

CAPÍTULO V. MARCO METODOLÓGICO

5.1. PLANTEAMIENTO METODOLÓGICO

5.2. CORPUS

5.2.1. Corpus histórico

5.2.2. Corpus sincrónicos

5.2.2.1. Peninsulares

- a) Alcalá de Henares
- b) Valencia
- c) Málaga
- d) Granada
- e) Sevilla
- f) Canarias

5.2.2.2. Americanos

- A) Ciudad de Guatemala y Cobán
- b) Ciudad de México
- c) Caracas
- d) Montevideo
- e) San Juan de Puerto Rico y el *Macro corpus de la Norma culta*

5.3. ANÁLISIS LÉXICO-SEMÁNTICO

5.4. VARIABLES, PREGUNTAS, MÉTODOS

5.4.1. Estudio dialectológico y sociolingüístico

5.4.2. *Ser y estar* con adjetivos de edad ¿un fenómeno de variacionismo?

5.4.3. Preguntas de investigación

5.4.4. Análisis: factores lingüísticos y sociales

5.4.5. Análisis estadístico

5.5. RESUMEN DE LA METODOLOGÍA

CAPÍTULO V. MARCO METODOLÓGICO

A continuación exponemos los lineamientos metodológicos generales que han guiado la investigación. La exposición contempla la descripción de los corpus y los criterios metodológicos con los que se constituyeron estas muestras de habla; asimismo, se presentan las preguntas de investigación y las variables del análisis.

5.1. PLANTEAMIENTOS METODOLÓGICOS GENERALES

El desarrollo de la presente investigación se basa en el estudio de dos corpus. Esto supone un acercamiento al estudio y sistematización del comportamiento lingüístico a través de un conjunto de datos, reales, organizados de acuerdo con características socioculturales y demográficas. Este estudio se inscribe dentro de la lingüística del corpus, que como plantea R. Caravedo (1999), consiste en:

(...) toda orientación que, en la formulación y en el desarrollo de su programa de investigación (comprendidos la teoría y el sistema de corroboraciones o refutaciones desprendido de la actividad analítica), depende de la observación de un conjunto de datos extraídos de la producción real de los individuos, y ordenados según criterios metodológicos diferentes pero explícitos de investigación. (Caravedo 1999: 19)

La lingüística del corpus supone la posibilidad de proyectar o generalizar las observaciones basadas en el análisis de los datos, nos referimos al “carácter predictivo de la observación” (Caravedo 1999:61). Así, las conclusiones sobre los corpus son, también, aplicables a las comunidades de habla en su conjunto, siempre que se mantengan los mismos factores condicionantes. Desde este enfoque, se estudia la extensión de *estar* como verbo copulativo, con datos históricos, producidos en un marco espacio-temporal dinámico y complejo, en el que las diferencias dialectales representarán estrategias particulares de atribución de la edad. El análisis de cada corpus sincrónico parte del presupuesto de la acción de algunos factores sociales en el proceso

de cambio lingüístico (cf. Hernández Campoy y Almeida, 2005; López Morales, 2004; Moreno Fernández 2005a; Silva-Corvalán, 2001).

5.2. CORPUS

5.2.1. CORPUS HISTÓRICO

El primer aspecto que se debe aclarar se refiere a *la unidad de análisis*. Siguiendo a Sánchez-Prieto Borja y Flores Ramírez (2005), denominaremos *documento histórico* a cada texto: documentos públicos u oficiales, jurídico-administrativos, privados, cartas, testimonios. En este corpus textual, la unidad de muestreo es el texto y no el sujeto (cf. López Morales 1994).

Para el análisis histórico se constituyó un corpus formado por documentos escritos que abarcan los siglos XVI, XVII, XVIII y XIX, de América, Canarias y España. Los textos, en cuestión, proceden de las siguientes fuentes:

- 1) *Archivo de la Real Chancillería de Valladolid.*
- 2) *Documentos para la historia lingüística de Hispanoamérica. Siglos XVI a XVIII I, II, III, III, IX.*
- 3) *Documentos Lingüísticos de la Nueva España. Altiplano-Central* (citado DLNE).
- 4) *Cartas privadas de emigrantes a Indias. 1540-1616* (citado Otte).
- 5) *Materiales para la historia de la lengua española en Centroamérica: Algunos documentos dieciochescos (1703-1758).*
- 6) *Textos para la historia del español II.*
- 7) *Textos para la historia del español III.*
- 8) *Textos para la historia del español IV.*
- 9) *Corpus Diacrónico del Español, CORDE*, de la RAE.

En cuanto a los *DHLH*, cabe destacar que la mayoría de los textos americanos y los canarios pertenecen a este conjunto de documentos escritos entre XVI al XVII. En cuanto a sus características discursivas, son textos variados, como señala Rojas Mayer (2000), y de una tipología que abarca actas capitulares, actas judiciales, denuncias, notificaciones, mercedes, testamentos, inventarios, licencias, padrones, cartas, memoriales, recibos, acuerdos. Buena parte de estos documentos son oficiales; no

obstante, muchas cartas expresan sentimientos y circunstancias personales, que muestran las vicisitudes de la vida en América y la complejidad de las relaciones sociales durante la época colonial. Dice Rojas Mayer al respecto:

Por tal razón, a través de la lectura de sus páginas no solamente podemos obtener una visión de la realidad lingüística en una de sus etapas americanas, sino también del comportamiento pragmlingüístico de los distintos elementos de la sociedad, españoles, criollos, indígenas, mestizos, negros, mulatos, etc., a través de las numerosas interacciones que se observan directamente o pueden presuponerse. De manera que en los últimos años se ha avanzado en los estudios de estos documentos, ya que en ellos no solamente se estudian los aspectos morfosintácticos, fonético-fonológicos o lexicales, sino que se han realizado también ricos estudios semánticos y pragmáticos encuadrados en el atrayente contexto que reconstruimos de su lectura (2000:25).

