nivel socioeconómico del hablante.

- alto 1.00 - 2.60
- medio-alto 2.61 - 4.20
- medio 4.21 - 5.80

medio-bajo 5.81 - 7.40  
bajo 7.41 - 9.00

Así, por ejemplo, si el resultado del cálculo era de 4.84 (como en el caso del hablante identificado con el código MDA3MA) se consideraba que este hablante pertenecía al nivel medio.

### **3. Los códigos de identificación**

Todas las grabaciones y transcripciones se identificaron de tal manera que, en el código asignado, se representarían las variables extralingüísticas que correspondían a cada hablante. Así,

**MD** Mérida

**A, B, C, D** Grupo generacional

**1, 2, 3, 4, 5** Nivel socioeconómico

**M, F** Sexo

**A, B** Ubicación del hablante en la casilla

De esta manera, MDA3MA, por ejemplo, debe leerse como sigue:

**MD** Mérida;

**A** primer grupo generacional;

**3** nivel socioeconómico medio;

**M** sexo masculino;

**A** primer hablante en la casilla.

El *CSM* quedó entonces organizado tal como puede verse en la tabla 1.

Tabla 1. *Corpus sociolingüístico de Mérida*: distribución de los hablantes de acuerdo a las variables extralingüísticas.

NS		GG I (A)	GG II (B)	GG III (C)	GG IV (D)
1	m	MDA1MA MDA1MB	MDB1MA MDB1MB	MDC1MA MDC1MB	MDD1MA MDD1MB
1	f	MDA1FA MDA1FB	MDB1FA MDB1FB	MDC1FA MDC1FB	MDD1FA MDD1FB
2	m	MDA2MA MDA2MB	MDB2MA MDB2MB	MDC2MA MDC2MB	MDD2MA MDD2MB
2	f	MDA2FA MDA2FB	MDB2FA MDB2FB	MDC2FA MDC2FB	MDD2FA MDD2FB
3	m	MDA3MA MDA3MB	MDB3MA MDB3MB	MDC3MA MDC3MB	MDD3MA MDD3MB
3	f	MDA3FA MDA3FB	MDB3FA MDB3FB	MDC3FA MDC3FB	MDD3FA MDD3FB
4	m	MDA4MA MDA4MB	MDB4MA MDB4MB	MDC4MA MDC4MB	MDD4MA MDD4MB
4	f	MDA4FA MDA4FB	MDB4FA MDB4FB	MDC4FA MDC4FB	MDD4FA MDD4FB
5	m	MDA5MA MDA5MB	MDB5MA MDB5MB	MDC5MA MDC5MB	MDD5MA MDD5MB
5	f	MDA5FA MDA5FB	MDB5FA MDB5FB	MDC5FA MDC5FB	MDD5FA MDD5FB

#### 4. La transcripción

Las grabaciones fueron transcritas de manera ortográfica. Tal como indicaba en la Introducción, me parece que de esta manera se ofrece un instrumento adecuado para que los investigadores tengan acceso al discurso de los hablantes y puedan identificar los fenómenos de su interés más rápidamente. Todo lingüista sabe que esto no basta, que debe recurrirse a la grabación en la casi totalidad de los casos, con mayor razón si se trata del análisis del segmento o del suprasegmento fónico, ahora bien, también se sabe que es imposible recuperar todas las apariciones de un determinado fenómeno solamente

escuchando la cinta, es necesario otro soporte (papel o disquete) que permita la revisión de la totalidad del texto sin necesidad de largas horas de escucha. Así, resulta evidente el que estas transcripciones son "genéricas" y que cada investigador las adaptará a sus intenciones de trabajo, esto, sin embargo, no disminuye la convicción de que una transcripción de este tipo es una ayuda inmensa para iniciar la investigación. Cualquier transcripción, "genérica" o no, traiciona, de una u otra manera, el discurso oral. Por el contrario, si algo se pone en evidencia a través de una transcripción es el punto de vista de quien la hace, transcribir revela no solo nuestra condición de hablantes de una lengua sino también nuestra condición de analistas y nuestras convicciones en lingüística. Así pues, no hay forma de poner en el papel todos los matices, variaciones, énfasis e inflexiones del discurso oral. Tratar de hacerlo, por otra parte, resultaría en un texto tan lleno de marcas, guías y señales que sería entonces ilegible y, por eso, inútil. La fidelidad de la transcripción del habla es un mito, no hay una técnica que permita recoger toda la riqueza de la oralidad. El recurso a las grabaciones será más necesario cada vez y cada investigador oír en ellas los rasgos específicos que interesan a su trabajo. Será él quien haga entonces esas marcas, también específicas. La transcripción es solo un apoyo inicial en el análisis. Lo dicho anteriormente no significa, sin embargo, que hayamos copiado las palabras del hablante sin otra indicación. Esto, de nuevo, haría ilegibles los textos y, por eso, inútiles. Para el *CSM*, hemos seguido los siguientes criterios de transcripción:

