

Fundación Universidad de las Américas, Puebla

Escuela de Artes y Humanidades

Departamento de Lenguas



Tesis presentada por Antonio Rico Sulayes como requisito parcial
para obtener el grado de Maestría en Lingüística Aplicada
Dirigida por el Dr. Roberto Herrera Herrera y aceptada por
el Departamento de Lenguas

Santa Catarina Mártir, Puebla, Otoño de 2004

El potencial de ejemplificación del léxico subestándar en un corpus electrónico
El caso del español en México

Esta propuesta ha sido leída y aprobada por los
miembros del Comité de Tesis de

Antonio Rico Sulayes

como requisito parcial para la obtención del grado de

Maestría en Lingüística Aplicada

Director

Dr. Roberto Herrera Herrera
Universidad de las Américas, Puebla

Asesores

Dr. Rafael Rodríguez Marín
Real Academia Española

Dra. Vanessa Marchand
Universidad de las Américas, Puebla

Contenido

Lista de tablas y figuras

1 Introducción

- 1.01 Proyecto de investigación
- 1.02 Relevancia del proyecto
- 1.03 Planteamiento del problema

2 Revisión bibliográfica

- 2.01 Componentes teóricos del léxico del subestándar
 - 2.01.01 Variación lingüística
 - 2.01.02 Dimensiones de la variación lingüística
 - 2.01.03 Variación social o diastrática
 - 2.01.04 Variación situacional o diafásica
 - 2.01.05 Variación lingüística y subestándar
- 2.02 Antecedentes de la lexicografía hispana del subestándar
- 2.03 Algunos trabajos de referencia de obras lexicográficas del español
- 2.04 Otra búsqueda de la lexicografía del subestándar en México
- 2.05 Corpus electrónicos en la lexicografía
- 2.06 Papel de la ejemplificación en la lexicografía
- 2.07 Planteamiento de la pregunta de investigación

3 Metodología

- 3.01 Fuentes lexicográficas
- 3.02 Fuentes secundarias
- 3.03 Fuentes primarias
- 3.04 Diseño
- 3.05 Expectativas y limitaciones
- 3.06 Procedimientos
- 3.07 Identificación del léxico del subestándar en los diccionarios fuente
 - 3.07.01 *DEUM*
 - 3.07.02 *DRAE*
 - 3.07.03 *DBM*

3.07.04 *DTV*

3.07.05 *PDLPM*

3.07.06 *AHDOM*

3.07.07 *DIME*

3.08 Manejo de las listas de fuentes secundarias

3.09 Organización del muestreo aleatorio estratificado

3.10 Documentación de las listas del muestreo

4 Análisis de los datos

4.01 Representación de las listas del muestreo en el corpus

4.02 Limitaciones en el conteo de resultados

4.03 Posibilidades de ejemplificación de las fuentes secundarias

4.04 Discusión

5 Bibliografía

Lista de tablas y figuras

Tablas

- Tabla 3.1. Lista de diccionarios fuente del léxico subestándar del español en México
- Tabla 3.2. Conteo por letra de los lemas con marcas del subestándar en el *DEUM*
- Tabla 3.3. Conteo por letra de los lemas con marcas del subestándar en el *DRAE*
- Tabla 3.4. Conteo por letra de los lemas con marcas del subestándar en el *DBM*
- Tabla 3.5. Conteo por letra de los lemas a utilizar del *DTV*
- Tabla 3.6. Conteo de entradas y formas lematizables en el *PDLPM*
- Tabla 3.7. Conteo por letra del lemario del *AHDOM*
- Tabla 3.8. Conteo por letra de los lemas con marcas del subestándar en el *DIME*
- Tabla 3.9. Fuentes secundarias del léxico subestándar del español en México
- Tabla 4.1. Resultados de los ejemplos obtenidos en el *CREA*
- Tabla 4.2. Resultados de la muestra y comportamiento potencial de la población
- Tabla 4.3. Conteo por letra del lemario del *DAE*
- Tabla 4.4. Conteo por letra del lemario del *FAE*
- Tabla 4.5. Conteo por letra del lemario del *ODMS*

Figuras

- Figura 2.1. Localización del subestándar dentro de la variación lingüística
- Figura 3.1. La *lista DRAE* en la primera etapa de trabajo con las fuentes secundarias
- Figura 3.2. La *lista DRAE* en la segunda etapa de trabajo con las fuentes secundarias
- Figura 3.3. La *lista DRAE sample* en la tercera etapa de trabajo con las fuentes secundarias
- Figura 3.4. Hoja de datos del programa SPSS para Windows
- Figura 3.5. Opciones para la obtención de un muestreo aleatorio en SPSS para Windows
- Figura 3.6. La *lista fusionada sample* utilizada durante la documentación
- Figura 3.7. Llenado de la *lista DRAE documentada*
- Figura 3.8. Página principal de la RAE y del *CREA*

Figura 3.9. Resultados de consulta en el *CREA*

1 Introducción

1.01 Proyecto de investigación

En este trabajo se pretende explorar hasta qué punto resulta factible llevar a cabo la ejemplificación de un diccionario de léxico subestándar por medio de un corpus electrónico. Por ello, comenzaremos revisando cómo se definen dentro de la lexicografía los conceptos de *subestándar*, de *corpus electrónico* y de *diccionario ejemplificado o con citas*.