A continuación, se describen los textos seleccionados para cada zona dialectal:

CUADRO 1. Distribución por país/zona y período de los *DHLH*

<i>País/zona</i>	<i>XV-XVI</i>	<i>XVII-XVIII</i>	<i>XVIII</i>	<i>XIX</i>	<i>Total</i>
Canarias	20	19	23		62
Cuba	3	4	4		11
Santo Domingo	14	14	2		30
Costa Rica	6	9	12		27
Venezuela	13	7	9		29
Reino de Nueva Granada	19	12	4		35
Ecuador		2	4		6
Lima	40				40
Santiago de Chile	19	15	13		47
Paraguay		1	4		5
México	7	4	24		35
Estados Unidos de Norteamérica			3		3
Granada			13		13
Venezuela: Antigua Provincia de Mérida	10				10
Bolivia: Alto Perú	39	26	22		87
Bolivia: Santa Cruz de la Sierra		3	4		7
Argentina: Buenos Aires	4	19	64		87
Argentina: Santa Fé			12		12
Argentina: Córdoba			2		2
Argentina: Mendoza	1	4	3		8
Argentina: Jujuy		1	1	1	3

Argentina: Salta		1			1
Argentina: Tucumán	17	32	18		67
Montevideo			12		12
Total	212	173	252	1	639

Para el estudio de México, se seleccionaron, además, los *Documentos Lingüísticos de la Nueva España. Altiplano-Central (DLNE)*, editados por Company Company (1994, 2001).

CUADRO 2. Distribución por período de los *DLNE*

<i>Documento/fecha</i>	<i>XVI</i>	<i>XVII</i>	<i>XVIII</i>	<i>XIX</i>	<i>Total</i>
	78	98	101	31	308

Company Company señala que estos textos pertenecen tanto al período colonial mexicano y, también, al de la transición entre la época colonial y la independentista; provienen del Altiplano Central, actualmente ciudad de México y los estados de Puebla, Querétaro, Tlaxcala, Guanajuato y Morelos. Es importante destacar que se trata de cartas que los parientes entregaban a las autoridades para poder pasar a Indias, cartas autógrafas de misioneros; denuncias y testimonios (juicios por asesinato, brujería, blasfemia; inventarios y testamentos; peticiones e informes). No obstante, hay distintos grados de esta coloquialidad, pues, por ejemplo, las peticiones e informes dirigidos a las autoridades españolas se escribían de modo muy distinto a las denuncias o a las misivas más personales. Company Company señala otro elemento importante de estos documentos y se refiere a que sus autores son mexicanos de nacimiento (con excepción de los primeros cincuenta años).

Los documentos publicados por Otte (de ahora en adelante *OTTE*), constituyen la tercera fuente para el corpus de nuestra investigación. Se trata de 650 misivas familiares enviadas por españoles desde variados puntos de la geografía americana y con distinto número de años residenciados en América. Se diferencian de los *DLNE* en el origen de quienes escriben y porque corresponden, en su conjunto, al período fundacional del español americano, poseen el registro más informal y espontáneo de todos los documentos que analizamos.

CUADRO 3. Distribución por origen OTTE

<i>Origen</i>	<i>Cartas</i>	<i>%</i>
Nueva España	240	37
América Central	77	12
Nuevo Reino de Granada	70	11
Quito	34	5
Perú	193	30
Chile	11	2
Río de La Plata	5	1
Venezuela	4	1
Antillas	14	2
Filipinas	2	0
Total	650	100

Son cartas privadas cuyos destinatarios eran principalmente las esposas, los sobrinos y hermanos. La temática epistolar es, por ello, íntima y en cada una de estas cartas, los pobladores indianos intentaban arreglar problemas de herencia o, también, convencer a sus familiares de que se fueran a América. Las profesiones, según Otte, eran diversas: militares, encomenderos, ganaderos, mineros, mercaderes, artesanos, carniceros. El tipo de discurso y el registro informal, espontáneo, de estas cartas convierten estas cartas en genuinas muestras sociolingüísticas. Adicionalmente, hay que destacar que el 36, 16% de las cartas fue enviado a distintas regiones de Andalucía y el 32, 35% a lugares de Castilla la Nueva.¹

Además de estas tres fuentes, para la constitución del corpus se emplearon 19 textos de los “*Materiales para la historia de la lengua española en Centroamérica: Algunos documentos dieciochescos*”, distribuidos como se indica a continuación:

¹ Las cartas han sido objeto de distintos estudios. Por citar solo algunos, tenemos el de Francisco Ruiz Fernández (1993), *Fonética del español de América. Cartas de emigrados andaluces a Indias*; dedicados al estudio de las formas de tratamientos se cuenta con los trabajos de P. García Mouton (1992), *Tratamientos en las cartas de indias* y P. Bentivoglio (2002-2004), *Formas de tratamiento en cartas de la segunda mitad del siglo XVI: una aproximación pragmática*. El estudio más general es el realizado por Alvar (1996) en *Los otros cronistas de América*.

CUADRO 4. “*Materiales para la historia de la lengua española en Centroamérica*”. (Ramírez Luengo 2006b)

Países	Documentos
<i>Guatemala</i>	6
<i>Nicaragua</i>	5
<i>Honduras</i>	5
<i>El Salvador</i>	3
Total	19

Según Ramírez Luengo, estos textos son documentos notariales, razón por la cual son muy formales: “En todos los casos se trata de textos caracterizados por un alto grado de formalidad que posee, además, una fuerte tradición retórica, todo lo cual ha de redundar indudablemente en el tipo de lengua que se emplea en ellos, de carácter *culto-estándar*” (2006b: 121). Con este material se logró aumentar el número de documentos de la zona centroamericana; adicionalmente, se analizaron otros documentos centroamericanos publicados en “*Contribución a la historia del español de Honduras: edición y estudio de documentos hondureños del siglo XVIII*”, del mismo autor.

