

Received 7 January 2019.

Accepted 15 March 2019.

CORDIA-VEN-sXX: UN SUBCORPUS DIALECTAL PARA EL ESTUDIO SOCIOLINGÜÍSTICO DEL ESPAÑOL HABLADO EN VENEZUELA

Krístel GUIRADO

Universidad Central de Venezuela-Universidad de Zaragoza *

kristelguirado@gmail.com

Resumen

La *Reingeniería de Corpus* es el proceso de reconfiguración de muestras de habla para su reutilización en diversos ámbitos (Guirado 2014, 2015, 2018). Esta investigación propone crear un subcorpus a partir de la reingeniería de corpus orales de tres ciudades venezolanas (Maracaibo, Caracas y Mérida), para realizar estudios dialectales en una microdiacronía del siglo XX. La metodología para el rediseño incluyó: i. describir la estructuración de los corpus originales; ii. evaluar los aspectos estables e inoperantes; iii. crear el nuevo diseño y estimar su representatividad. Como resultado se obtuvo el CORDIA-VEN-sXX, el cual consta de 144 entrevistas, distribuidas equitativamente según metrópolis, edad, nivel socioeconómico y sexo. En conclusión, la Reingeniería de Corpus resulta una práctica eficaz dentro de la Lingüística de Corpus para la diversificación de algunos materiales en nuevos contextos interpretativos.

Palabras clave

Reingeniería de Corpus, habla urbana en Venezuela, microdiacronía, Lingüística de Corpus

* Instituto de Filología "Andrés Bello", Biblioteca Central, Piso 11, Ciudad Universitaria, Caracas, 1050 / Departamento de Lingüística y Literatura Hispánicas, Despacho 431, calle San Juan Bosco, 7, Zaragoza, 50009.

CORDIA-VEN-SXX: A DIALECTAL SUBCORPUS FOR THE SOCIOLINGUISTIC STUDY OF SPANISH SPOKEN IN VENEZUELA

Abstract

Corpus Reengineering is the process of reconfiguration of speech samples for reuse in various scopes (Guirado 2014, 2015, 2018). This research proposes to create a subcorpus from the reengineering of oral corpora from three Venezuelan cities (Maracaibo, Caracas and Mérida), to carry out dialect studies in a twentieth century microdiachrony. The methodology for the redesign included: i. to describe the structuring of the original corpora; ii. to evaluate the stable and inoperative aspects; iii. to create the new design and estimate its representativeness. As a result, the CORDIA-VEN-SXX was obtained, which consists of 144 interviews, distributed equally according to metropolis, age, socioeconomic level and sex. In conclusion, *Corpus Reengineering* is an effective practice within Corpus Linguistics for the diversification of some materials in new interpretive contexts.

Keywords

Corpus reengineering, Spanish spoken in Venezuela, Grammar of spoken Spanish, corpus linguistic

1. El estudio dialectal del habla venezolana y la reingeniería de corpus

1.1 *El dialecto como unidad de análisis en el habla venezolana*

Hymes (1974) habla de dialecto como una “unidad local” caracterizada por la “interacción primaria” de sus miembros y Milroy (1980) lo propone como “una unidad de análisis”. Los límites del dialecto pueden ser locales, regionales, nacionales o incluso supranacionales. Moreno Fernández (2005) considera que el habla de una comunidad, circunscrita a una zona geográfica, puede ser considerada un *dialecto* si coinciden determinados factores de carácter extralingüístico en ella y si sus miembros se identifican y se sienten como integrantes de una condición o categoría social; esta conciencia debe extenderse a la actitud, es decir, los hablantes deben considerar que existe una substancial diferencia entre su variedad y otras. Sedano (2004) señala, además, la

necesidad de tomar en cuenta todo lo asociado con la prosodia, en particular con la entonación, pues esta resulta fundamental, en muchos casos, para identificar un dialecto.

El español que se habla en distintas regiones de Venezuela no corresponde a una única área dialectal. Armas & Céspedes ([1882]1977), Henríquez Ureña (1921) y Rona (1964), entre otros, coinciden en señalar que la variedad incluye dos zonas claramente diferenciadas: una andina y otra caribeña.¹ Posteriormente, Rona caracterizó una tercera zona dialectal correspondiente al Zulia.² Álvarez, Obediente & Rojas (2009: 8) también suscriben la división en dos grandes áreas, pero reconocen que la llamada *caribeña* aunque unificada dialectalmente por “rasgos típicos” asociados a fenómenos fónicos segmentales, presenta diferencias “en por lo menos dos aspectos fundamentales: en parte del léxico [...] y en la prosodia”. Por esta razón, proponen una subdivisión en dos subáreas, costera (Zulia, Falcón, centro y oriente) y del interior (centro-occidente, llanos y Guayana). Además de las clasificaciones mencionadas, se han propuesto otras divisiones que oscilan entre cuatro y siete zonas y en todas se mencionan las regiones central, andina y zuliana;³ en la ilustración 1 se observa la ubicación de las tres regiones respecto al mapa de Venezuela:

¹ Esta división dialectal coincide, en mayor o menor grado, con la división según el *origen de los colonizadores* de Rosenblat (1965). De acuerdo con los planteamientos de Zamora & Guitart (1982), el área andina formaría parte de los llamados dialectos conservadores y el área caribeña de los dialectos radicales. Desde el punto de vista fonético segmental, ambos se diferencian, por ejemplo, en la forma de pronunciar las consonantes en posición implosiva (final de sílaba): los dialectos conservadores las pronuncian de una manera muy similar a sus realizaciones prenucleares (comienzo de sílaba); mientras que, los radicales las pronuncian o de manera muy diferente a sus realizaciones prenucleares o las eliden.

² Serrón (1972: 15) reseña que: “al dictar un curso de “Morfosintaxis Dialectal” en el Instituto Pedagógico de Caracas (agosto 1968) señaló una nueva zona que correspondería al Zulia y así caracterizada: *yeísmo*: sí; *zeísmo*: no; *voseo*: sí; tipo A (-áis, -éis, -ís)”.

³ Cf. Alvarado ([1929]1955); Obregón (1981); Páez Urdaneta (1981); Tejera (1992); Mora (1997); Obediente (1998); Pérez (1998); Martínez & Alario (2004); Castro & Malaver (2016).

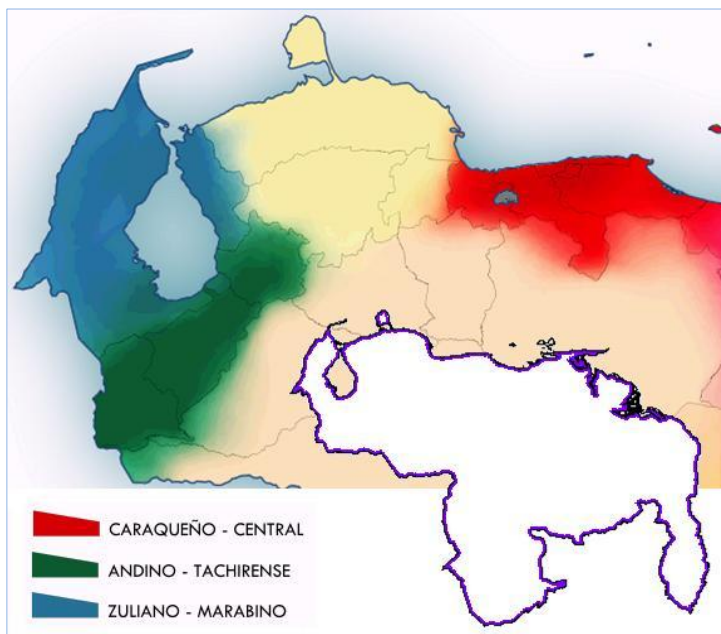


Ilustración 1. Ubicación de los dialectos caraqueño, andino y zuliano.