a. transcripción ortográfica utilizando los signos de puntuación de manera convencional: las pausas o cambios en la curva de entonación se marcan con comas o puntos de acuerdo con la longitud de la pausa y con el sentido de lo que se dice, se usan los signos de interrogación y de admiración convencionalmente; b. transcripción de hesitaciones y falsos arranques seguidos de puntos suspensivos; c. las repeticiones del hablante se copian todas de seguido; d. se usan los corchetes dobles para señalar los sonidos no lingüísticos: [[risas, toses, ruidos, intervención de terceras personas, diálogos del hablante con terceras personas, interrupciones en la grabación]]; e. paréntesis exclusivamente para indicar que (*no se entiende*) una palabra o una cadena indeterminada; f. transcripción convencional de las interjecciones: *ajá, mjm, mm, uf, uy, ay, ey*; g. los extranjerismos se transcriben de acuerdo con el DRAE cuando este los ha incluido con una ortografía hispanizada, cuando no es este el caso entonces se transcriben de acuerdo con la ortografía original y se hace una notación fonética entre corchetes, como en *country club* [kontriklú], *full* [fúl]; h. los corchetes se utilizan también en casos como los de *base ball* o *foot ball* que, de acuerdo con el DRAE, se escriben *béisbol* y *fútbol* y que los diferentes hablantes de la muestra pueden leer como [béisbol] o [beisból], o bien, [fútbol] o [futból]; i. igualmente, se usan los corchetes para indicar la realización efectiva del hablante en los casos de discordancia con la norma general, por ejemplo en *diabetes* [diábetes], *desavenencia* [desabeniéncia], *inaugurar* [inagurár]; j. se copia el uso del hablante cuando este implica alteración dialectal de la morfología de la palabra, como en *íbanos* o *haiga*; k. se copia exactamente lo que dice el hablante cuando usa expresiones populares en las cuales la morfología de alguna palabra se ha alterado, por ejemplo, *darle un parao, ojo pelao, echar pa'lante*; l. se copia solo la inicial del nombre de las personas citadas por el hablante a menos que se trate de personalidades reconocidas; m. no se comenta la producción de los hablantes mediante el uso de *sic* ni se "corrige" lo que el hablante dice: inconcordancias, uso incorrecto de los términos, etc. n. se hace la separación de las voces superpuestas cuando las hay.

Las transcripciones se realizaron con la ayuda de un equipo SONY BM810 que permite retroceder y reescuchar fácilmente la cinta. Este trabajo se hizo directamente en un procesador de palabras (*Word Perfect 5.1* para DOS), de modo que fuera más fácil la realización de cualquier corrección necesaria y, de hecho, fueron necesarias varias correcciones sobre cada transcripción: i) la primera consistió, simplemente, en "pasar" el corrector ortográfico del *Word Perfect* sobre cada texto. Este procedimiento nos permitió empezar a notar algunos problemas de la transcripción como, por ejemplo, el que en nuestra versión original habíamos copiado exactamente la realización del hablante y no lo que en efecto decía. Fue así como decidimos empezar a distinguir entre lo que decía (y esto se copió en el texto) y lo que "pronuncia" (esto se copió entre corchetes simples), cfr. los criterios de transcripción h-i. ii) se realizó una segunda corrección que consistió en la reescucha de toda la grabación para revisar las decisiones del transcriptor y aplicar los nuevos criterios. Esta segunda escucha estuvo a cargo de alguno de los asistentes, con la condición de que no hubiera sido el transcriptor de la primera versión. iii) hubo una tercera, y última, corrección que consistió en la reescucha de todas las grabaciones para revisar finalmente la "fidelidad" de la transcripción y la consistencia en la aplicación de los criterios. Esta última fue realizada en todos los casos por mí. Para ejemplificar ampliamente el resultado de la aplicación de estos criterios aquí se incluye, en anexo, la transcripción integral de la grabación realizada a ese hablante que ya he mencionado varias veces en este artículo: MDA3MA (ver anexo I).

### 5. Para terminar

Me parece que este largo relato no hace muy necesario insistir en aquello que señalaba al principio en relación con el tamaño y las exigencias de la tarea de recoger una muestra de habla que permita sustentar la investigación. Parafraseando a Bentivoglio 1996 "la muestra no es la investigación, pero la investigación presupone la muestra", sobre todo si, como ahora, nos proponemos una descripción cabal de nuestra variedad del español. Por suerte, para este trabajo no estamos solos. El proyecto de recolección del CSM se inició para responder a una iniciativa colectiva empeñada en producir la Gramática de Referencia del Español Hablado en Venezuela (GREHV). Esta iniciativa, que propone la descripción sistemática, sincrónica y funcional del habla nacional, basada en una muestra que incluya la variable sociolingüística, es ahora el reto al que debemos responder. Los datos están disponibles, pero los datos son solo el comienzo.

### Referencias bibliográficas

- Bentivoglio, Paola y Mercedes Sedano. 1993. Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana. En *Boletín de lingüística* 8 (3-35). Caracas: Universidad Central de Venezuela.
- Bentivoglio, Paola. 1996. La gramática del español hablado en Venezuela. En *Lengua y Habla* 1,1 (21-39). Mérida: Universidad de Los Andes.
- Real Academia Española. 1992. *Diccionario de la lengua española*. Madrid: Espasa-Calpe.