En cuanto a los documentos peninsulares (como se aprecia en el cuadro 1), de los *DHLH* se analizaron textos canarios y granadinos y los documentos publicados en *Textos para la historia del español II*, *Textos para la historia del español III* y *Textos para la historia del español IV*.

En el caso del volumen II, se trata de una recopilación de 37 documentos, principalmente administrativos, de los siglos XVI al XVII, depositados en el Archivo Municipal de Guadalajara (AMGU):

Se ha procurado recoger toda la variedad de tipos documentales del AMGÚ, al tiempo que los asuntos más variados, con la intención de que así estuvieran representados distintos registros de lengua. En especial, se reproducen los pocos textos de índole privada, por reflejar un tipo de lengua que, aun con los rasgos inherentes al uso escrito, se aparta de los formulismos oficiales, y ocasionalmente puede hacerse eco de la lengua coloquial (Sánchez-Prieto Borja, 1995: 18).

En el volumen III hay 39 documentos del Archivo Municipal de Alcalá de Henares, desde 1252 hasta 1684; sobre su tipología, nos dice Paredes García: “(...) El grupo más numeroso de textos es el constituido por los documentos jurídico-administrativos, entre los que podemos encontrar privilegios reales o arzobispaes (...),

un testamento y las sucesivas modificaciones del mismo (...) Junto a este abundante grupo, ofrecemos algunas cartas de tipo privado o semi privado” (2005: 14).

El volumen IV contiene documentos del Archivo Regional de la Comunidad de Madrid: “(...) pertenecen al antiguo fondo de la inclusa de Madrid, propiamente llamada «Hospital de los niños expósitos» (doc. 3) y también «hospital de la inclusa» (5) o aun «obra pía de los niños expósitos de Nuestra Señora de la Enclusa» (6)” (Sánchez-Prieto Borja y Flores Ramírez, 2005: 9).

Se trata de 11 documentos administrativos, notas de abandono y notas de entrega de niños: “Muchas de las notas de abandono son de mano de personas con un grado de instrucción muy escaso, que no han logrado un dominio siquiera pasable de la escritura. Esto las convierte en un material precioso para comprender el proceso de alfabetización entre finales del siglo XVI y el primer tercio del s. XVII” (Sánchez-Prieto Borja y Flores Ramírez, 2005: 11). Finalmente, se consultó el *Corpus Diacrónico del Español*, de la RAE, para los siglos XVII, XVIII y XIX, en este último período tanto para España como para América.

Se extrajeron todas las expresiones de edad; se elaboró una base de datos con el Programa informático ACCESS y se empleó el EXCEL y el SPSS para el análisis cuantitativo. El corpus histórico quedó constituido por 1269 textos.

CUADRO 5. Corpus histórico para el estudio de *ser* y *estar* con adjetivos de edad

<i>Textos americanos</i>	891
<i>Textos peninsulares</i>	316
<i>Canarias</i>	62
<i>Total</i>	1269

No debemos finalizar sin enfatizar la naturaleza del trabajo con textos históricos, sus limitaciones y también sus enormes posibilidades, pues constituyen un punto de partida sólidamente atestiguado de los cambios que se dan en la sincronía de la lengua:

Debe correrse el riesgo, pues, de elegir un corpus para el análisis, con las ventajas y los inconvenientes que tal libertad de elección comporta, sabiendo que el corpus escogido sólo convencionalmente se delimita y que en sentido estricto tampoco podrá ser “orgánico”, cuando en no pocos corpus que para el análisis sincrónico se manejan mucho

habría que discutir de semejante condición. Los caracteres de “cerrado” y “estático” que al texto histórico se le atribuyen de alguna manera son susceptibles de matizarse con la apertura y dinamismo mediante la comparación de los pertenecientes a varias sincronías. (Frago Gracia 2002: 115)

5.5.2. CORPUS SINCRÓNICOS

El estudio del español contemporáneo se hace mediante el análisis de diez corpus de habla sincrónicos, cinco peninsulares y cinco americanos, integrados en el macro corpus *PRESEEA: Proyecto para el estudio sociolingüístico del español de España y de América* (Moreno Fernández 1997, 2006),² creado a partir de 1993 y cuyo objetivo es reunir muestras de habla de las distintas comunidades hispánicas. Se trata de muestras orales, conversaciones grabadas, semi-dirigidas, de 45 minutos de grabación, aproximadamente, hechas a miembros de las comunidades, seleccionados con base en un conjunto de lineamientos metodológicos comunes para la recolección de los corpus específicos, y entre los cuales resumimos los siguientes:

- 1) las comunidades sometidas a estudio son centros urbanos (monolingües o bilingües);
- 2) los informantes deben haber nacido en la ciudad;
- 3) la selección de los informantes se lleva a cabo según un muestreo por cuotas con afijación uniforme, definidas de acuerdo con tres variables sociales: edad, sexo/género y nivel de instrucción;
- 4) para las comunidades con poblaciones de entre medio millón y un millón de habitantes, es suficiente reunir 54 informantes; si el centro urbano es mayor, las muestras pueden tener 72 o 108 hablantes;
- 5) la edad se distribuye en tres generaciones: *Generación 1*, de 20 a 34 años; *Generación 2*, de 35 a 54 años; *Generación 3*, de 55 años en adelante;
- 6) el nivel social o socio-educativo quedó formalizado como *grado de instrucción*, definido en tres variantes: *Grado 1*. Analfabetos, sin estudios. Enseñanza Primaria (hasta 10-11 años de edad aprox.), 5 años aproximadamente de escolarización; *Grado 2*. Enseñanza Secundaria (hasta 16-18 años de edad aprox.), 10-12 años

² <http://www.linguas.net/preseea>

aproximadamente de escolarización; *Grado 3*. Enseñanza Superior (universitaria, técnica superior) (hasta 21-22 años de edad aprox.), 15 años aproximadamente de escolarización;

- 7) se emplea la técnica de la entrevista semidirigida, con una duración mínima de 45 minutos.