En la actualidad, las diferencias podrían estar siendo mitigadas por la acción uniformadora de los medios de comunicación, las redes sociales, las TIC y por la continua movilidad poblacional hacia las urbes industrializadas del centro del país; no obstante, el reconocimiento de estas tres zonas dialectales se mantiene en la actualidad, tal como lo demuestra el reciente estudio perceptivo de Castro & Malaver (2016: 40): “El hecho de que la región zuliana y la región de los Andes sean las zonas más percibidas, se explica, en nuestra opinión, porque son zonas que poseen rasgos muy diferenciados respecto del habla caraqueña”. En la ilustración 2 se muestra el ejemplo de un hablante que identifica las tres regiones en el mapa mudo de Venezuela:

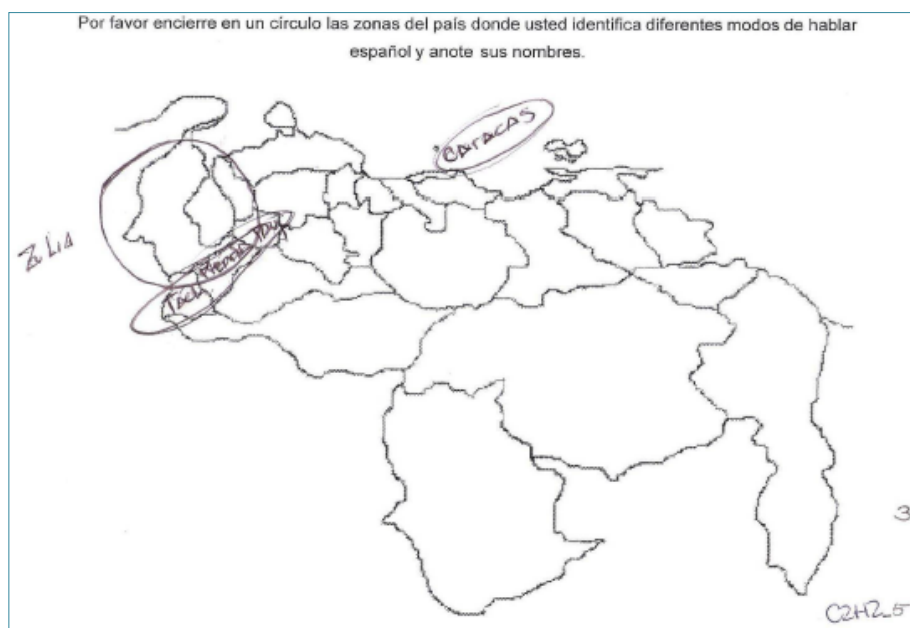


Ilustración 2. Mapa mudo en el que se reconocen las tres zonas dialectales (cf. Castro & Malaver 2016: 33).

El concepto de *región histórica* de Briceño de Bermúdez (2007: 120-121) resulta útil a la hora de justificar la independencia dialectal de estas zonas:

Llamaremos *región histórica* al espacio que gira en torno de una ciudad, con una dinámica propia que lo hace singular y coherente, con una personalidad definida, cuyos habitantes han tejido vínculos y nexos de solidaridad a lo largo del tiempo, así como un sistema de valores y representaciones sociales, que se expresan en [...] los nexos de pertenencia, las afinidades, las prácticas sociales, lingüísticas, internalizadas en el transcurrir del tiempo.

Asimismo, Álvarez (2007: 68) señala que podemos asumir heurísticamente una región como una unidad de análisis si “esa comunidad se cohesiona y se identifica a través de sus expresiones espirituales, de su mundo cultural, de su «mentalidad colectiva» expresada en sus valores, su lengua, sus costumbres [...] solidificada muchas veces con un claro sentido de pertenencia”.

Cardozo Galué (2013: 48) comenta que las regiones históricas venezolanas: “se formaron en la larga duración desde el remoto pasado aborigen [...] tuvieron existencia autónoma durante el período hispánico [...] y se consolidaron en el período republicano”.

Briceño Irigorry (1982: 51) advierte sobre el alcance que adquiere la división en jurisdicciones, “si se mira la significación que aquel antiguo aislamiento y la sujeción de unas ciudades a otras llegaron a tener en relación con la vida interior de la Colonia”. Rosenblat (1989) también observa que, para la época, Venezuela no era más que un conjunto de gobernaciones o provincias que dependían en lo político, militar y jurídico, unas veces del Nuevo Reino de Granada y otras de la Audiencia de Santo Domingo. Tejera (199: 2131) señala que la división geopolítica que se mantiene hasta finales del siglo XVIII será determinante “para las variaciones que tendrá el español en estas tierras y que le dará su carácter especial”.⁴ Los siguientes mapas ilustran la división en provincias para el año 1840:

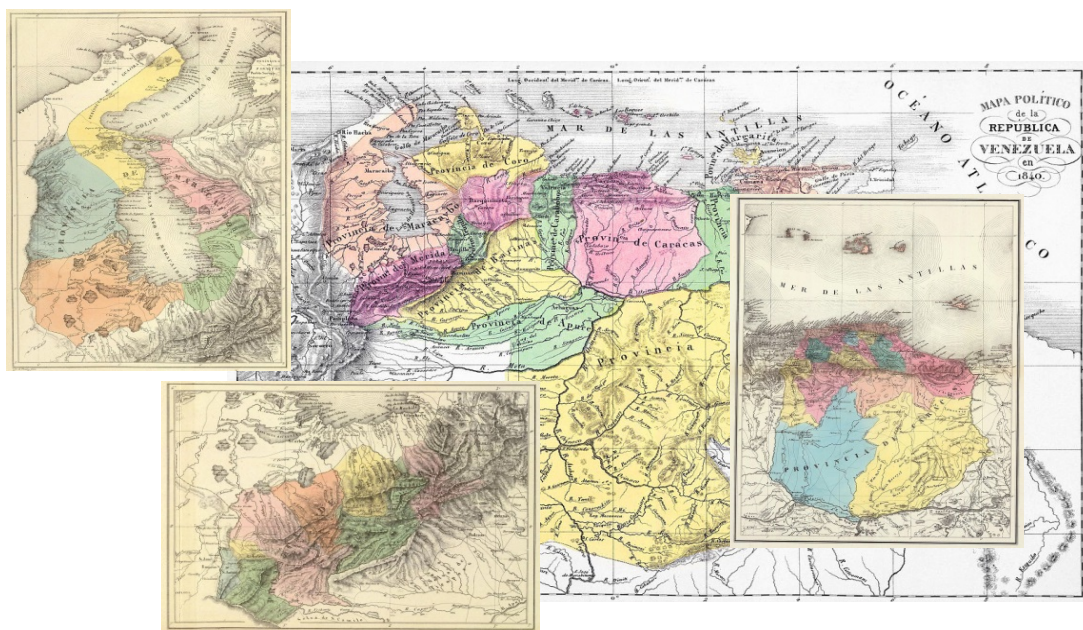


Ilustración 3. Conjunto de mapas de Venezuela en 1840 y provincias de Maracaibo, Mérida y Caracas (Codazzi 1840).

⁴ “En 1528, el rey de España, Carlos V, delimita la Provincia de Venezuela entre el Cabo de la Vela y el Cabo de Macarapana. Este territorio le es cedido a los Welser para explotarlo bajo el mando de Ambrosio Alfinger, el primer gobernador. Ya se había constituido –en 1525– la Gobernación de Margarita; después, en 1568, se fundara la de Nueva Andalucía, que en el siglo XVIII se llamó Provincia de Cumana. Más tarde (1570 y 1608) se delimita la Provincia de La Grita, que posteriormente se anexará a Mérida para constituir la Provincia y Capitanía General de Mérida, en 1622. Hacia 1625 se deslinda la Provincia de Maracaibo y en 1762 se separa la de Guayana de la de Cumana. Básicamente, esta misma división en provincias y capitanías generales se mantiene hasta la Independencia” (Tejera 1991: 213).

De esta forma, es posible suponer que la autonomía política y los accidentes geográficos que dificultaban la comunicación entre las provincias contribuyeron a dibujar la actual fisonomía dialectal de buena parte de las regiones históricas del país.

Finalmente, las variedades dialectales, además de contribuir a determinar la región como unidad de análisis geohistórica, puede operar como una categoría de análisis sociolingüística. Moreno Fernández (2005: 23-24) entiende por *comunidad de habla* aquella en la que sus hablantes “comparten efectivamente, al menos una lengua” y “un conjunto de normas y valores de naturaleza sociolingüística [...] una comunidad de consenso, de sintonía entre grupos e individuos diferentes, donde el conflicto está minimizado”.