Entre los distintos tipos de obras lexicográficas se encuentran los llamados diccionarios de la lengua hablada o diccionarios del subestándar. El subestándar en lexicografía es entendido como un ámbito de la lengua en donde se mezclan una serie de niveles, tales como el coloquial, el familiar, el popular o el vulgar. Además, en las obras lexicográficas hispanas es común encontrar bajo el mismo rubro un número de vocablos de origen jergal, es decir sociolectal. Estos vocablos proceden de una variedad de sociolectos, como el de la delincuencia, el de los jóvenes estudiantes o el de grupos humanos marginales (Haensch, 1997, p. 94). En este sentido, hay que aclarar que un sociolecto es el conjunto de elementos que identifican la forma de hablar de un grupo de hablantes que comparten ciertas características socioculturales (Briz, 1996, pp. 15). De esta manera, en la mayoría de los trabajos lexicográficos del subestándar, es común que coexista léxico perteneciente a diferentes esferas. Por un lado, nos encontramos con un léxico que tiene una marcación diastrática, es decir, que caracteriza la manera de hablar de ciertos grupos sociales. Por otro lado, aparece un tipo de léxico que tiene una marcación diafásica, es decir, un léxico cuyo uso depende de la situación en que es utilizado. En este mismo

sentido, también es común encontrar incluido en los diccionarios del subestándar el léxico que tiene una marcación propia del tabú lingüístico, es decir, que forma parte de las groserías. Respecto a este último tipo de léxico, es pertinente mencionar que éste no corresponde a un nivel de lengua o diastrato, sino que representa una connotación especial en el uso social; es más bien una variación diafásica o pragmática, y por tanto, su utilización depende de las características del contexto en que se usa y de la adecuación entre el significado de las voces tabuizadas y la intención del hablante. De cualquier manera, siendo parte de la variación diafásica, es muy frecuente que se incluya el léxico tabuizado en el léxico del subestándar (Haensch, 1982b, p. 144). Así pues, la consideración de estos dos tipos de marcas, las diastráticas y las diafásicas, como un solo conglomerado, al menos dentro del campo de la lexicografía, parece normal e incluso justificada (Haensch, 1997, p. 94). Ahora bien, como hace notar Haensch (1982a), el problema que enfrenta la lexicografía en el tratamiento tanto del léxico subestándar como del léxico tabuizado está en el hecho de que muchos de los vocablos que conforman estos dos tipos de léxico, incluyendo aquellos vocablos que tienen una frecuencia de aparición muy alta, se hallan representados de manera sumamente irregular en los distintos textos escritos. En consecuencia, documentarlos de forma escrita es una tarea que en muchos casos ha dependido más bien de la casualidad (Haensch, 1982a, p. 442). Cabe recordar que la *causalidad*, según la Real Academia Española (RAE), es la “combinación de circunstancias que no se pueden prever ni evitar” (2001, p. 477). Es decir, lo que Haensch critica como “el gran problema de muchos diccionarios generales descriptivos” es la falta de regularidad que se ha tenido para la “documentación escrita” de “las unidades léxicas propias del lenguaje subestándar (familiar, popular, vulgar, tabuizado o jergal)”, que por tanto ha resultado “en gran parte fruto del azar” (1982a, p. 442).

En cuanto a la documentación de las voces del diccionario, una de las herramientas tecnológicas utilizadas para la búsqueda y el hallazgo de posibles instancias de una palabra es el corpus electrónico. Para entender qué es un corpus electrónico, hay que comenzar definiendo el concepto de *corpus* en lexicografía. El corpus es el conjunto de textos, tanto orales como escritos, en que aparecen las palabras que se tomarán en cuenta para su inclusión en el diccionario. Estos textos conforman los materiales lexicográficos en que se basará la elaboración de la obra (Porto, 2002, p. 84). Para el almacenamiento de estos materiales, en la lexicografía tradicional es común el empleo de fichas textuales (Sinclair, 1985, p. 83), las cuales conforman, en conjunto, el llamado fichero de referencia. Ahora bien, en informática un corpus es simplemente una cadena de textos electrónicos enlazados de forma secuencial y etiquetados de manera que puedan ser identificados según una serie de datos. Los datos que se incluyen en el etiquetado, como el autor, el título de la obra, entre otros, son aquellos que se consideran necesarios para la búsqueda e identificación posteriores de los materiales que conforman el corpus (Porto, 2002, p. 131). Así, a través de recursos informáticos, un corpus electrónico tiene la capacidad de localizar y desplegar todos los contextos en que una palabra ha sido usada en una gran base de datos (Biber, Conrad y Reppen, 1998, p. 22). Esto hace que el corpus electrónico supere cuantitativamente al uso de fichas textuales. En este sentido, hay que tener presente que las fichas textuales representan tan sólo los contextos que un lector humano, o un grupo de éstos, puede registrar sobre una palabra dada. Tales contextos, además, corresponden en muchos casos a los usos menos comunes de un vocablo, y eliden los usos estadísticamente más significativos (Biber, Conrad y Reppen, 1998, p. 26).

Por todo lo anterior, el corpus electrónico constituye una herramienta única para la lexicografía, proveyéndole con un material no sólo cuantioso sino además vigente, pues la

mayoría de los corpus que se crean en la actualidad están en constante actualización. Esto último hace que el corpus pueda ser representativo de un estado de lengua determinado (Lara, 2002, p. 5). Una ventaja más del corpus electrónico es que, debido a que es de cuño reciente y se apoya en el uso de las computadoras, puede recabar y almacenar una gran cantidad de datos de las más variadas fuentes, incluyendo materiales orales (Biber, Conrad y Reppen, 1998, p. 22). Por todo esto, los elementos provenientes de los corpus representan usos lingüísticos naturales y espontáneos que pueden contrastarse con las intuiciones del autor, como hablante nativo, tanto a nivel de uso del término como del significado o significados del mismo (Biber, Conrad y Reppen, 1998, pp. 24-25). Con todo, no hay que olvidar que los corpus electrónicos también presentan algunas limitaciones desde el punto de vista de la labor lexicográfica.