Se trata, en síntesis, de muestras de habla estratificadas según el sexo/género, la edad y el grado de instrucción y son representativas de la población de la que han sido extraídas. Son, en consecuencia, metodológicamente afines.³

5.2.2.1. Peninsulares

El español de España quedó cubierto por los corpus sociolingüísticos de las ciudades de Alcalá de Henares, Valencia, Granada, Málaga y Sevilla, que representan tres áreas dialectales: la de Castilla la Nueva, Andalucía y la Comunidad Valenciana. A excepción de la muestra de Sevilla, todos los corpus pertenecen a *PRESEEA*, por lo que comparten las características ya expuestas. Asimismo, del *Macrocorpus de la Norma culta* se analizó el habla de Canarias, con lo que sumamos otra modalidad dialectal al estudio sincrónico. Veamos la distribución de cada una de las muestras.⁴

- a) Alcalá de Henares (Moreno Fernández, Francisco, Ana María Cestero Mancera, Isabel Molina Martos y Florentino Paredes García (2002, 2004, 2007).

El corpus de esta ciudad del centro-norte de España está compuesto por 54 entrevistas, cuya distribución se indica en el Cuadro 6.

³ Los corpus regionales que pertenecen a *PRESEEA* son los de Alcalá, Valencia, Málaga, Granada, Ciudad de México, Ciudad de Guatemala y Cobán, Montevideo y Caracas.

⁴ Nos limitaremos a la descripción general de las casillas; para conocer la descripción exhaustiva de los proyectos *PRESEEA* de cada comunidad: datos demográficos, perfil sociolingüístico, antecedentes teóricos, investigaciones en curso, se puede consultar la bibliografía citada.

CUADRO 6: Distribución del corpus sociolingüístico de *PRESEEA-Alcalá de Henares*

	GENERACIÓN 1 (20 A 34 AÑOS)		GENERACIÓN 2 (35 A 54)		GENERACIÓN 3 (54 O MÁS)		TOTAL
	H	M	H	M	H	M	
GRADO DE INSTRUCCIÓN 1	3	3	3	3	3	3	18
GRADO DE INSTRUCCIÓN 2	3	3	3	3	3	3	18
GRADO DE INSTRUCCIÓN 3	3	3	3	3	3	3	18
TOTAL	9	9	9	9	9	9	54

b) Valencia (Gómez Molina 2001, 2005, 2007)

La muestra valenciana está compuesta por 72 entrevistas; como señala Gómez Molina, el corpus fue recogido entre diciembre de 1996 y febrero de 2006, con un total de 2.497 minutos de grabación (41 horas y 73 minutos) y 432.115 palabras; las horas de grabación se distribuyen en: sociolecto alto, 16 horas y 10 minutos; sociolecto medio, 15 horas y 30 minutos; sociolecto bajo, 15 horas y 50 minutos.

CUADRO 7. Distribución del corpus sociolingüístico de *PRESEEA-Valencia*

	GENERACIÓN 1 (20 A 34 AÑOS)		GENERACIÓN 2 (35 A 54)		GENERACIÓN 3 (54 O MÁS)		TOTAL
	H	M	H	M	H	M	
GRADO DE INSTRUCCIÓN 1	4	4	4	4	4	4	24
GRADO DE INSTRUCCIÓN 2	4	4	4	4	4	4	24
GRADO DE INSTRUCCIÓN 3	4	4	4	4	4	4	24
TOTAL	12	12	12	12	12	12	72

c) Málaga (Vida Castro 2007; Ávila Muñoz, Lasarte y Villena Ponsona 2008)

El corpus malagueño consta de unas 100 horas y 58 minutos de grabación; la muestra se recogió en 2003 y cada grabación osciló entre una hora y hora y media.

CUADRO 8. Distribución del corpus sociolingüístico de PRESEEA-Málaga

	GENERACIÓN 1 (20 A 34 AÑOS)		GENERACIÓN 2 (35 A 54)		GENERACIÓN 3 (54 O MÁS)		TOTAL
	H	M	H	M	H	M	
GRADO DE INSTRUCCIÓN 1	4	4	4	4	4	4	24
GRADO DE INSTRUCCIÓN 2	4	4	4	4	4	4	24
GRADO DE INSTRUCCIÓN 3	4	4	4	4	4	4	24
TOTAL	12	12	12	12	12	12	72

d) Granada (Moya Corral 2007)

Respecto del corpus granadino hay que aclarar que no se analizaron los informantes del grado de instrucción 1, por lo que si bien este corpus tiene 54 entrevistas, en nuestro análisis solo se revisaron 36.

CUADRO 9. Distribución del corpus sociolingüístico de PRESEEA-Granada

	GENERACIÓN 1 (20 A 34)		GENERACIÓN 2 (35 A 54)		GENERACIÓN 3 (54 O MÁS)		TOTAL
	H	M	H	M	H	M	
GRADO DE INSTRUCCIÓN 2	3	3	3	3	3	3	18
GRADO DE INSTRUCCIÓN 3	3	3	3	3	3	3	18
TOTAL	6	6	6	6	6	6	36

e) Sevilla (Lamíquiz y Pineda 1983; Lamíquiz y Roperó 1985; Ollero y Pineda 1992).