En lo que concierne a esta propuesta, interesan especialmente las comunidades de habla marabina, caraqueña y merideña en virtud de que:

- i. Cada comunidad de habla hace vida en una ciudad ubicada en una de las zonas dialectales ya mencionadas: Maracaibo, región zuliana; Caracas, región central; Mérida, región andina.
- ii. Cada ciudad es capital de una entidad político-territorial (estado Zulia, Distrito Capital y estado Mérida, respectivamente).
- iii. Cada ciudad puede ser considerada una entidad de influencia con zona metropolitana, en virtud de la población censada durante el periodo que se propone para el corpus.⁵
- iv. Asimismo, cada comunidad presenta diferencias fonetológicas, morfosintácticas y léxicas en el uso de la lengua; principalmente, las caracteriza un “acento”, “deje” o “cantaíto” específico,⁶ y sus hablantes manifiestan una fuerte identidad regional.
- v. Por último –y fundamental para su estudio–, de cada una se dispone de un corpus de habla elaborado y transcrito bajo una metodología sociolingüística semejante.⁷

⁵ En los censos de población y vivienda de Venezuela de 1981 y 1991, “a las ciudades capitales de entidades de 50.000 y más habitantes se les han definido áreas metropolitanas. Para el año 1990 existen 19 áreas metropolitanas” (Ramírez 1991: 148). Caracas es la primera, seguida en importancia por Maracaibo, mientras que, para entonces, Mérida ocupa el décimo lugar.

⁶ Álvarez, Obediente & Rojas (2009) explican que este rasgo prosódico “resultó ser un fenómeno más complejo en el que se juntan el *recorrido tonal* o la *curva de F0* + la *duración silábica* en un todo indisoluble que llamamos *entonación*, compleja realidad acústica que da origen a un ritmo determinado que es lo que se percibe como el tonillo propio de cada variedad dialectal” (20). Para una consulta más amplia sobre el tema, cf. Álvarez (1999), Álvarez & Medina (1999), Obediente (1999), Álvarez, Martínez & Urdaneta (2001).

⁷ Con esto, no se pretende afirmar que el estudio del dialecto de otras regiones no sea también de suma importancia para la descripción del español hablado en Venezuela. Además de los citados, existen corpus

1.2 El proyecto Gramática del Español hablado en Venezuela (GREHV)

En 1987 Paola Bentivoglio tuvo conocimiento del proyecto de la *Gramática do português falado*: “cuya base documental es el inmenso corpus grabado en cinco ciudades del país (Porto Alegre, Recife, Rio de Janeiro, Salvador, y São Paulo) para el proyecto del habla culta” (Bentivoglio 1996: 24). Con esta idea, la referencia de gramáticas similares en inglés (Quirk, Greenbaum, Leech & Svartvik 1985) e italiano (Renzi, Salvi & Cardinaletti 1988) y la existencia de muestras de habla recogidas con criterios sociolingüísticos similares en varias ciudades del país,⁸ la investigadora le propone a la comunidad lingüística nacional: “producir una gramática sincrónica (período 1986-1995) del español hablado en Venezuela, basada en grabaciones de habla según una metodología sociolingüística” (Bentivoglio 1996: 27).⁹

Los equipos de investigación adscritos al proyecto GREHV siguieron los mismos criterios técnicos de grabación y normas de transcripción en las entrevistas, así como parámetros similares de estratificación social de las muestras. La metodología empleada se puede resumir como sigue:

- i. Tipo de entrevista: conversación semidirigida de 30 minutos, entre un hablante y dos investigadores según lo descrito en Silva Corvalán (1989).
- ii. Distribución de las entrevistas: equitativa entre cuatro grupos etarios (15-29, 30-45, 46-60, 61 o más años), cinco niveles socioeconómicos (alto, medio-alto, medio, medio-bajo, bajo) y el sexo del hablante (hombre, mujer).¹⁰

sociolingüísticos del habla urbana de Cumaná (coordinado por Alida Velásquez, de la Universidad de Oriente), Margarita (cf. Murguey 1982), Puerto Cabello y Valencia (cf. Navarro 1995), Porlamar (cf. García 2003), Maracay (cf. Villalobos 2005), los cuales no pueden ser utilizados en la presente propuesta porque algunos siguen criterios de estratificación diferentes al proyecto GREHV (cf. §1.2), o no están completamente transcritos, o fueron recogidos al inicio del siglo XXI.

⁸ Para detalles específicos de los corpus, cf. Bentivoglio (1998: 40).

⁹ La propuesta tiene lugar en Caracas, durante las Jornadas de la Asociación de Lingüística y Filología de América Latina-Asociación Venezolana para el Avance de la Ciencia - ALFAL-ASOVAC 1989.

¹⁰ El nivel socioeconómico de los informantes se determinó a través de un índice creado especialmente para el análisis de fenómenos sociolingüísticos (Contasti 1980), utilizado anteriormente en el diseño del *Corpus Sociolingüístico de Caracas* 1977 (cf. Bentivoglio, D’Introno & Sosa 1977). Fue esencialmente el mismo método con una pequeña variante: “la escala de nueve valores (1, 2, 3, 4, 5, 6, 7, 8, 9) fue reducida a cinco

En la tabla 1 se observa la matriz de datos de la estratificación:

| GRUPO ETARIO | A: 14-29 años | | B: 30-45 años | | C: 46-60 años | | D: 61 o + años | |
|---------------------------|---------------|-------|---------------|-------|---------------|-------|----------------|-------|
| NIVEL SOCIOECONÓMICO/SEXO | hombre | mujer | hombre | mujer | hombre | mujer | hombre | mujer |
| alto (1) | | | | | | | | |
| medio alto (2) | | | | | | | | |
| medio (3) | | | | | | | | |
| medio bajo (4) | | | | | | | | |
| bajo (5) | | | | | | | | |

Tabla 1. Matriz de distribución de entrevistas para los corpus del GREHV.

Para que la muestra fuera representativa y proporcional al número de habitantes de cada ciudad, se acordó que la cantidad de entrevistas debía ser de cuatro (4) hablantes por casilla para el corpus de Caracas y dos (2) hablantes para el habla del resto de las urbes.

La mayoría de los temas tratados en las conversaciones tienen que ver con la infancia de los hablantes (educación, juegos, disciplina familiar), las tradiciones y costumbres regionales (festividades, celebraciones religiosas), la historia política de la ciudad y, ocasionalmente, la ocupación de la persona entrevistada, ámbito profesional, viajes, entre otros.

En su momento, las grabaciones fueron transliteradas de forma mecanográfica. La transcripción se hizo de forma ortográfica y se utilizaron todos los signos de puntuación convencionales que se consideraron necesarios para recuperar la organización discursiva del texto, el sentido y la intencionalidad.¹¹

De esta forma, la construcción de una *Gramática del Español hablado en Venezuela* constituye el más ambicioso proyecto de investigación propuesto en la historia lingüística venezolana.¹² Si bien el desarrollo de la propuesta ha sido diferente por parte de cada una

(1, 3, 5, 7, 9) con el fin de ajustarla al modelo de cinco valores utilizado por el *Proyecto Venezuela* (Bentivoglio & Sedano, 1993: 8).

¹¹ Para una explicación más amplia de los métodos de recolección y transcripción de los corpus, cf. Bentivoglio & Sedano (1993) y Domínguez & Mora (1998).

¹² Es importante señalar que, desde finales de los años 60, la mayoría de los investigadores venezolanos se ha preocupado por fundamentar científicamente la investigación del lenguaje al estudiar y documentar la extensión y pertinencia real de estos fenómenos en amplias muestras de la lengua en uso. Un ejemplo de

de las instituciones responsables y la GREHV seguramente no llegue a publicarse, el proyecto ha contribuido a formar nuevos investigadores y ha originado estudios de gran valía.¹³

1.3 La noción de Reingeniería de Corpus y el estudio dialectal como propósito

En trabajos anteriores se ha propuesto el término *Reingeniería de Corpus* para designar las tareas relacionadas con la reconfiguración de materiales de habla (orales y escritos) recopilados y estructurados en diversas bases y cuerpos de datos cuando su productividad se torna vulnerable (cf. Guirado 2014, 2015, 2018). Así, la creación de corpus específicos a partir del rediseño y la reconcepción de uno o varios corpus responde a la necesidad de incrementar y diversificar los ámbitos de análisis de los fenómenos del lenguaje.

En el presente trabajo parto de la definición de *corpus* de Parodi (2008) porque es una de las nociones más completas; en la misma, el autor señala la pertinencia que tiene la recuperación de los datos de origen en la investigación posterior; para el autor, el corpus debe incluir en su configuración “detalles relevantes acerca de su recolección y procedencia. De modo más específico, se espera se almacene en conjunto con otros corpus diversos con el fin que se permita su comparación e, idealmente, su contraste” (Parodi 2008: 107).