Una de las principales carencias de los corpus electrónicos la constituye la posible falta de representación de vocablos. Esta falta de representación puede tener lugar aun cuando dichos vocablos y sus acepciones sean de los más comunes. Así pues, muchas veces los diccionarios que han sido elaborados dando una marcada preponderancia a la intuición del autor registran vocablos y acepciones que no se encuentran en los corpus electrónicos. Esto se da de tal forma que puede incluso haber casos en que una acepción en concreto sea de las primeras en aparecer en la mente de los hablantes, y a pesar de ello no sea frecuente en el uso diario (Biber, Conrad y Reppen, 1998, p. 41). Si bien es posible que estos casos resulten aislados, su existencia debe ser tomada en cuenta. Debido a lo anterior, es necesario reconocer que las intuiciones del lexicógrafo no deberían ser descartadas en su totalidad, sino que más bien deberían considerarse como un complemento a los resultados obtenidos por medio de los corpus electrónicos (Biber, Conrad y Reppen, 1998, p. 41). Siguiendo este razonamiento, Biber, Conrad y Reppen nos dicen que en el caso concreto de

los diccionarios diafásicos y, por extensión, de los diccionarios del léxico subestándar, la utilización de fichas textuales podría continuar siendo recomendable (1998, p. 26). Ahora bien, sea cual sea el medio con que se documenten los vocablos del diccionario, la búsqueda de la documentación es condición necesaria para llevar a cabo la elaboración de un diccionario ejemplificado. Para entender mejor en qué consiste este tipo de diccionario, habría que comenzar por explicar primero en qué consisten los llamados *diccionarios con citas*.

Un diccionario con citas es aquel en que los textos en que ocurren las palabras que forman parte del diccionario aparecen en el mismo como ejemplos (Porto, 2002, p. 84). Las citas son referencias textuales utilizadas para justificar, por un lado, la inclusión de un vocablo en un diccionario y, por otro, la definición dada de dicho vocablo (Bajo, 2000, p. 48). El término de *diccionario ejemplificado*, por su parte, es un concepto más general que engloba tanto al diccionario en que los ejemplos han sido creados por quien lo redacta, como a aquel en que los ejemplos han sido extraídos de algún corpus (Bajo, 2000, p. 49). En este trabajo en particular, cuando se habla de la elaboración de un diccionario ejemplificado, se está haciendo referencia a esta última posibilidad. Es decir, por diccionario ejemplificado aquí se alude a un diccionario con citas procedentes de un corpus, específicamente de un corpus electrónico.

En cuanto al mérito de los diccionarios ejemplificados, una de sus principales ventajas está en el hecho de que a partir de un ejemplo concreto del discurso de un individuo es posible hacer toda una serie de conclusiones sobre el modo de empleo de los elementos utilizados en esa realización particular del discurso (Werner, 1982, p.75). En otras palabras, el usuario del diccionario puede inferir una gama de informaciones sobre el uso de un vocablo a partir de un ejemplo concreto de dicho uso. Biber, Conrad y Reppen

afirman que el aprendiz de una lengua necesita algo más que una definición del significado y una lista de sinónimos más o menos aproximados; quien busca el significado de una palabra en un diccionario necesita saber además cómo se usa de forma real el vocablo en cuestión (1998, p. 53), y una forma de conseguir que lo descubra es proporcionándole un ejemplo. Otros posibles beneficios de la ejemplificación, mencionados por Bajo (2000), son el recordatorio de la norma lingüística, de algunas formas difíciles y del funcionamiento sintáctico; la indicación de combinaciones habituales; y la inclusión de información enciclopédica (p. 50). Sin embargo, esta autora menciona también la existencia de varios riesgos implícitos en la ejemplificación, tales como el cuidado en la selección de ejemplos para voces con marcas de restricciones de uso, la manipulación de definiciones *ad hoc* y la proyección de ideología en la selección de los ejemplos (Bajo, 2000, pp. 50-51).

Para este estudio, cuyo contexto es el subestándar, el principal inconveniente es, como se ya se mencionó, la dificultad para obtener los ejemplos a través de un corpus electrónico (Biber, Conrad y Reppen, 1998, p. 41). Esta situación se deriva del hecho de que la lingüística moderna quizá todavía no ha desarrollado una forma suficientemente sistemática, exhaustiva y verificable para el reconocimiento de este tipo de léxico (Lara, 1997, p. 249). Sin embargo, como Sinclair hace notar, la utilización de recursos electrónicos como los corpus no debería posponerse en la labor lexicográfica. Por el contrario, se deben aprovechar aquellos recursos electrónicos que se tengan disponibles e intentar obtener el mayor rendimiento posible de ellos (1985, p. 86).

1.02 Relevancia del proyecto

Así pues, a pesar de las limitantes, la realización de obras lexicográficas que registran un léxico subestándar o tabuizado, es decir, popular o grosero, es completamente

factible. Haensch hace énfasis en este hecho, y comenta la existencia de este tipo de obras en otras lenguas como el francés (1982b, p. 150). A pesar de ello, este mismo autor llama la atención acerca de los raquíuticos avances de la lexicografía hispana respecto de este tipo de diccionarios diastrático-diafásicos (1982b, p. 150; 1997, p. 95). En cuanto al caso concreto de este tipo de diccionarios en México, hay que señalar que es posible encontrarse con un par de trabajos lexicográficos que resultan al menos ricos en materiales, el de Colín (1987/2001) y el de Jiménez (1999). Sin embargo, si a la observación de Haensch acerca de la pobreza en los avances de la lexicografía hispana del subestándar le añadimos la posibilidad de contar con un diccionario ejemplificado de este tipo de léxico para el español en México, el progreso en ese sentido resulta simplemente nulo. Por otro lado, la relevancia de la elaboración de un diccionario ejemplificado del subestándar no se desprende únicamente de la carencia de un trabajo similar previo, sino de la importancia, recién mencionada, que tiene la ejemplificación en sí para cualquier diccionario (Bajo, 2000, pp. 50-51; Bber, Conrad y Reppen, 1998, p. 53; Werner, 1982, p. 75).