El corpus de Sevilla se diferencia de los que ya hemos descrito porque pertenece al *Proyecto de estudio de la norma lingüística culta*. Así, la variable *grado de instrucción* aparece como *nivel culto*, *medio* y *popular*. En cuanto a la edad, el corpus sevillano aunque está estratificado en tres generaciones, sus límites son distintos a los establecidos por *PRESEEA*.

CUADRO 10. Distribución del corpus sociolingüístico de Sevilla

	GENERACIÓN 1 (MENOS DE 30)		GENERACIÓN 2 (30 A 45)		GENERACIÓN 3 (45 O MÁS)		TOTAL
	H	M	H	M	H	M	
NIVEL CULTO	3	3	3	3	3	3	18
NIVEL POPULAR	3	3	3	3	3	3	18
NIVEL MEDIO	3	3	3	3	3	3	18
TOTAL	9	9	9	9	9	9	54

De igual manera que en los corpus *PRESEEA*, el *nivel sociolingüístico* de los hablantes sevillanos se decidió según la cantidad de estudios formales cursados: el nivel culto corresponde a aquellos con estudios universitarios superiores; el nivel medio, a estudios de grado medio; del nivel popular, señalan al respecto Ollero y Pineda: “Aquí pueden incluirse individuos que han cursado los estudios primarios e incluso los que han iniciado cursos de bachillerato o formación profesional (...) hemos seleccionado informantes con grados muy distintos de instrucción: desde los que han iniciado la enseñanza secundaria hasta informantes sin apenas estudios o analfabetos.” (1992: 17). Otro aspecto que debemos señalar, finalmente, tiene que ver con el tiempo en que se recogió este corpus y que se ubica en la década de los 80.

f) Canarias (Samper Padilla, José Antonio, Clara Eugenia Hernández Cabrera y Magnolia Troya. 1998).

Del *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico*⁵ se analizaron las 14 entrevistas pertenecientes a hablantes canarios de nivel culto (periodistas, abogados, arquitectos, médicos, profesores universitarios), 7 hombres y 7 mujeres, de edades comprendidas entre 26 y 68 años). La finalidad de analizar esta muestra de habla canaria fue representar en el corpus sincrónico la mayor variedad dialectal del español contemporáneo.

5.2.2.2. Americanos

Hemos analizado tres corpus completos, correspondientes a ciudad de México, Ciudad de Guatemala y Cobán; además, analizamos una muestra bastante más reducida de los corpus sociolingüísticos de Caracas y de Montevideo.⁶ De igual modo, se incorporaron otros corpus de algunas variedades americanas con la finalidad de comparar y registrar los usos de *ser* y *estar* en EE en el español americano.

⁵ Estos materiales pertenecen al “Proyecto de estudio coordinado de la norma lingüística culta de las principales ciudades de Hispanoamérica” (PILEI), dirigido por Lope Blanch: “Surgió con el propósito de determinar cuáles son los hechos lingüísticos propios de cada norma geográfica, de cada dialecto culto hispánico, identificando los que caracterizan a cada norma y los que las diferencian” (Moreno Fernández 2005:320).

⁶ Nos referimos al número total de informantes por comunidad.

a) Ciudad de Guatemala y Cobán, Guatemala. (Verdugo de Lima, Palma Chacón, Fong de Rivera Acevedo-Halvick 2007).

De Guatemala se tomaron dos corpus sociolingüísticos, el de Ciudad de Guatemala y Cobán; como señalan las editoras, el proyecto general (*Proyecto para el estudio sociolingüístico del español de las ciudades de Guatemala*), que contempla cinco regiones, se llevó a cabo en tres fases en las que se cubrieron: i) Ciudad de Guatemala; ii) las ciudades monolingües de las regiones del oriente y sur del país; iii) las ciudades bilingües y multilingües de las regiones de occidente y del norte del país. En ciudad de Guatemala, monolingüe en español, se realizaron 72 entrevistas y 54 en Cobán, una región del norte del país caracterizada por el bilingüismo español- Q'eqchi', aunque la muestra *PRESEEA* es monolingüe en español. Se tienen, en total, 126 entrevistas estratificadas por sexo, edad y grado de instrucción y que representan, sin duda alguna, una extraordinaria muestra para el estudio del español centroamericano.

CUADRO 11. Distribución del corpus sociolingüístico de *PRESEEA-Guatemala*

CIUDAD DE GUATEMALA	GENERACIÓN 1 (20 A 34 AÑOS)		GENERACIÓN 2 (35 A 54)		GENERACIÓN 3 (54 O MÁS)		TOTAL
	H	M	H	M	H	M	
GRADO DE INSTRUCCIÓN 1	4	4	4	4	4	4	24
GRADO DE INSTRUCCIÓN 2	4	4	4	4	4	4	24
GRADO DE INSTRUCCIÓN 3	4	4	4	4	4	4	24
TOTAL	12	12	12	12	12	12	72

CUADRO 12. Distribución del corpus sociolingüístico de *PRESEEA-Guatemala (2)*

COBÁN	GENERACIÓN 1 (20 A 34 AÑOS)		GENERACIÓN 2 (35 A 54)		GENERACIÓN 3 (54 O MÁS)		TOTAL
	H	M	H	M	H	M	
GRADO DE INSTRUCCIÓN 1	3	3	3	3	3	3	18
GRADO DE INSTRUCCIÓN 2	3	3	3	3	3	3	18
GRADO DE INSTRUCCIÓN 3	3	3	3	3	3	3	18
TOTAL	9	9	9	9	9	9	54

b) Ciudad de México (Lastra y Martín-Butragueño 2007)⁷

De Ciudad de México se analizaron 108 grabaciones, parte del *Corpus Sociolingüístico de la ciudad de México 1997-2007*.