Guirado (2014) propone una serie de acciones necesarias para la construcción de un nuevo corpus a partir de la reingeniería de otros: i. definir el/los corpus objeto de la reingeniería. ii. establecer la nueva finalidad; iii. comprobar la representatividad; iv. registrar la reconfiguración; v. recodificar las muestras; vi. diagramar la distribución; vii. digitalizar la muestra en diversos formatos; viii. informar sobre su disponibilidad.

ello lo constituye el corpus recogido por los investigadores del Instituto Pedagógico de Caracas, el cual ha servido para la realización de varios estudios de tipo dialectal (Cf. Barrera Linares 1978). Para una explicación detallada acerca de las diferencias entre el GREHV y otros proyectos similares, cf. Bentivoglio 1996.

¹³ En el proyecto participaron: Universidad Central de Venezuela, UCV (Caracas); Universidad del Zulia, LUZ (Maracaibo), Universidad de Los Andes (Mérida), ULA; y Universidad de Oriente (Cumaná), UDO. Los equipos contaron con el apoyo financiero del Consejo de Desarrollo Científico y Humanístico de cada una de las universidades para la recolección y transcripción de los corpus.

Asimismo, Guirado (2015) advierte que se deben valorar los siguientes aspectos antes de escoger un rediseño:

- i. *Atributos quebrantados*: depreciación de la sincronía, desequilibrio distribucional, formato, etc.
- ii. *Propiedades sólidas*: estratificación social de la muestra, comparabilidad, procesamiento informatizado, datos de procedencia, etc.
- iii. *Procesos factibles*: cambio y/o creación de nuevas de unidades de análisis, reestratificación, sinergia con otras muestras, disponibilidad, etc.

Resulta obvio, entonces, que la posibilidad de recuperar los metadatos de la arquitectura de uno o varios corpus constituye una condición fundamental para poder llevar a cabo cualquier proyecto de reingeniería.

De acuerdo con el criterio de representatividad, los corpus pueden ser objeto de diversas tipologías. Atkins, Clear & Ostler (1992) diferencian entre *corpus* y *subcorpus*. Para Sinclair (1996) y Torruella & Llisterri (1999), un subcorpus es cualquier porción seleccionada de un corpus mayor. No obstante, para Pearson (1998), los corpus especializados no constituyen un subcorpus porque “they are not designed to have all of the properties of a larger corpus” (Pearson 1998: 48). El autor prefiere identificar con el término *corpus de propósito especial (especial purpose corpus)* toda selección de materiales que responde a un objetivo específico de investigación: “whose composition is determined by the precise purpose for which it is to be used [...] is the reason for creating or selecting the corpus” (Pearson 1998: 48).

El objetivo de la presente investigación fue crear un subcorpus para el estudio dialectal del español hablado en Venezuela en una microdiacronía del siglo XX,¹⁴ a partir de la reingeniería de tres corpus de esta variedad, recogidos y transcritos durante la década 1986-1996, en tres ciudades del país (Maracaibo, Caracas y Mérida).¹⁵

¹⁴ Cf. Pons Bordería (2014) para la noción de *microdiacronía* y su importancia en el estudio de algunos fenómenos durante el siglo XX.

¹⁵ En el objetivo explícito del presente trabajo subyace el propósito de honrar la tenacidad de Paola Bentivoglio para llevar a cabo la ardua empresa de crear corpus, su capacidad para convocar e integrar

He dividido este artículo en cinco partes. En la sección que acabo de exponer, introduje brevemente una perspectiva dialectal del habla en Venezuela, la memoria de un proyecto destinado a su estudio, la noción de *Reingeniería de Corpus* y el objetivo de la presente investigación. En el apartado siguiente describiré la configuración de los tres corpus que serán objeto de reingeniería. Luego relataré el proceso de reestructuración del corpus marabino. Posteriormente, presentaré la estimación de representatividad y el rediseño de las muestras para la creación del nuevo corpus. Finalmente, cerraré el artículo con algunas consideraciones sobre el proceso de reingeniería documentado.

2. Los corpus objeto de reingeniería

2.1 *Corpus sociolingüístico de Maracaibo 1986 (CSMB-86)*

El proyecto del CSMB-86 estuvo coordinado por las profesoras Bertha Chela-Flores de la Universidad Simón Bolívar (USB) y Jeannette Gelman de la Universidad del Zulia (LUZ). Los entrevistados son todos oriundos de la ciudad de Maracaibo e hijos de padres marabinos. Aunque el diseño del corpus sigue el modelo establecido en las bases metodológicas del Proyecto GREHV, se usaron parámetros diferentes en algunas entrevistas, lo que influyó en la configuración final del corpus.

Según el informe del proyecto entregado al CDCH-LUZ (cf. Chela-Flores & Gelman 1988), se grabaron ochenta y dos (82) entrevistas a un total de ochenta y siete (87) hablantes. Esta divergencia responde al hecho de que cinco (5) de las grabaciones corresponden a entrevistas semidirigidas entre un (1) encuestador y dos (2) entrevistados. Adicionalmente, en algunas casillas se grabaron más de dos hablantes, lo que impide la distribución equitativa en las diferentes categorías, tal como se visualiza en la tabla 2:

diversos equipos en torno a un proyecto de carácter nacional como el GREHV y su convicción en los estudios coordinados del habla venezolana.

| GE | <u>14-29 años</u> | | <u>30-45 años</u> | | <u>46-60 años</u> | | <u>61 o + años</u> | | Total |
|------------|-------------------|----------|-------------------|----------|-------------------|----------|--------------------|----------|-------|
| NS/s | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | |
| alto | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| medio alto | 4 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 19 |
| medio | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| medio bajo | 4 | 4 | 2 | 2 | 2 | 2 | 2 | 2 | 20 |
| bajo | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| Total | 14 | 12 | 11 | 10 | 10 | 10 | 10 | 10 | 87 |
| | 26 (-1) | | 21 | | 20 (-3) | | 20 | | |

Tabla 2. Distribución original de los hablantes del CSMB-86.

Debido a los compromisos con el proyecto GREHV, las profesoras Chela-Flores y Gelman entregaron una copia del informe presentado al CDCH-LUZ al Instituto de Filología “Andrés Bello” de la Universidad Central de Venezuela (IFAB-UCV), en el cual se incluía gran parte de las copias mimeografiadas de las entrevistas. En la relación del material recibido se reporta la ausencia de cuatro (4) transcripciones que aparecen como realizadas en el informe del proyecto. En el año 2000, los profesores Carlos Sánchez y Rosa Sánchez Doreste, junto a un equipo de estudiantes de la LUZ, emprendieron la tarea de digitalizar la totalidad del corpus. Los investigadores trasladaron la mayor parte del corpus pero no lograron ubicar diecisiete (17) transcripciones, incluidas las ya citadas. No obstante, como se describe en §3, gracias al trabajo conjunto entre los equipos de LUZ-UCV, fue posible recuperar buena parte de estos materiales.

En la actualidad, las transcripciones están disponibles en formato impreso y digital en el Departamento de Lengua Modernas de la Facultad de Humanidades y Educación de la LUZ y en el Departamento de Dialectología del IFAB-UCV.¹⁶

2.2 Corpus sociolingüístico de Caracas 1987 (CSCA-87)

Las creadoras y coordinadoras del CSCA-87 fueron Paola Bentivoglio y Mercedes Sedano de la UCV; posteriormente se unió al equipo coordinador Alexandra Álvarez de la

¹⁶ En la introducción del informe de investigación, Chela-Flores & Gelman (1988) informan que el material de audio está a la disposición en las oficinas del CDCH-LUZ; sin embargo, actualmente se desconoce el destino de las grabaciones.

Universidad de Los Andes (ULA), (cf. Bentivoglio & Sedano 1993). El corpus reúne 160 entrevistas hechas a hablantes caraqueños, hijos de padres caraqueños, residenciados en la ciudad. La distribución de las grabaciones se ilustra en la tabla 3:

| GE NS/S | 14-29 años | | 30-45 años | | 46-60 años | | 61 o + años | | Total |
|-------------------|------------|----------|------------|----------|------------|----------|-------------|----------|------------|
| | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | |
| alto | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 32 |
| medio alto | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 32 |
| medio | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 32 |
| medio bajo | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 32 |
| bajo | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 32 |
| Total | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 160 |
| | 40 | | 40 | | 40 | | 40 | | |

Tabla 3. Distribución de los hablantes del CSCA-87

Las transcripciones de todas las grabaciones están incorporadas al CREA y disponibles en audio y texto digital en el Departamento de Dialectología del IFAB-UCV.

