

# *Corpus sociolingüístico del habla de Mérida-Venezuela*

Carmen Luisa Domínguez y Elsa Mora (coordinadoras).

Depósito Legal: If 2371998400562  
ISBN: 980-11-0208-X

## **Criterios para la constitución del *Corpus sociolingüístico de la ciudad de Mérida-Venezuela*<sup>1</sup>**

(por Carmen Luisa Domínguez Mujica)

### Introducción

1. Ubicación de los hablantes. Realización de las grabaciones.
  2. Las variables extralingüísticas. El cálculo de los índices.
  3. Los códigos de identificación.
  4. La transcripción.
- Referencias.

Cada vez que un lingüista se propone el análisis sistemático y sistémico de su objeto de estudio debe proponerse también, necesariamente, la búsqueda de los datos que sustenten sus intuiciones y permitan la verificación científica de sus hipótesis. Esta búsqueda, 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, se convierte en obstáculo insalvable en nuestro quehacer. Así, muchas veces nos conformamos con estudios-sondeo, a la espera de ser corroborados en una muestra mayor, o bien, desistimos.

Para salvar este obstáculo, los centros de investigación en lingüística se han propuesto, cada vez más consistentemente, la recolección de muestras extensas que faciliten las investigaciones en nuestra ciencia. Con esta intención, en 1990 se inicia la

---

<sup>1</sup> La recolección del *Corpus sociolingüístico de la ciudad de Mérida* fue posible gracias al financiamiento otorgado por el Consejo de Desarrollo Científico y Humanístico de la Universidad de Los Andes a través de los proyectos H-152-89 y H-226-92. La publicación en papel (1998) fue financiada por el mismo Consejo a través del proyecto H-670-02-06-A.

recolección del *Corpus sociolingüístico de la ciudad de Mérida*.<sup>2</sup> Este corpus está constituido por cuarenta horas de grabación de entrevistas semielicitadas con ochenta hablantes merideños nativos (media hora cada uno) y es el resultado de cuatro 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. La distribución de estos ochenta hablantes puede verse en el cuadro siguiente:

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

<sup>2</sup> El *Corpus sociolingüístico de la ciudad de Mérida*, en sus varios formatos (cintas, papel, disquetes, CD) se encuentra, en su totalidad, en los archivos del Grupo de Lingüística Hispánica de la Universidad de Los Andes en Mérida-Venezuela.

<sup>3</sup> La clave para la asignación de estos códigos se presenta más adelante, en la exposición de los criterios de selección y clasificación de los hablantes.

Hemos decidido difundir, por este medio, la transcripción de una selección de la muestra general, constituida por veinticuatro de esas entrevistas, escogidas de manera que sean representativas de las variables extralingüísticas que hemos tomado en cuenta en la constitución del *Corpus*. En el cuadro siguiente aparece la distribución de estos 24 hablantes identificados esta vez por el número con el cual aparecen en esta edición electrónica:

Nivel →	ALTO		MEDIO		BAJO	
	M	F	M	F	M	F
Grupo A (14 a 29 años)	1	2	3	4	5	6
Grupo B (30 a 45 años)	7	8	9	10	11	12
Grupo C (46 a 60 años)	13	14	15	16	17	18
Grupo D (60 años ...)	19	20	21	22	23	24

Pensamos que los datos que presentamos aquí serán de interés para los estudiosos del lenguaje en sus aspectos fonético y fonológico pues, a pesar de no estar transcritos en caracteres especiales, puede, sin duda, servir de ayuda en la localización de los contextos que puedan interesarles para facilitar el recurso obligatorio a la grabación original; los estudiosos de la morfología, sintaxis, significación, realizaciones discursivas y pragmáticas encontrarán aquí, esperamos, un importante apoyo para sus investigaciones.