CUADRO 13. Distribución del corpus sociolingüístico de *PRESEEA-Ciudad de México*

	GENERACIÓN 1 (20 A 34 AÑOS)		GENERACIÓN 2 (35 A 54)		GENERACIÓN 3 (54 O MÁS)		TOTAL
	H	M	H	M	H	M	
GRADO DE INSTRUCCIÓN 1	6	6	6	6	6	6	36
GRADO DE INSTRUCCIÓN 2	6	6	6	6	6	6	36
GRADO DE INSTRUCCIÓN 3	6	6	6	6	6	6	36
TOTAL	18	18	18	18	18	18	108

c) Caracas (Bentivoglio y Malaver 2006)

De Caracas se analizaron 36 grabaciones, distribuidas en los grados de instrucción 1 y 3, con igual cantidad de informantes por generación y sexo. El análisis de Caracas-*PRESEEA* representa un eslabón determinante en el estudio de la evolución de *estar* en EE en el dialecto caribeño.

CUADRO 14. Distribución del corpus sociolingüístico de *PRESEEA-Caracas*

	GENERACIÓN 1 (20 A 34 AÑOS)		GENERACIÓN 2 (35 A 54)		GENERACIÓN 3 (54 O MÁS)		TOTAL
	H	M	H	M	H	M	
GRADO DE INSTRUCCIÓN 1	3	3	3	3	3	3	18
GRADO DE INSTRUCCIÓN 3	3	3	3	3	3	3	18
TOTAL	6	6	6	6	6	6	36

d) Montevideo (Elizaincín, Urse y Rodríguez 2009)⁸

La selección del habla de Montevideo está compuesta por 20 hablantes, 4 hombres y 16 mujeres, de distintos grados de instrucción y edades. No tomaremos en cuenta los factores sociales para el análisis, dada la distribución desigual de la muestra. Para el desarrollo de la investigación resultaba esencial incorporar datos de las

⁷ <<http://www.lef.colmex.mx>>

⁸ www.lenguas.net/equipos/Montevideo

variedades dialectales del español austral,⁹ pues constituye una modalidad de la que carecemos estudios sobre *ser* y *estar* y que puede ilustrar la evolución del fenómeno en el cono Sur.

e) San Juan de Puerto Rico, Bogotá, otras ciudades del *Macrocorpus de la Norma culta*

Se reunieron distintos corpus, cuya composición no alcanzaba la cantidad mínima requerida por muestra para garantizar la comparabilidad de los datos, pero, que son confiables pues forman parte de macroproyectos como *PRESEEA*, en el caso de Puerto Rico; el *PILEI*, *Macrocorpus de la Norma culta* (se trata de las ciudades de Buenos Aires, La Paz, Santiago de Chile, San José de Costa Rica, Lima);¹⁰ los materiales de *El español hablado en Bogotá* (Montes Giraldo 1997, 1998), con el habla de Bogotá. En su conjunto, estos corpus arrojan información geo-social sobre el fenómeno de *estar* con adjetivos de edad.

El corpus oral *PRESEEA* analizado consta de 572 grabaciones, 288 pertenecientes a hablantes peninsulares (50%) y 284 a hablantes americanos (50%), todas de estilo semiformal y que representan 590 horas de grabación (45min. cada una como duración mínima); todas estas características hacen posible un estudio dialectológico y sociolingüístico, los enfoques que conforman las dimensiones analíticas para la descripción de *ser* y *estar* con adjetivos de edad. Como se aprecia, los corpus peninsulares representan tres áreas dialectales de España: las hablas del centro-norte, las hablas andaluzas y la comunidad valenciana; de América se cuenta con la muestra centroamericana, el español de México, el caribeño y el del Cono Sur.

⁹ *Español austral*: Argentina, Paraguay y Uruguay.

¹⁰ Excluimos Caracas, Ciudad de México, Bogotá y San Juan.

5.3. ANÁLISIS LÉXICO-SEMÁNTICO

Se identificaron todas las unidades con función de adjetivo de edad (sustantivos y adjetivos) con la finalidad de describir su evolución a partir de los datos del corpus. Las acepciones se consultaron en los siguientes diccionarios:

- 1) Real Academia Española (2002). *Diccionario de la lengua española* [citado *DRAE*];
- 2) Real Academia Española (2005). *Diccionario Panhispánico de Dudas* [citado *DPD*];
- 3) Real Academia Española (1984). *Diccionario de Autoridades* [citado *DA*];
- 4) Corominas Joan y J. A. Pascual. (1980-1991). *Diccionario crítico etimológico castellano e hispánico*. [citado *DCE*];
- 5) Covarrubias, Santiago de. (2006). *Tesoro de la lengua castellana o española* [citado *TLCE* o *Covarrubias*];
- 6) Moliner, María. (2002). *Diccionario de uso del español* (3ª edición). [citado *DUE*];
- 7) Seco, Manuel, Andrés Olimpia y Gabino y Ramos (1999). *Diccionario del español actual* [citado *DEA*];
- 8) Suances, Jaimes. (2000). *Diccionario del Verbo Español, Hispanoamericano y Dialectal* [citado *DVEHD*]

Además, se consultó el *Léxico hispanoamericano del siglo XVI*, de Boyd-Bowman (1972). Para la descripción de los adjetivos, tomamos el estudio de Corrales Zumbado (1982), *El campo semántico ‘edad’ en español*, investigación en la cual se describe diacrónica y sincrónicamente el campo semántico ‘edad’;¹¹ sus resultados revisten especial importancia para nuestra investigación pues se documentan las apariciones de los sustantivos y adjetivos que forman el campo semántico de ‘edad’ en obras literarias (del siglo XV al XX). Es, por esto, que se ha tomado como una obra de consulta para la descripción de los adjetivos extraídos del corpus histórico.