2.3 Corpus sociolingüístico de Mérida 1995 (CSMD-95)

El proyecto para la recolección del CSMD-95 estuvo a cargo de las profesoras Elsa Mora hasta el año 1992 y Carmen Luisa Domínguez, ambas investigadoras de la ULA (Cf. Domínguez & Mora 1998). El corpus consta de ochenta (80) entrevistas realizadas a hablantes merideños, en su mayoría nativos de esa ciudad o procedentes del estado Mérida pero que han vivido gran parte de su vida en la capital del estado. En la tabla 4, se muestra la distribución:

| GE NS/S | 14-29 años | | 30-45 años | | 46-60 años | | 61 o + años | | Total |
|-------------------|------------|----------|------------|----------|------------|----------|-------------|----------|-----------|
| | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | |
| alto | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| medio alto | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| medio | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| medio bajo | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| bajo | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| Total | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 80 |
| | 20 | | 20 | | 20 | | 20 | | |

Tabla 4. Distribución de los hablantes del CSMD-95.

Todas las entrevistas están incorporadas al CREA y los archivos en formato digital (audio y texto) deben solicitarse al Grupo de Lingüística Hispánica de la ULA, GLH-ULA.

3. Reingeniería del corpus de habla marabina

En principio, los corpus sociolingüísticos de Maracaibo, Caracas y Mérida comparten una arquitectura similar que, en teoría, permitiría el estudio dialectal de fenómenos gramaticales del español hablado en Venezuela. No obstante, las inconsistencias estructurales que presenta el CSMB-86 –señaladas en el apartado §2.1, (casillas con una sola entrevista y grabaciones con dos entrevistados)– limitan algunos análisis comparativos.¹⁷ En tal sentido, a continuación se presenta una propuesta de reingeniería con el propósito de potenciar un producto lingüístico subutilizado. Primero se relatará una memoria de las actividades de pesquisa de los materiales faltantes y luego se explicará la propuesta de reestructuración.

3.1 La recuperación de los materiales

En el año 2006, con motivo de un estudio sobre el uso variable de la preposición *para~pa'* (cf. Guirado 2007), el equipo del IFAB solicitó el permiso de la profesora Rosa Sánchez Doreste para intentar completar y reorganizar el CSMB-86 con los materiales que estaban resguardados en el instituto. Gracias a ello, se recuperaron e incluyeron trece (13) de las diecisiete (17) entrevistas que faltaban.¹⁸ En la tabla 5, se presenta la

¹⁷ Por ejemplo, en Guirado (2007) solo se usaron los hablantes de los grupos etarios B y D, ya que era la única distribución que permitía hacer una comparación equitativa de los tres corpus.

¹⁸ Las entrevistas 05, 11, 13, 25, 32, 34, 35, 37, 49, 52, 73, 76, 83 se digitalizaron en el IFAB (cf. tabla 5). Agradezco especialmente al profesor José Alejandro Martínez Lara por su colaboración. Sin su empeño y meticulosidad no hubiera sido posible recuperar la mayor cantidad de los materiales.

distribución de las entrevistas y se identifican los hablantes grabados con dígitos del 01 al 87:¹⁹

| GE | <u>14-29 años</u> | | <u>30-45 años</u> | | <u>46-60 años</u> | | <u>61 o + años</u> | |
|--------------|-------------------|----------|-------------------|----------|-------------------|----------|--------------------|----------|
| NS/s | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> |
| alto | mb-31 | mb-56 | mb-41 | mb-32 | mb-38 | mb-55 | mb-47 | mb-75 |
| | mb-60 | mb-58 | mb-61 | mb-49 | mb-39 | [mb-86] | mb-85 | mb-84 |
| medio | mb-26 | mb-74 | mb-40 | mb-34 | mb-72 | mb-78 | mb-25 | mb-65 |
| | mb-30 | mb-83 | mb-44 | mb-37 | [mb-80] | mb-79 | mb-69 | mb-71 |
| alto | mb-36 | | mb-54 | | | | | |
| | mb-57 | | | | | | | |
| medio | mb-23/24 | mb-23/24 | mb-03 | mb-05 | mb-45 | mb-04 | mb-48 | mb-53 |
| | mb-29 | [mb-81] | mb-73 | mb-15 | mb-68 | mb-70 | mb-59 | mb-87 |
| medio | mb-10 | mb-06 | mb-09 | mb-07 | mb-02 | mb-01 | mb-17/18 | mb-43 |
| | mb-19/20 | mb-08 | mb-13 | mb-12 | mb-33 | mb-11 | mb-35 | mb-51 |
| bajo | mb-19/20 | mb-42 | | | | | | |
| | mb-21-22 | mb-21/22 | | | | | | |
| bajo | mb-27/28 | mb-27/28 | mb-14 | mb-16 | mb-77 | mb-63 | mb-46 | mb-17/18 |
| | mb-62 | mb-66 | mb-76 | mb-64 | [mb-82] | mb-67 | mb-50 | mb-52 |

Tabla 5. Inventario de las transcripciones del CSMb'86 (cf. Chela-Flores & Gelman 1988).

Luego de este inventario, se duplicaron las copias digitales de las transcripciones con dos hablantes. En cada duplicado se eliminó la intervención de uno de los entrevistados con el propósito de ubicar al otro hablante en la casilla socioeconómica que le corresponde y evitar así el solapamiento de información tanto lingüística como extralingüística. De esta forma, gracias a la labor conjunta entre los equipos de la LUZ y UCV fue posible culminar el trabajo de transliteración electrónica emprendido por el profesor Carlos Sánchez, quien previó la necesidad de ofrecer el material del corpus en un soporte apropiado para el análisis informatizado de los datos.

¹⁹ La numeración corresponde a la presentada en el informe del proyecto. En el caso de las entrevistas con dos hablantes (17/18; 19/20; 21/22; 23/24; 27/28), se destacó en negritas el número que corresponde al hablante asignado a esa casilla. Las marcadas entre corchetes ([]) indican las grabaciones que aparecen como realizadas pero de las cuales no se ha logrado ubicar ni la transcripción ni el archivo de audio (80, 81, 82 y 86).

3.2 La reclasificación de las entrevistas

La primera decisión que se tomó para emprender la reingeniería de los materiales fue unificar la cuota de afijación del corpus a dos (2) hablantes por casilla, de acuerdo con la metodología del proyecto GREHV. Por tal motivo, se retiraron las transcripciones 26, 30 y 54 del nivel medio alto, y las 06, 20, 21 y 22 del medio bajo.²⁰ Asimismo, se decidió declarar inexistentes las entrevistas carentes de respaldo físico del audio de grabación (80, 81, 82 y 86) y eliminarlas de la matriz. A continuación, se grafican las acciones de reestructuración:

| GE | <u>14-29 años</u> | | <u>30-45 años</u> | | <u>46-60 años</u> | | <u>61 o + años</u> | | cambios |
|-------------------|-------------------|----------|-------------------|----------|-------------------|----------|--------------------|----------|---------------------------|
| NS/S | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | |
| alto | mb-31 | mb-56 | mb-41 | mb-32 | mb-38 | mb-55 | mb-47 | mb-75 | mb-86 |
| | mb-60 | mb-58 | mb-61 | mb-49 | mb-39 | | mb-85 | mb-84 | |
| medio alto | mb-36 | mb-74 | mb-40 | mb-34 | mb-72 | mb-78 | mb-25 | mb-65 | → mb-26/mb-30/mb-54 |
| | mb-57 | mb-83 | mb-44 | mb-37 | | mb-79 | mb-69 | mb-71 | mb-80 |
| medio | mb-23 | mb-24 | mb-03 | mb-05 | mb-45 | mb-04 | mb-48 | mb-53 | mb-81 |
| | mb-29 | | mb-73 | mb-15 | mb-68 | mb-70 | mb-59 | mb-87 | |
| medio bajo | mb-10 | mb-08 | mb-09 | mb-07 | mb-02 | mb-01 | mb-18 | mb-43 | → mb-06/mb-20/mb-21/mb-22 |
| | mb-19 | mb-42 | mb-13 | mb-12 | mb-33 | mb-11 | mb-35 | mb-51 | |
| bajo | mb-27 | mb-28 | mb-14 | mb-16 | mb-77 | mb-63 | mb-46 | mb-17 | mb-82 |
| | mb-62 | mb-66 | mb-76 | mb-64 | | mb-67 | mb-50 | mb-52 | |

Tabla 6. Reestructuración de las transcripciones del CSMB'86

Ahora bien, en virtud de que siguen faltando cuatro entrevistas que impiden comparar a plenitud los corpus de las tres ciudades porque están ubicadas en niveles vertebrales de la categorización socioeconómica (alto, medio, bajo), se decidió indagar en