1.03 Planteamiento del problema

En la busca de subsanar la carencia de un diccionario ejemplificado del español en México, este trabajo pretende utilizar uno de los pocos corpus electrónicos que existen para el español actual, de hecho, el único disponible al público en Internet, el *Corpus de referencia del español actual*, CREA (Porto, 2002, p. 129). Por medio de la utilización de este corpus, se intenta explorar hasta qué punto resulta viable llevar a cabo la ejemplificación de una lista exhaustiva de palabras o lemas pertenecientes al léxico subestándar del español en México. Esta lista de lemas se ha obtenido a partir de todas las fuentes secundarias, u obras lexicográficas predecesoras, que han surgido en los últimos

veinticinco años de labor lexicográfica respecto del español en México. La pregunta acerca de la posibilidad de llevar a cabo la ejemplificación de dichas fuentes se intenta contestar por medio de la documentación de un muestreo aleatorio de las mismas. Con la documentación del muestreo se pretende estimar el volumen potencial de la ejemplificación total del conjunto de lemas del subestándar provenientes de las obras lexicográficas predecesoras de este trabajo (ver sección 2.07).

Aquí cabe aclarar que por lema se debe entender la reducción de “todas las formas de una palabra a la forma paradigmática considerada fundamental (el infinitivo para todas las formas verbales del mismo paradigma, el singular para los sustantivos...)” (Bajo, 2000, p. 16). El lema es por tanto el encabezamiento de la *entrada* o unidad autónoma mínima del diccionario en la cual aparece cada una de las unidades léxicas incluidas en el mismo. Para mostrar mejor este concepto he decidido poner a continuación un ejemplo de entrada tomado del diccionario de Moliner (1998, p. 154):

amachinarse (de «a-²» y «machín», nombre aplicado al dios cupido) **1** (Am. C., Col., Méj.) prnl. recípr. **Amancebarse*. **2** (Guat., Pan.) prnl. **Abatirse*, *acobardarse*.

En esta entrada, nos encontramos con que el lema corresponde a la palabra *amachinarse* que aparece en letras redondas, minúsculas y resaltadas en negritas. La entrada está constituida por el párrafo, separado del resto del texto, que conforma el ejemplo en su totalidad. Si bien esta entrada, con tres líneas en su fuente original, se antoja corta, algunas otras entradas pueden ser mucho más largas. En este mismo diccionario, por ejemplo, nos encontramos la entrada *alzar*, aparecida poco antes de la entrada *amachinarse*, que alcanza casi las 60 líneas (Moliner, 1998, pp. 153-154). Igualmente, nos podemos

encontrar con la entrada *Am* con apenas cinco palabras en total. Ahora bien, aludiendo al concepto de entrada, nos encontramos con que en ésta, además del lema, aparece una definición del mismo, así como cualquier otra información relativa a la unidad léxica (Haensch, 1997, pp. 39-40). También habría que apuntar que en la definición del lema puede aparecer uno o más significados o acepciones. En el ejemplo de Moliner (1998) aquí presentado, constituido por la entrada de *amachinarse*, podemos ver que la definición contiene dos acepciones que están marcadas por los números uno y dos en negritas. Las acepciones están destacadas en este diccionario con letras cursivas, y el comienzo de cada una de ellas está indicado por un asterisco. En este mismo ejemplo nos encontramos con algunas de las informaciones relativas a la unidad léxica que Haensch (1997) menciona. Después del lema *amachinarse* aparece un paréntesis con información etimológica, relativa a los orígenes de la palabra¹. De manera similar, entre el número que indica el conteo progresivo de cada una de las acepciones del lema y la definición de cada acepción en sí, aparece otro paréntesis, de ámbito geográfico, que indica los lugares en donde tiene efecto la acepción subsiguiente². Además, las informaciones que se encuentran después de estos paréntesis de ámbito geográfico y antes de la definición misma son informaciones de carácter gramatical³. Así pues, una vez explicados los conceptos de lema y entrada, se puede entender más fácilmente el concepto de *lemario*, el cual es el conjunto de todos los lemas de un diccionario, y cuya contabilización equivale al total de entradas incluidas en el

¹ Respecto de este paréntesis, en él aparecen dos envíos o remisiones a otras entradas del diccionario. En el caso del primer envío, además, se incluye un superíndice, el número dos. Este superíndice indica qué número de acepción de la entrada a la que se remite es el que corresponde al significado con que se utiliza el término enviado en el paréntesis.

² En la primera acepción aparece el grupo de abreviaturas Am. C., que significa América Central, y las abreviaturas simples, Col. (Colombia) y Méj. (México). En la segunda acepción aparecen las abreviaturas Guat. (Guatemala) y Pan. (Panamá).

³ La abreviación *prnl.* que aparece en las dos acepciones presentadas en el ejemplo corresponde a la indicación *pronominal* que se puede hacer respecto de un verbo. La indicación *recípr.*, aparecida únicamente en la primera acepción, significa *recíproco*.

diccionario. El leuario, junto con sus entradas correspondientes, constituye por tanto la parte más substancial del cuerpo del diccionario.

La lista utilizada en este trabajo constituye tan sólo un leuario tentativo para un diccionario del español subestándar en México. Este leuario provisorio ha sido obtenido haciendo una revisión de todos los diccionarios (fuentes secundarias) aparecidos en los últimos veinticinco años, que por cuya construcción contuvieran lemas pertenecientes al léxico del subestándar del español en México. A este respecto debemos recordar que el subestándar aquí debe ser entendido como la combinación de la variación diastrática y la variación diafásica que se mencionó con anterioridad.

2 Revisión bibliográfica

2.01 Componentes teóricos del léxico del subestándar

Para comenzar esta sección, se hará una revisión de la teoría lingüística que subyace al concepto del subestándar. En este sentido, se abordará primero el estudio de la variación lingüística. Después, se mencionarán algunas de las propuestas teóricas que se han elaborado para establecer las distintas dimensiones en que dicha variación puede tener lugar. De estas dimensiones, se abordarán de manera independiente dos de ellas, las cuales son las relacionadas directamente con el subestándar, la dimensión de la variación social o diastrática y la dimensión de la variación situacional o diafásica. Finalmente, de estas dos dimensiones, se verá cuales son los aspectos y las categorías derivadas de ellas que forman parte del subestándar propiamente hablando.