¹¹ Este trabajo se inscribe en el campo de la semántica estructuralista; Corrales Zumbado establece los términos que componen el campo semántico y documenta su uso en textos latinos y textos del siglo XIII al XX.

5.4. VARIABLES, PREGUNTAS, MÉTODOS

5.4.1. Estudio dialectológico y sociolingüístico

Nuestro estudio trata sobre el proceso general de cambio de *estar* como verbo copulativo; es, en este sentido, un estudio histórico sobre la extensión de esta cópula con determinados adjetivos calificativos.

Es, también, el estudio del fenómeno de *estar* con adjetivos de edad en dos conjuntos dialectales: el español de América, el español de España y, a la vez, es el estudio de estos dos complejos dialectales en algunas de sus variedades regionales, esto es, el español de Alcalá de Henares o el de ciudad de México (las cuales, de cara al conjunto mayor al que pertenecen, poseen sus propias características), contrastando los datos de acuerdo con las características sociales y los factores lingüísticos que inciden en el mismo.

Se estudia una muestra diacrónica, diatópica y sociolingüísticamente diferenciada del español; los resultados del análisis del corpus sincrónico nos permiten caracterizar sociolingüísticamente el fenómeno de la alternancia de *ser* y *estar* en EE, en un conjunto de comunidades de habla, estableciendo frecuencias de uso y correlaciones estadísticas que describen la manera en que se da el fenómeno y su estado de evolución.

Este análisis de cada modalidad inscribe nuestro estudio en la sociolingüística y en la dialectología, dos niveles de la variación lingüística: la dimensión espacio-temporal y la dimensión socio-situacional: “La longitud de cada una de las dimensiones viene determinada por la extensión geográfica e histórica de cada lengua y por su complejidad sociolingüística” (Moreno Fernández 2005a: 135).

5.4.2. *Ser* y *estar* con adjetivos de edad: ¿un fenómeno de variacionismo?

Nuestro trabajo se inscribe en el marco general de la sociolingüística, en la medida en que el objetivo de nuestra investigación es describir la influencia de los factores sociales en la alternancia de *ser* y *estar* con una clase de los adjetivos calificativos, a la luz del proceso del cambio lingüístico. La edad, instrucción o sexo constituyen las variables explicativas para interpretar los resultados como un fenómeno de variación estable o de cambio en curso.

La decisión de analizar sólo un tipo de adjetivo en una muestra histórica y sincrónica se basa en la necesidad de caracterizar el continuo evolutivo del proceso de extensión de *estar* en distintas modalidades geosociales de la lengua. Nuestra finalidad es constatar la existencia de las diferencias temporales, geográficas, sociolectales y diafásicas y analizar, consecuentemente, un estado de la evolución. La variación que estudiamos es el resultado de un cambio que culminó su primer ciclo en el siglo XVI, período en el que se puede considerar la penetración o extensión de *estar* como innovadora, frente a *ser*.

Este proceso produjo una diferenciación semántico-pragmática que se instaló en el sistema y que lo hace singular en comparación con otras lenguas romances. Así, el adjetivo que, acompañado con *ser*, se interpreta de una manera determinada, acompañado con *estar*, puede interpretarse de otra. Asistimos a una penetración paulatina de *estar* con adjetivos calificativos, proceso que produce, a su vez, diferenciaciones semántico-pragmáticas cada vez más complejas de sistematizar. Desde esta perspectiva, y de partida, el adjetivo de edad con *ser* no quiere decir lo mismo que con *estar*, más aún si toman en cuenta todos los modificadores que pueden aparecer y los distintos adjetivos. Por otra parte, en Malaver (2001) revisamos las limitaciones del enfoque variacionista, pues la reducción del contexto de variación supone una serie de escollos metodológicos que se resumen en la obtención del número suficiente de casos por variantes.

En este estudio, en conclusión, no definimos la alternancia de *ser* y *estar* en EE como un caso de o en variación (desde el enfoque sociolingüístico variacionista) pero sí como un fenómeno de variación en la lengua, cuya descripción debe tomar en cuenta el papel de los modificadores, el tipo de adjetivo, los factores sociales, como la edad o el grado de instrucción y otros, más complejos y generales, vinculados a los cambios sociodemográficos y culturales de las comunidades, aspecto que nos parece fundamental y que R. Caravedo resalta acertadamente cuando afirma:

(...) además de las características fundamentales para identificar los fenómenos lingüísticos variables en la perspectiva de la variación [a saber, contextualizada lingüística y extralingüística (...)], hay que tener en cuenta los marcos históricos, espaciales, culturales más amplios, y los psicológicos-cognitivos particulares, en que

tales fenómenos ocurren, pues pueden ser determinantes en su configuración y fisonomía. (Caravedo 2003a: 547)

Por lo anterior, para el análisis empleamos un conjunto de variables: i) el verbo; ii) el adjetivo de edad; iii) la complejidad sintáctica de la EE (presencia de modificadores); iv) el tiempo y aspecto verbal; v) la persona. De igual manera, hablaremos de *variables extralingüísticas* o la edad, sexo, grado de instrucción.