²⁰ Las entrevistas mb-26 y mb-54 tienen menor longitud respecto a sus vecinas en la casilla. La mb-30 se excluye porque la entrevista fue clasificada por error en otro grupo etario (el hablante es hermano gemelo de mb-31; entre ambas, mb-30 tiene menor extensión). Se eliminan mb-21 y mb-22 porque pertenecen a una misma entrevista y sobran en la casilla. Entre mb-19 y mb-20 hubo que escoger la de mayor longitud para que ocupara la casilla (mb-19). Finalmente, la valoración de los datos socioeconómicos de la hablante de la entrevista mb-06, tal como se verá inmediatamente, permite reclasificar la misma y sustituir con ella una de las grabaciones faltantes.

la ficha de identificación de los hablantes –almacenada en la base de datos original–,²¹ con la intención de obtener información clave adicional sobre las características sociales de algunos entrevistados, la cual permitiera reclasificarlos en el espacio vacío de las casillas faltantes; en concreto, se reanalizaron los datos socioeconómicos para verificar una virtual reubicación de hablantes en un nivel anterior o posterior inmediato. Las decisiones tomadas fueron:

- i. Incorporar y subir la entrevista excluida mb-06, de una estudiante universitaria, del nivel medio bajo del grupo A a la casilla incompleta del nivel medio [~~mb-81~~].
- ii. Mover la entrevista mb-33, perteneciente a un cobrador sin estudios formales, del nivel medio bajo del grupo C a la casilla vacía del nivel bajo [~~mb-82~~].
- iii. Subir la entrevista mb-79, de una profesional universitaria con estudios en el exterior, del nivel medio alto del grupo C a la casilla faltante del nivel alto [~~mb-86~~].

Finalmente, en la tabla 7, se puede apreciar la distribución ajustada de las entrevistas, la cual se sugiere usar como muestra completa del CSMb'86:

| GE NS/S | 14-29 años | | 30-45 años | | 46-60 años | | 61 o + años | | Total |
|-------------------|------------|---------------|------------|----------|---------------|---------------|-------------|----------|-----------|
| | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | |
| alto | mb-31 | mb-56 | mb-41 | mb-32 | mb-38 | mb-55 | mb-47 | mb-75 | 16 |
| | mb-60 | mb-58 | mb-61 | mb-49 | mb-39 | mb-79↑ | mb-85 | mb-84 | |
| medio alto | mb-36 | mb-74 | mb-40 | mb-34 | mb-72 | mb-78 | mb-25 | mb-65 | 14 |
| | mb-57 | mb-83 | mb-44 | mb-37 | | | mb-69 | mb-71 | |
| medio | mb-23 | mb-24 | mb-03 | mb-05 | mb-45 | mb-04 | mb-48 | mb-53 | 16 |
| | mb-29 | mb-06← | mb-73 | mb-15 | mb-68 | mb-70 | mb-59 | mb-87 | |
| medio bajo | mb-10 | mb-08 | mb-09 | mb-07 | mb-02 | mb-01 | mb-18 | mb-43 | 15 |
| | mb-19 | mb-42 | mb-13 | mb-12 | | mb-11 | mb-35 | mb-51 | |
| bajo | mb-27 | mb-28 | mb-14 | mb-16 | mb-77 | mb-63 | mb-46 | mb-17 | 16 |
| | mb-62 | mb-66 | mb-76 | mb-64 | mb-33↓ | mb-67 | mb-50 | mb-52 | |
| Total | 10 | 10 | 10 | 10 | 8 | 9 | 10 | 10 | 77 |
| | 20 | | 20 | | 17 | | 20 | | |

Tabla 7. Distribución ajustada de las transcripciones del CSMb'86.

²¹ Vale la pena aclarar que dicha base de datos consiste en un registro manual de la información en una planilla mimeografiada diseñada para tales fines. Esto ha permitido caer en cuenta de la necesidad de digitalizar esta información en un futuro, de modo que sea posible su procesamiento informatizado.

De esta forma, la reingeniería del CSMB-86 ha permitido estructurar un corpus más coherente e integral; si bien no ha sido posible completar las 80 muestras de habla, la redistribución de los materiales, tal como se observa en la tabla 7, amplía las probabilidades de explotación del corpus porque ofrece la opción de una comparación equitativa, bien entre sus propios componentes o de sus componentes con los de otros corpus, es decir, se pueden comparar todos los grupos etarios al menos en los niveles socioeconómicos alto-medio-bajo.

4. EL CORDIA-VEN-sXX

Una vez reestructurado el CSMB-86, se presenta la propuesta de elaborar un subcorpus para el estudio sociolingüístico de ciertos fenómenos del español hablado en Venezuela en tres regiones de esta variedad. El mismo debe estar formado por 72 casillas con afijación uniforme, creadas a partir del producto de las variantes de las variables dialecto, grupo etario, nivel socioeconómico y sexo ($3 \times 4 \times 3 \times 2 = 72$).

El siguiente paso consiste en asignar la cuota por casilla. En la tabla 8, se puede observar el cálculo de la proporción de hablantes del nuevo corpus respecto a la totalidad de la población de referencia. Los datos expresados en cada columna corresponden a: i. relación corpus-ciudad-año de grabación; ii. población estimada de la metrópoli según el censo correspondiente;²² iii. número de hablantes por muestra según la distribución por casilla; y, vi. cálculo de la proporción de habitantes oriundos mayores de 14 años por hablante ($f = \text{población} / \text{total de hablantes}$).

²²Los datos aportados por el censo de 1981 se usan de referencia para las muestras procedentes de Maracaibo (CELADE 1991: §cuadros, I.12) y Caracas (Delgado Linero 2005: 204), y los del Censo 1990 para los de Mérida (Ramírez 1991: 145). Los datos incluyen el total de habitantes de las capitales, así como del entorno metropolitano de cada una. Al área metropolitana de Caracas se suman las parroquias foráneas del Distrito Federal, de los municipios del Departamento Sucre del estado Miranda y de las capitales de los municipios altomirandinos; el área de Maracaibo incluye los restantes municipios que integran el Distrito Maracaibo y las ciudades y centros poblados de La Concepción, Santa Cruz, El Carmelo, Chiquinquirá, Potrerito, Las Cruces y Palmarejo; y el área metropolitana de Mérida contempla todos los municipios de los cuales Mérida es capital y la ciudad de Ejido.

| CORPUS (ciudad-año) | HABITANTES POR METRÓPOLI | Nº DE HABLANTES (72/3 = 24 casillas x ciudad) | PROPORCIÓN DE HABITANTES POR ENTREVISTADO (f=población/hablantes) |
|------------------------|--------------------------------|--|---|
| CSMb-86 | 989.935 | 2 x casilla = 48 | |
| CSCa-87 | 2.970.405 | 2 x casilla = 48 | |
| CSMd-95 | 237.575 | 2 x casilla = 48 | 1:17.500→2.520.000 |
| Total | 4.197.915 – 20% (2.518.749) | 144 | |

Tabla 8. Estimación de la proporción de la población por hablante en casilla.

Como se ha señalado en otras oportunidades, en el caso de los corpus de propósito especial, la representatividad puede “restringirse a la del dominio de estudio específico para el que son creados” (Pérez Hernández & Moreno Ortiz 2009: 76). En virtud de la restricción de origen establecida para las grabaciones (hablantes nacidos y residentes en la región, de padres también nacidos en la zona) y del límite inferior de edad de los hablantes (14 años), el total de habitantes de las metrópolis se redujo en un 20%, ya que este margen permite sustentar con más fuerza la representatividad del diseño de la muestra;²³ asimismo, se usó una cifra levemente mayor para obtener un número “redondo” (2.518.749→2.520.000/144=17.500). El resultado implica que 144 hablantes sería una proporción representativa de esta población ($f\ 0,0175 < 0,025$).