2.01.01 Variación lingüística. En los fundamentos de los estudios del habla se encuentra el presupuesto de que los individuos usan la lengua de forma distinta entre ellos, e incluso, respecto de sí mismos en diferentes situaciones (Sherzer y Darnell, 1978/2000, p. 65). Según una serie de factores que se han examinado en este tipo de estudios, los individuos introducen con frecuencia variaciones en su forma de utilizar la lengua. Al estudiar estas variaciones en la forma de hablar de los individuos, los investigadores han podido detectar que la variación lingüística es un fenómeno que puede tener lugar entre lenguas distintas no emparentadas, entre dialectos de una misma lengua, entre distintos estilos de hablar, así como también entre sólo algunos elementos muy específicos de estos dialectos o estilos (Sherzer y Darnell, 1978/2000, p. 66). De acuerdo con esto, y haciendo a un lado el caso de la variación entre distintas lenguas, existen dos ámbitos primordiales de

la variación lingüística. Por un lado, la variación en el habla de los individuos es un fenómeno que se puede presenciar en los miembros de un grupo social respecto de los miembros de otro grupo social distinto (Holmes, 2001, p. 123), en cuyo caso estamos hablando de una variación a nivel de dialecto. Cabe aclarar que aquí dialecto se debe entender como un sistema lingüístico derivado de una lengua, de la cual se pueden derivar otros tantos sistemas similares, y que se diferencia tanto de la lengua como de los otros posibles sistemas según criterios geográficos o sociales; todo ello sin llegar a convertirse en una estructura lingüística con categoría de lengua en sí (RAE, 2001, p. 815). Al mismo tiempo, hay que mencionar que ninguna definición de dialecto es absoluta, pues la distinción entre una lengua, o idioma, y los posibles dialectos que de éste se derivan se basa comúnmente en la presencia o ausencia de diferencias léxicas, de diferencias de pronunciación y de la inteligibilidad mutua entre los hablantes de varios dialectos pertenecientes a una misma lengua. Sin embargo, todos estos criterios de delimitación no siempre aparecen de manera regular y por ello es difícil esgrimir cualquiera de ellos en particular como prueba definitiva de la existencia de un dialecto o lengua independiente (Holmes, 2001, pp. 128-130). Por otro lado, la variación se puede encontrar también en un mismo individuo que varía su forma de hablar una misma lengua en distintas situaciones, en este caso se dice que se trata de una variación a nivel de estilo (Milroy y Gordon, 1997, p. 198). El estilo, por tanto, debe ser entendido como la variación lingüística que refleja cambios en la situación en que se encuentra el hablante; situación que implica factores tales como el interlocutor, el lugar en el que se habla, el tema al que se refiere, entre otros (Holmes, 2001, p. 246). Por su parte, Sherzer y Darnell (1978/2000, p. 66) añaden que tanto la variación a nivel de dialecto como la variación a nivel de estilo bien pueden manifestarse en tan sólo algunos elementos muy particulares de cualquiera de los dos niveles. Así pues,

como consecuencia de estudiar el uso del habla, los investigadores han llegado a la conclusión de que los individuos en general tienen la capacidad de utilizar la lengua haciendo uso de una serie de variedades lingüísticas (Sherzer y Darnell, 1978/2000, p. 65), y que dichas variedades lingüísticas se manifiestan según diferentes factores sociales y situacionales.

Ahora bien, aquí hay que aclarar que aunque para este trabajo no es tan importante la variación lingüística que se presenta entre distintas lenguas, ésta ha tenido una gran relevancia para los estudiosos del habla. Así por ejemplo, si bien la posibilidad de encontrar la variación lingüística a nivel de diferentes dialectos no es comentada por Ryan, Giles y Sebastian, estos autores llaman la atención sobre la existencia de la variación lingüística entre varias lenguas, además de la variación a nivel estilístico (1982, p. 1). Esta variación entre lenguas diferentes, claro está, se ha hecho patente en el estudio de las sociedades y los individuos multilingües. Holmes, que también da cuenta de la posibilidad de la variación en la que se involucran lenguas distintas, nos dice que durante la variación lingüística los individuos se enfrentan a una selección de código o variedad lingüística, y que dicho código o variedad lingüística bien puede ser una lengua, un dialecto o un estilo (2001, p. 223).

Respecto de la variación lingüística en general, Labov pone de manifiesto que la variación social y la variación estilística representan la posibilidad que tienen los hablantes de comunicar un mismo mensaje de distintas maneras sin afectar su valor referencial o su contenido, sino únicamente alterando su valor social o su calidad expresiva (1983, p. 338). Así pues, con la finalidad de estudiar la variación lingüística en su contexto social, el mismo Labov elabora el concepto de *variable sociolingüística*. Esta variable sociolingüística es descrita como aquel elemento o rasgo lingüístico que se encuentra

relacionado con algún elemento o variable no lingüística, que forma más bien parte del contexto social, como los interlocutores por ejemplo (1983, p. 299). Este concepto, actualmente, es designado simplemente como *variable lingüística* y sigue siendo el constructo metodológico central para el análisis y la comparación de datos lingüísticos social y estilísticamente diferenciados (Milroy y Gordon, 2003, p. 88). Utilizando este concepto de la variable lingüística, Ferguson establece uno de los principios básicos en el estudio de la variación lingüística. De acuerdo con este autor, los conjuntos de marcadores o variables lingüísticas que identifican a un dialecto o a un estilo (que este autor designa con el término registro) conforman un sistema. Estos marcadores o variables lingüísticas presentan distintos grados de cohesión entre sí. De dicha cohesión, señala Ferguson, depende la percepción que los hablantes tengan del sistema como una entidad independiente o no (1994, p. 23).