5.4.3. Preguntas de investigación

- 1) ¿Cuáles son los mecanismos predicativos de la edad con *ser* y *estar*?
- 2) ¿Los usos de *ser* y *estar* con adjetivos de edad representan un avance en el proceso de cambio de *estar* como verbo copulativo?
- 3) ¿Cómo se manifiesta la variación de las EE con *ser* y *estar* en la geografía de la lengua?; ¿puede considerarse *estar* con adjetivos de edad como un americanismo sintáctico?
- 4) ¿Cuál es la relación entre los factores sociales y los usos de las cópulas en cada comunidad de habla?

5.4.4. Análisis: factores lingüísticos y sociales

a) *Cópula*

La presencia de una cópula u otra distingue el tipo de predicado en: predicado ‘gnómico’, con *ser*, y predicado ‘episódico’, con *estar* (o ‘EE gnómicas’ y ‘EE episódicas, respectivamente):

Yo *soy* viejo → EE ‘gnómica’

Yo *estoy* viejo → EE ‘episódica’

Se excluyen todos los aquellos en los que el adjetivo aparece con un determinante: *Yo soy un viejo* y en los que el adjetivo modifica a un sustantivo: *yo soy un hombre mayor, ella era ya una joven adolescente*.¹²

¹² Al final de cada EE, aparece el código de identificación del hablante. Las primeras cuatro letras indican la ciudad, a saber: ALCA: Alcalá de Henares; MEXI: Ciudad de México; GUAT: Ciudad de Guatemala; GRAN: Granada; MALA: Málaga; VALE: Valencia; CARA: Caracas; MONT: Montevideo. El número

b) *Tipo de adjetivo*

Se construyó un listado de sustantivos y adjetivos con función de adjetivo de edad; se tomaron en cuenta distintos diccionarios, manuales dialectológicos y el conjunto de investigaciones previas ya reseñadas. El listado consta de 20 voces, con sus respectivos diminutivos y aumentativos, en forma singular y plural (propias de los dialectos peninsulares y americanos): *adulto, anciano, chamaco, chaval, chavo, chico, chiquito, chiri, grande, joven, hombre, maduro, mujer, muchacho, mayor, menor, niño, patojo, pequeño, viejo*.

c) *Complejidad sintáctica: EE sustanciales y EE modificadas (con modificadores)*

Las EE se clasificaron según la presencia de los modificadores, en *EE sustanciales* y *EE modificadas*, como se esquematizan a continuación:

Clasificación de la expresión de edad construida con <i>ser</i> y <i>estar</i>	
Modificadas con <i>ser</i>	Modificadas con <i>estar</i>
Yo <i>soy</i> viejo	Yo <i>estoy</i> viejo
Yo <i>soy</i> <u>mu</u> y viejo	Yo <i>estoy</i> <u>mu</u> y viejo
Yo <i>soy</i> <u>tan</u> viejo	Yo <i>estoy</i> <u>tan</u> viejo
Yo <i>soy</i> <u>ya</u> tan viejo	Yo <i>estoy</i> <u>ya</u> <u>tan</u> viejo

La finalidad de incorporar el papel de estas partículas es analizar su efecto sobre las EE: 1) determinar de qué modo se combinan con *ser* y 2) analizar si su presencia se relaciona con la extensión de *estar* en este contexto adjetival.

d) *Tiempo verbal y persona*

Se analizaron el tiempo verbal de la EE y el sujeto, para determinar su relación con la presencia de cada cópula y los modificadores.

que sigue a continuación indica la edad del hablante (Generación 1, 2 y 3) y el último dígito señala su grado de instrucción (como se especifica en los cuadros de cada corpus).

e) *Sexo, edad y grado de instrucción*

Las variables que se toman en cuenta para el estudio son las de las distintas muestras de habla; como ya se vio en el cap. II, la edad, el sexo y el grado de instrucción son esenciales para el análisis sociolingüístico de la investigación.

Asimismo, al comparar distintas modalidades, estudiamos las diferencias dialectales en el uso de *ser* y *estar* con adjetivos de edad, por lo que combinaremos, entonces, la sociolingüística y dialectología para dar cuenta de la evolución de *estar* como verbo copulativo.

5.4.5. Análisis estadístico

Los datos se analizaron con el programa SPSS (*Statistical Package for the Social Sciences*), con el cual obtuvimos las frecuencias y se calculó la X^2 . Como consecuencia de la argumentación expuesta en 5.3.2. se ha prescindido del análisis multivariable. En cuanto al análisis cuantitativo, la tabulación cruzada permite sistematizar la acción de los factores sociales en el comportamiento lingüístico y dar cuenta del modo en que ambos interactúan.

5.5. RESUMEN DE LA METODOLOGÍA

En esta investigación se constituyó un corpus textual, para el estudio de los siglos XVI al XIX, con base en un conjunto de documentos históricos, producidos en España y América. Asimismo, se ha constituido un macrocorpus sincrónico con muestras de habla -estratificadas según el grado de instrucción, sexo y edad de los informantes- pertenecientes a cinco modalidades dialectales peninsulares y cinco americanas, recogidas todas durante la última década del siglo XX.

En este conjunto de materiales escritos y orales se analizó la alternancia de *ser* y *estar* con adjetivos de edad en una construcción del tipo *María ES/ESTÁ muy mayor*, a la que se ha denominado *expresión de edad*.

Se ha partido del período de consolidación del sistema atributivo de la lengua, a partir de la extensión de los usos de *estar*, de verbo locativo a verbo copulativo y se estudió la evolución de *estar* con estos adjetivos en el continuo dialectal de la lengua española; analizadas las muestras sincrónicas, se determinaron los factores lingüísticos

y sociales relacionados con el fenómeno que nos ocupa. Se ha emprendido, en síntesis, un estudio dialectológico y sociolingüístico, diacrónico y sincrónico, de los mecanismos de predicación de la edad con *ser* y *estar*.