De esta forma, el *Corpus dialectal para el estudio sociolingüístico del español hablado en Venezuela a finales del siglo XX* (CORDIA-VEN-sXX) está constituido por 144 grabaciones de habla de 3 ciudades venezolanas, distribuidas equitativamente, de acuerdo con la estratificación de los hablantes, en cuatro (4) grupos etarios, tres (3) niveles socioeconómicos y sexo. En la tabla 9 se puede observar esta distribución:

²³ Según los datos aportados por el Instituto Nacional de Estadística, el porcentaje de personas nacidas en Venezuela, cuya entidad de nacimiento es la misma donde residen al momento del empadronamiento oscila entre el 71% y el 76% en los censos realizados en los años 1981 y 1991 respectivamente, es decir, que al menos un 25% de la población censada no es oriunda de la región (cf. INE 2014). Este dato había sido señalado ya en un informe realizado por el Centro Latinoamericano y Caribeño de Demografía para el Banco Interamericano de Desarrollo: “Respecto a la migración interna, la información acerca de los cambios de residencia entre Entidades de nacimiento y de residencia actual muestra que algo más de la cuarta parte de la población nativa del país residía en 1981 en una Entidad distinta a la de su nacimiento” (CELADE 1991: 10). Por otra parte, entre 1985 y 1995, una media del 13,85% de la población oscila entre 0 y 4 años y un 34,6% entre 15 y 19 años (CELADE 1991). De forma similar, el INE informa que en 1990 el 37,3% de la población oscila entre los 0 y 14 años (INE 2014). Esta misma cifra suele ser mencionada por los analistas, cuando se refieren a la estructura de la población censada en tres rangos de edad (0 a 14, 15 a 64 y 65 o más años), para respaldar su tesis del *bono demográfico*, todos con fuente en los Censos de Población y Vivienda de la OCEI 1950-1990 e INE 2001-2011.

| GE | | 14-29 años | | 30-45 años | | 46-60 años | | 61 años o + | | Total |
|-------------------|--------|------------|----------|------------|----------|------------|----------|-------------|----------|-------|
| CIUDAD | NS/SEX | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | <i>h</i> | <i>m</i> | |
| MARACAIBO 1986 | alto | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| | medio | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| | bajo | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| CARACAS 1987 | alto | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| | medio | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| | bajo | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| MÉRIDA 1995 | alto | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| | medio | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| | bajo | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 |
| Total | | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 144 |
| | | 36 | | 36 | | 36 | | 36 | | |

Tabla 9. Distribución ajustada de las transcripciones del CORDIA-VEN-sXX.

Finalmente, las grabaciones se identificaron con un código alfanumérico, convenido en el proyecto GREHV, el cual permite recuperar el lugar del hablante en la casilla de estratificación y, por ende, los datos asociados al análisis sociolingüístico del texto de la entrevista:

- i. ciudad: Maracaibo (MB-); Caracas (CA-); Mérida (MD-).
- ii. grupo etario: 15-29 años (A), 30-45 años (B), 46-60 años (C), 61 o más años (D).
- iii. nivel socioeconómico: alto (1), medio (3), bajo (5).
- iv. sexo: hombres (M) y mujeres (F).
- v. lugar en la casilla: A, B.
- vi. año de referencia del corpus: CSMb 1986 (.86); CSCa 1987 (.87); CSMd 1995 (.95)

A continuación, se muestra cómo opera la codificación con tres ejemplos tomados de Guirado (2016) del uso ponderativo del posesivo antepuesto con diminutivo en el habla de las tres ciudades:

- (1) a. me doy **mi** ... *zambulladita* <sic> de vez en cuando (MB-B3FA.86)
- b. friego **mis** *coroticos del café*, hago **mi** *cafecito*, cuando lo tengo (CA-D5FB.87)
- c. Yo pienso seguir tocando, *matando mis tigrillos* y demás (MD-A1MA.95)

El ejemplo (1a) pertenece a una hablante marabina entre 30 y 45 años de edad, de nivel socioeconómico medio; en el (1b) la hablante es una caraqueña con más de 60 años del nivel bajo; y, finalmente, el ejemplo (1c) identifica a un joven merideño entre 14 y 29 años de nivel socioeconómico alto.

5. Consideraciones finales

La construcción de un subcorpus dialectal a partir de los corpus sociolingüísticos del proyecto GREHV requirió de una revisión fundamental de los procesos que confluyeron en la construcción de cada uno, –y, en algunos casos, la intervención en la configuración inicial–, para diagnosticar los elementos potenciables y reubicarlos en una nueva arquitectura. Como resultado, el CORDIA-VEN-sXX constituye una fuente útil de textos reales para el estudio exploratorio del español hablado en Venezuela en una microdiacronía específica. Su representatividad garantiza la proyección de algunas tendencias de uso y, en algunos casos, el análisis exhaustivo de fenómenos del lenguaje asociados a factores socioculturales específicos, ya que su proporción permite hacer generalizaciones e incluso acercamientos desde algunos modelos estadísticos inferenciales. Así, el producto final del proceso de reingeniería confirma que un corpus sociolingüístico debidamente estratificado, transcrito y digitalizado se puede instrumentalizar en otra base de datos con fines y enfoques similares.

Referencias bibliográficas

- ALVARADO, Lisandro ([1955]1929) *Glosario del bajo español en Venezuela*, Caracas: Ministerios de Educación.
- ÁLVAREZ, Alexandra & Anderzon MEDINA (1999) “‘I love Caracas’. Language attitudes in High School Students from the Venezuelan Andes”, en *Memorias de AVEPLEFE*, Mérida: Universidad de Los Andes.

- ÁLVAREZ, Alexandra (1999) "Comunidad de habla e identidad en Venezuela: el Centro y Los Andes", en Matthias Perl & Klaus Pört (eds.), *Identidad cultural y lingüística en Colombia, Venezuela y en el Caribe hispánico*, Tübingen: Niemeyer, 197-212.
- ÁLVAREZ, Alexandra, Enrique OBEDIENTE & Nelson ROJAS (2009) "Subdialectos del español caribeño de Venezuela: prosodia e identidad regional", *Revista Internacional de Lingüística Iberoamericana*, 2(14), 7-20.
- ÁLVAREZ, Alexandra, Hernán MARTÍNEZ & Lino URDANETA (2001) "Actitudes lingüísticas en Mérida y Maracaibo: Otra cara de la identidad", *Boletín Antropológico* II(52), 145-166.
- ÁLVAREZ, Víctor (2002) "Historia local, historia regional", en Arístides Medina Rubio (comp.), *Lecturas de historia regional y local*, Caracas: Biblioteca Nacional., 59-70.
- ARMAS & CÉSPEDES, Juan Ignacio de ([1882]1977) "Orígenes del lenguaje criollo", en Gladys Alonso & Ángel Luis Fernández (eds.), *Antología de lingüística cubana I*, La Habana: Ciencias Sociales, 115-186.
- ATKINS, Sue, JEREMY Clear & NICHOLAS Ostler (1992) "Corpus Design Criteria", *Literary and Linguistic Computing*, 7(1), 1-16.
- BARRERA LINARES, Luis (1978) "Las áreas dialectales de Venezuela", *Letras*, 34-35, 18-31.
- BENTIVOGLIO, Paola & Mercedes SEDANO (1993) "Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana", *Boletín de Lingüística*, 8, 3-35.
- BENTIVOGLIO, Paola (1996) "La gramática del español hablado en Venezuela", *Lengua y habla*, 1(1), 21-39.
- BENTIVOGLIO, Paola (1998) "La variación sociofonológica", *Español Actual*, 69, 29-42.
- BENTIVOGLIO, Paola, Francesco D'INTRONO & Juan M. SOSA (1977) *Proyecto de investigación para un análisis sociolingüístico del habla de Caracas*, Material mimeografiado, Caracas: Universidad Central de Venezuela.
- BRICEÑO DE BERMÚDEZ, Tarcila (2002) "De la ciudad a la región histórica", en Arístides Medina Rubio (comp.), *Lecturas de historia regional y local*, Caracas: Biblioteca Nacional, 113-128.
- BRICEÑO IRAGORRY, Mario (1982) *Tapices de Historia Patria. Ensayo de una Morfología de la Cultura Colonial*, Caracas: Talleres Litográficos de Impresos Urbina.
- CARDOZO Galué, Germán (2013) "Significaciones y alcances del concepto región histórica", *Cambios y Permanencias*, 4, 45-61.
- CASTRO, Thais & Irania MALAVER (2016) "División dialectal del español venezolano: Estudio dialectológico perceptivo", *Dialectología*, 16, 19-43.
- <www.publicacions.ub.edu/revistes/dialectologia16/>