2.01.02 Dimensiones de la variación lingüística. Respecto a cómo dimensionar las distintas posibilidades en que puede ocurrir la variación lingüística, los estudiosos del tema han aludido a una amplia serie de factores que la influyen. En este sentido, Holmes habla de dos posibles grupos de factores. Por un lado menciona aquellos factores que están relacionados con los diferentes usos que se hacen del lenguaje, y por otro lado señala a los usuarios mismos de la lengua (2001, p. 223). Scherzer y Darnell hablan, por ejemplo, de la necesidad de diferenciar las variedades lingüísticas de acuerdo con las diversas funciones que éstas cumplen en la sociedad. Al mismo tiempo estos autores mencionan, si bien lo hacen sólo de paso, algunos posibles binomios para establecer las dimensiones de la variación. De hecho, son tres los binomios que Scherzer y Darnell (1978/2000) nombran, el de lo formal frente a lo no formal, el de lo público frente a lo privado, y el de la pertenencia al grupo frente a la exclusión del grupo (p. 66). Lamentablemente, estos autores no

profundizan más allá en cómo operan estos binomios respecto de las posibles dimensiones que de ellos se derivarían. Por su parte, Labov hace algo similar en un primer momento y menciona, dentro de algunos de los axiomas derivados de su investigación sociolingüística, ciertos posibles ejes para estudiar las variedades lingüísticas. Dichos ejes son tres: la atención, que puede resultar en un discurso cuidado frente a uno casual; lo vernáculo, que genera un discurso menos cuidado frente a uno más sistemático; y la formalidad, que opone un discurso cuidado y sistemático frente a uno más casual y amistoso (Labov, 1983, p. 265). Sin embargo, a diferencia de Scherzer y Darnell (1978/2000), Labov señala más adelante en su estudio dos posibles ámbitos para distribuir las variables lingüísticas, y establece la existencia de dos tipos principales de estratificación, una social y otra estilística (Labov, 1983, p. 299). Con esto, Labov comienza a apuntar a una distribución de las variaciones lingüísticas que resulta relevante de manera directa para este trabajo, enfocado hacia el léxico subestándar. Esto se hace particularmente cierto si se toma en cuenta que, como ya se dijo, este tipo de léxico representa una colisión entre la variación diastrática o social y la variación diafásica o situacional. De cualquier forma, una revisión de algunos de los diferentes autores que tratan el tema de la variación lingüística permite ver que muchos de ellos hacen mayor énfasis en ciertos aspectos y dejan de lado otros al momento de intentar dimensionar dicha variación. Todo esto, por supuesto, obedece a los intereses particulares que guía a cada uno de ellos.

Ahora bien, aquí quiero comenzar haciendo de lado tanto las diferencias en las dimensiones utilizadas por los distintos autores en el estudio de la variación, así como la orientación particular de este trabajo. Quiero hacer esto para presentar dos caracterizaciones, complementarias entre sí, con las cuales se puede conformar una tipología para el estudio sistemático de la variación lingüística. Me refiero a la

caracterización de la variación lingüística de Ferguson (1994) y a la de Briz (1996). El primero de estos autores menciona que hay cuatro ámbitos de la variación lingüística cuya exploración resulta necesaria para entender los presupuestos de la labor sociolingüística. El primer tipo de variación es la dialectal, cuya unidad de representación es el dialecto. Este tipo de variación es descrita por Ferguson como la variación, manifiesta en los cambios de habla o escritura de los individuos, que depende del lugar de origen de los mismos (1994, p. 16). El segundo tipo de variación sobre el que llama la atención este autor es la variación relacionada con la ocasión o situación en que se usa la lengua. A la unidad de este tipo de variación Ferguson la llama registro (1994, pp. 16-17). Sobre el significado de este término, hay que hacer mención que se corresponde exactamente con lo que hasta ahora hemos llamado estilo (Holmes, 2001, p. 246); más adelante explicaremos el empalme que se presenta comúnmente entre el término *registro* y el término *estilo* (ver sección 2.01.04). En tercer lugar, Ferguson coloca la variación que se puede presentar entre distintos géneros, que de acuerdo con él son los distintos tipos de mensaje que se presentan constantemente en una comunidad y que desarrollan por tanto regularidades en su estructura interna (1994, pp. 17, 21). Finalmente, este autor menciona otro tipo de variación llamada diacrónica, que bien puede referirse a la forma en que cambia el uso de la lengua en un individuo a medida que varía su edad dentro de una misma sociedad, o bien puede representar los cambios de ese mismo individuo según entra en contacto con otras comunidades a lo largo de su vida (Ferguson, 1994, p.18). Así pues, los ámbitos de la variación para Ferguson son el del dialecto o de orden geográfico, el del registro o de orden situacional, el del género o de orden textual, y el diacrónico o de orden vital.

Otro esquema muy completo en el establecimiento de las dimensiones de la variación lingüística es el de Briz (1996). Para este autor, el estudio de los fenómenos de la

variación lingüística se divide en cuatro tendencias. Primero está el estudio de la variedad diacrónica, que según él es aquella que implica una diferencia en el tiempo y que ha sido abordada por la historia de la lengua. La segunda tendencia es la del estudio de la variedad diatópica, que es aquella que se diferencia en el espacio y que ha estado a cargo de la dialectología. El tercer tipo de estudio es el de la variedad diastrática, la cual involucra las características socioculturales del usuario y ha sido objeto de estudio de la sociolingüística. Por último se encuentra la variación diafásica, que está determinada por la situación de uso y el contexto comunicativo (Briz, 1996, pp. 15). Respecto de la disciplina que se encarga de estudiar este último tipo de variación, aunque Briz no la menciona directamente, podemos deducir que se trata de la pragmática (1996, p. 24). Excluyendo la dimensión diacrónica, que es la proyección de la variación a través del tiempo, las tres últimas variedades constituyen estados sincrónicos, o de un momento histórico dado, y se representan por las siguientes unidades: el dialecto, para la variedad diatópica; el sociolecto, en el caso de la variedad diastrática; y el registro, para la variación diafásica (Briz, 1996, pp. 15). En este último caso, la terminología de Briz coincide con la de Ferguson (1994) y por ello, su uso del término registro es equivalente también al concepto antes presentado bajo la etiqueta de *estilo*.