#### 1. UBICACIÓN DE LOS HABLANTES. REALIZACIÓN DE LAS GRABACIONES.

Tal como mencionamos en la Introducción, el *Corpus sociolingüístico de la ciudad de Mérida* 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, en general, es un perfecto desconocido. En efecto, se trató 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 tratados. 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 evitar 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. Por esto, 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) porque el hablante no se adecuaba a los criterios de selección establecidos, o porque, en el transcurso de la grabación, nuestro 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; b) porque la grabación resultaba inaudible 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. Tal como ya mencionamos, aquí se presentan las transcripciones correspondientes a veinticuatro hablantes, es decir, 12 hombres y 12 mujeres, seis hablantes por grupo de edad y ocho hablantes del nivel alto, ocho del nivel medio y ocho del nivel bajo.

Para identificar las variables extralingüísticas según las cuales se agrupó a los hablantes en niveles socioeconómicos, se siguieron los siguientes parámetros:<sup>4</sup>

*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.

---

<sup>4</sup> Estos parámetros coinciden con los utilizados por Paola Bentivoglio y Mercedes Sedano para la constitución del *Corpus sociolingüístico del habla de Caracas* en 1987 (cfr. Bentivoglio y Sedano 1993). A su vez, el método para calcular el índice socioeconómico fue elaborado por Max Contasti, profesor de la Universidad Central de Venezuela y la Universidad Nacional Abierta.

*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).<sup>5</sup>*

- 1 305.001 o más
- 3 130.001 a 305.000
- 5 35.001a 130.000
- 7 12.001 a 35.000
- 9 hasta 12.000

*Ingreso familiar promedio (julio 1993).<sup>6</sup>*

- 1 75.001 o más
- 3 28.001 a 75.000
- 5 6.801 a 28.000
- 7 4.001 a 6.800
- 9 hasta 4.000

---

<sup>5</sup> En la primera etapa de este proyecto se utilizaron los valores de cálculo que eran adecuados para esa época. Estos son: **1:** 38.000 Bs. o más; **3:** 14.000 a 38.000; **5:** 7.100 a 14.000; **7:** 2.600 a 7.100; **9:** 0,00 a 2.600.

<sup>6</sup> Para el ingreso promedio, los valores asignados en la primera etapa fueron: **1:** 6.100 Bs.. o más; **3:** 2.200 a 6.100; **5:** 1.280 a 2.200; **7:** 570 a 1.280; **9:** 0,00 a 499.

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 se consideraba que este hablante pertenecía al nivel medio.

### 3. LOS CÓDIGOS DE IDENTIFICACIÓN.

En la muestra total, 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,

<b>MD</b>	Mérida
<b>A, B, C, D</b>	Grupo generacional
<b>1, 2, 3, 4, 5</b>	Nivel socioeconómico
<b>M, F</b>	Sexo
<b>A, B</b>	Ubicación del hablante en la casilla

De esta manera, MDA3MA debe leerse como sigue:

<b>MD</b>	Mérida;
<b>A</b>	primer grupo generacional;
<b>3</b>	nivel socioeconómico medio;
<b>M</b>	sexo masculino;
<b>A</b>	primer hablante en la casilla

### 4. LA TRANSCRIPCIÓN.

Las grabaciones fueron transcritas de manera ortográfica. Tal como indicábamos en la Introducción, pensamos 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 de 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, nos resulta evidente el que estas transcripciones son "genéricas" y que cada investigador las adaptará a sus intenciones de trabajo, pero también sabemos 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. 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 necesario y cada investigador oirá en ellas los rasgos específicos que interesan a su trabajo y hará 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 *Corpus sociolingüístico de la ciudad de Mérida*, 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 usando los 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*);
- 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 enseguida una notación entre corchetes que represente lo mejor posible la realización efectiva del hablante, 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 diferentes hablantes 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 *ibanos* 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.

#### REFERENCIAS

- 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.
- Real Academia Española. 1992. *Diccionario de la lengua española*. Madrid: Espasa-Calpe.