- CELADE-CENTRO LATINOAMERICANO Y CARIBEÑO DE DEMOGRAFÍA (1991) *Venezuela: Caracterización demográfica y su impacto sobre los servicios sociales. Informe BID*, Santiago de Chile: CELADE-CEPAL-ONU.
- CHELA-FLORES, Bertha & Jeannette GELMAN (1988) *El habla de Maracaibo: Materiales para su estudio*, Material mimeografiado, Maracaibo: Universidad del Zulia.
- CODAZZI, Agustín (1840) *Atlas Físico y Político de la República de Venezuela*, Paris: Lithographie de Thierry Frères.
- CONTASTI, Max (1980) "Metodología para la medición del nivel socio-económico para la población venezolana", *Boletín de la AVEPSO*, 3(2), 13-17.
- DELGADO LINERO, Manuel (2005) "Crecimiento de la población y proceso de urbanización en el Distrito Metropolitano de Caracas: efectos ambientales", en Anitza Freitez, María Di Brienza, Genny Zúñiga, Rhayza Carvallo, Mauricio Phélan & Thaís García (ed.), *Cambio demográfico y desigualdad social en Venezuela al inicio del tercer milenio. II Encuentro Nacional de Demógrafos y Estudiosos de la Población*, Caracas: AVEPO, 197-212.
- DOMÍNGUEZ, Carmen Luisa & Elsa MORA (1998) *El habla de Mérida*, Mérida: Universidad de Los Andes.
- GARCÍA GONZÁLEZ, Carmynel (2003) *El corpus del español hablado en la ciudad de Porlamar*, Tesis de grado, Mérida: Universidad de Los Andes.
- GUIRADO, Krístel (2007) "La alternancia *para~pa'* en tres comunidades de habla de Venezuela", *Interlingüística*, 17, 455-464.
- GUIRADO, Krístel (2014) "Corpus Diacrónico del Habla de Caracas 1987/2013", *Boletín de Lingüística*, 41/42, 19-44. <http://saber.ucv.ve/ojs/index.php/rev_bl/article/view/9578>
- GUIRADO, Krístel (2015) "*Reingeniería de Corpus* en Venezuela: una propuesta metodológica para diversificar el análisis de los corpus del español hablado en Caracas", *Letras*, 92, 55-87. <<http://revistas.upel.edu.ve/index.php/letras/article/view/5962/3239>>
- GUIRADO, Krístel (2016) "*Los muchachos juegan que sus metras, que su volantín, que sus trompos: los posesivos y el realce expresivo del enunciado en tres dialectos venezolanos*", en XXXII *Encuentro Nacional de Docentes e Investigadores de la Lingüística-ENDIL* (ponencia inédita), Universidad de Carabobo, Venezuela.
- GUIRADO, Krístel (2018) "Habla culta de Caracas 1973-2011: Un subcorpus de propósito especial para el estudio diacrónico del habla caraqueña", *CHIMERA. Romance Corpora and Linguistic Studies*, 5(1), 111-116. <<https://revistas.uam.es/index.php/chimera/article/view/9884/10304>>.

- HENRÍQUEZ UREÑA, Pedro (1921) "Observaciones sobre el español de América", *Revista de Filología Española*, 8, 357-390.
- HYMES, Dell (1974) *Foundations in sociolinguistics. An ethnographic approach*, Philadelphia: University of Pennsylvania Press.
- INE-INSTITUTO NACIONAL DE ESTADÍSTICA (2014) *XIV Censo Nacional de Población y Vivienda. Resultados Total Nacional de la República Bolivariana de Venezuela*, Caracas: INE.
<<http://www.ine.gov.ve>>
- MARTÍNEZ, M^a Eugenia & Antonietta ALARIO (2004) *Léxico básico del español en Venezuela*, Trabajo de Ascenso a Agregado, Caracas: Universidad Central de Venezuela.
- MEDINA RUBIO, Arístides (comp.) (2002) *Lecturas de historia regional y local*, Caracas: Biblioteca Nacional.
- MILROY, Lesley (1980) *Language and social networks*, Baltimore: University Park Press.
- MORA, Elsa (1997) "División prosódica dialectal de Venezuela", *OMNIA*, 2, 93-99.
- MORENO FERNÁNDEZ, Francisco (2005) *Principios de sociolingüística y sociología del lenguaje*, Barcelona: Ariel.
- MURGUEY, Antonio (1982) *El habla popular de Margarita*, Nueva Esparta: Universidad de Oriente.
- NAVARRO, Manuel (1995) *El español hablado en Puerto Cabello*, Valencia: Universidad de Carabobo.
- OBEDIENTE, Enrique. (1998) "Fonetismo segmental", *Español Actual*, 69, 11-18.
- OBEDIENTE, Enrique (1999) "Identidad y dialecto: el caso de los Andes venezolanos", en Matthias Perl & Klaus Pört (eds.), *Identidad cultural y lingüística en Colombia, Venezuela y en el Caribe hispánico*, Tübingen: Niemeyer, 213-219.
- OBREGÓN MUÑOZ, Hugo (1981) *Hacia el estudio de la entonación dialectal del español de Venezuela*, Maracay: IUPEMAR.
- PÁEZ URDANETA, Iraset (1981) *Historia y geografía hispanoamericana del voseo*, Caracas: La Casa de Bello.
- PARODI, Giovanni (2008) "Lingüística de corpus: una introducción al ámbito", *Revista de lingüística teórica y aplicada*, 46(1), 93-119.
- PEARSON, Jennifer (1998) *Terms in Context. Studies in Corpus Linguistics* 1, Amsterdam/Philadelphia: John Benjamins.
- PÉREZ, Francisco Javier (1998) "Léxico", *Español Actual*, 69, 83-92.
- PÉREZ HERNÁNDEZ, Chantal & Antonio MORENO ORTIZ (2009) "Lingüística Computacional y Lingüística de Corpus. Potencialidades para la investigación textual", en Nuria Rodríguez Ortega (dir.),

Teoría y literatura artística en la sociedad digital: construcción y aplicabilidad de colecciones textuales informatizadas, Gijón: TREA, 67-96.

PERL, Matthias & Klaus PÖRT (eds.) (1999) *Identidad cultural y lingüística en Colombia, Venezuela y en el Caribe hispánico*, Tübingen: Niemeyer.

PONS BORDERÍA, Salvador (2014) “El siglo XX como diacronía: intuición y comprobación en el caso de o sea”, *RILCE*, 30(3), 985-1016.

QUIRK, Randolph, Sidney GREENBAUM, Geoffrey LEECH & Jan SVARTVIK (1985) *A comprehensive language of the English language*, London: Longman.

RAMÍREZ, Erasmo (1991) “La población de Venezuela según los primeros resultados del Censo '90”, *Boletín de la Academia Nacional de Ciencias Económicas*, 7(14), 127-153.

RENZI, Lorenzo, Giampaolo SALVI & Anna CARDINALETTI (1988) *Grande grammatica italiana di consultazione*, Bologna: Il Mulino.

RONA, José Pedro (1964) “El problema de la división del español en zonas dialectales”, *Presente y futuro de la lengua española*, PFLE I, 215-226.

ROSENBLAT, Ángel ([1956]1989) *Estudios sobre el habla de Venezuela. Buenas y malas palabras*, Caracas: Monte Ávila.

ROSENBLAT, Ángel (1965) *El castellano de España y el castellano de América: unidad y diferenciación*, Caracas: IFAB.

SEDANO, Mercedes (2004) “Unidad y diferenciación del idioma español”, *Textura* 3, 29-47.

SERRÓN, Sergio (1972) “Para determinar las zonas dialectales de Venezuela”, *Revista Imagen*, 56, 15.

SILVA-CORVALÁN, Carmen (1989) *Sociolingüística. Teoría y Análisis*, Madrid: Alhambra.

SINCLAIR, John (1996) “Preliminary recommendations on Corpus Typology. EAG-TCWG-CTYP/P”, en *EAGLES*. (<http://www.ilc.cnr.it/EAGLES96/corpus/typ/corpus/typ.html>)

TEJERA, M^a Josefina (1991) “El castellano por las tierras de Venezuela”, *Presencia y destino. El español de América hacia el siglo XXI. Encuentro Internacional sobre el Español de América*, I, Santafé de Bogotá: Instituto Caro y Cuervo, 209-227.

TEJERA, M^a Josefina (1992) “El léxico como elemento diferenciador”, *Cuadernos Lagoven. El idioma español de la Venezuela actual*, Caracas: Lagoven, 72-91.

TORRUELLA, Joan & Joaquim LLISTERRI (1999) “Diseño de corpus textuales y orales”, en José Manuel Blecua, Gloria Clavería, Carlos Sánchez & Joan Torruella (eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos*, Barcelona: Milenio, 45-77.
<http://liceu.uab.es/~joaquim/publicacions/Torruella_Llisterri_99.pdf>

VILLALOBOS, Ana Margarita (2005) *Corpus Sociolingüístico de la Ciudad de Maracay*, Proyecto de investigación, Maracay: Universidad Pedagógica Experimental Libertador.

ZAMORA, Juan & Jorge GUITART (1982) *Dialectología hispanoamericana*, Salamanca: Almar.