Comparando los dos esquemas de las dimensiones de la variación lingüística de Ferguson (1994) y de Briz (1996), podemos notar que ambos coinciden en la existencia de una variación de orden cronológico, que ambos llaman diacrónica, de una variación de orden geográfico, que para ambos está representada por los dialectos, y de una variación de orden situacional, cuya unidad es denominada por los dos como registro. Sin embargo, Briz (1996) menciona un tipo de variación que Ferguson no trata (1994), y que Labov (1983) ya había señalado antes, la variación de orden social o diastrática. Estos dos últimos tipos de

variación, la situacional y la social, como mencionamos antes, son las que requieren un comentario aparte en el caso concreto del estudio del subestándar.

2.01.03 Variación social o diastrática. Como Briz hace notar, la sociolingüística es la disciplina encargada de estudiar la variación lingüística que se deriva de las características socioculturales del hablante (1996, p. 15). Uno de los fundadores principales de esta disciplina es William Labov, quien desde sus primeros estudios, iniciados en los años sesenta, propuso intentar correlacionar los diferentes elementos lingüísticos observables en el habla de los individuos con elementos del contexto social (Labov, 1983, p. 299). Así, a partir de varios de sus estudios de la pronunciación en distintos grupos de hablantes, este autor propone considerar el estatus social de los individuos como una dimensión de la variación lingüística (Labov, 1983, p. 154). En este sentido, para poder introducir el estatus socioeconómico de los individuos como una variable de sus estudios, Labov se basó en los complicados índices de estatus socioeconómico que construían las investigaciones sociológicas del momento (1983, p. 154). Los indicadores en que se basaban dichas investigaciones, y que Labov utilizó en aquellos estudios pioneros, son los tres indicadores del estatuto productivo: la profesión o el trabajo de la persona que sostiene a la familia a la cual pertenece el hablante, el nivel de educación del hablante mismo y el ingreso familiar en el hogar del que el hablante forma parte (Labov, 1983, p. 157). Conforme ha evolucionado la sociolingüística, podemos encontrar que los estudios en que se intenta relacionar la lengua y el contexto social han propuesto otras dimensiones, a parte del estatus socioeconómico, para expresar esta relación. Holmes señala cuatro escalas o dimensiones (2001, p. 373). La primera de estas dimensiones es la solidaridad o la distancia social, es decir el grado de conocimiento y confianza que hay entre los interlocutores, así como su afinidad de actitudes y valores. La segunda dimensión está dada por el estatus o el

poder, es decir el estatus social, económico y cultural del grupo social al que pertenece el hablante (Holmes, 2001, p. 374). En tercer lugar está la formalidad del lugar o la situación en que se usa la lengua, opuesta al uso de ésta en una conversación casual o relajada. Finalmente, se encuentra la dimensión representada por la función de la interacción de los hablantes (Holmes, 2001, p. 375), es decir, si dicha interacción tiene como fin principal expresar sentimientos, dar a conocer información, lograr que el otro haga algo, comentar algo sobre el mensaje mismo, entre otros posibles fines (Holmes, 2001, pp. 258-260). De estas cuatro dimensiones, sin embargo, la segunda dimensión, del estatus social (en que Holmes y Labov coinciden), está relacionada directamente con la variación social, diastrática o de grupo. Las otras tres dimensiones son más bien situacionales; la primera dimensión depende de los interlocutores, la tercera del lugar o la situación y la cuarta de la interacción misma. Así pues, aquí nos hemos de centrar en la dimensión del estatus que sigue siendo la dimensión más importante de la variación social.

El eje de categorización de la variedad lingüística que se deriva del estatus sociocultural de sus hablantes es el de la lengua estándar frente a la no estándar. La variante estándar de una lengua es aquella que se define por un grupo de normas a las que se tilda de ‘correctas’ y que se incluyen en diccionarios, gramáticas y manuales de estilo. Este proceso de inclusión de ciertas formas lingüísticas es designado como *codificación*. La codificación de una variedad lingüística depende comúnmente del poder del grupo social que la habla. Así, se puede decir que la variedad estándar de una lengua es simplemente aquella que pertenece a los grupos sociales que detentan el poder. La aceptación de esta variedad, en detrimento de las otras variedades habladas en la misma sociedad, es promovida especialmente por ciertas instituciones sociales como el gobierno, la educación y los medios de comunicación (Ryan, Giles y Sebastián, 1982, p. 3). Las herramientas o pasos

comunes a seguir en la codificación de una variedad estándar son el desarrollo de un sistema de escritura y una ortografía, la confección de una gramática de la variedad en cuestión y la elaboración de un diccionario que registre su vocabulario (Holmes, 2001, p. 103). La variante o las variantes no estándar, por su parte, son la que habla el grupo, o grupos sociales menos privilegiados. En este sentido, todas las formas lingüísticas no pertenecientes a la variedad estándar, la cual es aquella que habla el grupo socialmente poderoso, son definidas como no estándar. Dichas formas lingüísticas bien pueden ser parte del vocabulario, de la gramática o de la pronunciación (Holmes, 2001, p. 133). Otros términos que se han utilizado para designar las categorías de las variedades lingüísticas derivadas del estatus social son el término de *nivel de lengua alto* o *elevado*, para la variedad estándar, y el término de *nivel de lengua bajo* o *popular*, para la variedad no estándar (Hodge y Kress, 1997, p. 53). Estos términos se derivan del concepto original de diglosia, el cual describe una comunidad en la cual se hablan dos variedades de una misma lengua (el nivel de lengua alto y el nivel de lengua bajo, antes mencionados), que se usan para distintas funciones, de manera complementaria, pero donde el nivel de lengua alto no es usado para la conversación diaria (Holmes, 2001, p. 27). Con todo, cómo la designación de no estándar, o bajo, ha acumulado presupuestos negativos que equiparan, por ejemplo, a estas formas lingüísticas con desviaciones inadecuadas de la lengua, algunos lingüistas proponen llamarlas formas vernáculas (Holmes, 2001, p. 133). Cabe aclarar aquí que la estandarización de una variedad lingüística en una sociedad no se basa en ningún tipo de propiedad intrínseca de dicha variedad, sino simplemente en su tratamiento social (Ryan, Giles y Sebastián, 1982, p. 3). De igual manera, ninguna de las variedades no estándar o vernáculas, de ninguna sociedad, se pueden caracterizar como lingüísticamente inferiores respecto de las variedades estándar con las cuales coexisten (Holmes, 2001, p. 133), es

decir, entre todas estas variedades existe al menos una igualdad potencial para desarrollarse en cualquier tipo de práctica lingüística, con la salvedad de que en muchas variedades aún falta el vocabulario, los patrones de discurso y los textos mismos para llevar a cabo tales prácticas (Hymes, 1996, p. 211).

Finalmente, para hacer una aclaración sobre el título mismo de este trabajo que emplea el término subestándar, habría que decir que este término, en un primer momento, es equiparable con el de no estándar o vernáculo. Sin embargo, aquí he utilizado el término subestándar pues para Haensch, lexicógrafo especialista en tipología de diccionarios, este término bien podría utilizarse como un concepto aglutinante para designar a los diccionarios que tratan ciertas variedades lingüísticas de orden diastrático y diafásico, incluyendo al léxico tabuizado (Haensch, 1982b, p. 144; 1997, p. 94).

Ahora bien, antes de terminar con esta sección sobre la variación diastrática, perteneciente al subestándar, habría que hablar de las variedades sociales denominadas jergas. Las jergas son “lenguajes especiales” o variedades lingüísticas que usan en común los hablantes que comparten ciertas profesiones u oficios (RAE, 2001, p. 1318), o simples intereses de grupo (Holmes, 2001, p. 246), como por ejemplo sucede entre los miembros de los grupos sociales marginales. Por ello, las jergas pueden ser vistas como un tipo de variación lingüística diastrática o sociolectal. De hecho, Haensch así las considera (1997, p. 94). Esta clasificación de las jergas resulta aún más entendible si recordamos que los dialectos o sociolectos son variedades lingüísticas que se definen de acuerdo con características propias del usuario (Briz, 1996, p. 15), y más aún de grupos de usuarios (Biber y Finegan, 1994, p. 4; Holmes, 2001, p. 123). Así pues, si bien las jergas no conforman ningún eje dicotómico de categorización, como lo hace la dimensión del estatus social con la variedad estándar frente a la variedad vernácula, no por ello dejan de ser otra

categoría de la variación social o diastrática. Debido a ello aquí he decidido tratar a las jergas como variedades sociales. Sin embargo, hay que comentar también que en este trabajo se buscó identificar, en todas las fuentes posibles, el léxico subestándar del español en México con excepción de lo jergal. Dicha exclusión se llevó a cabo debido a que se consideró que la representación de este tipo de léxico en el corpus electrónico aquí utilizado, el *CREA*, era poco probable dada la composición del mismo (ver sección 3.02).

2.01.04 Variación situacional o diafásica. En cuanto a los ejes de categorización de las variedades lingüísticas propios de la variación situacional, me parece pertinente comenzar aclarando la existencia de dos conceptos cercanos que han sido utilizados para referirse a la unidad de este tipo de variación: el registro y el estilo. Como ya dijimos antes, Ferguson (1994, pp. 16-17) y Briz (1996, p. 15) le llaman registro a la unidad de variación, o variedad lingüística, que está relacionada con la situación en que se habla. Además de estos dos autores hay otros, como Mey (1993/2001, p. 41), que utilizan este término en un sentido similar. Sin embargo, también nos podemos encontrar con autores que hablan del estilo y de la variación estilística para referirse a esta misma variación situacional, como por ejemplo Labov (1983, p. 299); Ryan, Giles y Sebastian (1982, p. 1); Sherzer y Darnell (1978/2000, p. 66); y Milroy y Gordon (1997, p. 198). Igual que sucede con el término registro, además de estos autores ya antes mencionados, existen otros más que se adhieren al uso del término estilo, tal es el caso de Bell (1997, p. 240). Ahora bien, sea cual sea el término utilizado para designar la unidad o sistema representativo de la variación situacional, para los distintos estudiosos de este tipo de variación, el término registro o estilo, según sea el caso, constituye el conjunto de formas lingüísticas que marcan una diferencia en el habla de un mismo individuo cuando diferentes factores situacionales lo influyen para cambiar su forma de hablar. Así pues, en un primer momento se podría

pensar que los términos registro y estilo son un par de términos con substitución mutua, cuya aparición de uno sobre el otro depende de las preferencias de cada autor, sin embargo la coexistencia de estos términos es aún más compleja.

Un ejemplo de esta complejidad lo encontramos en Holmes, quien utiliza el término registro para referirse al tipo de variación social que antes denominamos como jerga. Es decir, para esta autora el registro es la variedad lingüística que un grupo con oficios o intereses comunes comparte, ya sea en las situaciones asociadas con dichos oficios o intereses, ya sea en sus interacciones en general (Holmes, 2001, p. 246). Al mismo tiempo, Holmes utiliza el término estilo para designar la variación lingüística relacionada con factores situacionales (2001, p. 246). Además, esta autora hace notar que la distinción entre los términos registro y estilo no es del todo clara y que muchos sociolingüistas prefieren ignorarla. Ella misma, por ejemplo, al anunciar que explicará cómo los requerimientos funcionales, situacionales y ocupacionales de la comunicación pueden llegar a moldear la variación lingüística, habla indistintamente de estilos y registros para referirse a dichos requerimientos (Holmes, 2001, p. 224). De manera opuesta, otros autores prefieren tomar partido por un término sobre otro. Así lo hacen, por ejemplo, Crystal y Davy (como se cita en Biber, 1991, p. 39). Estos dos autores, según nos dice Biber, rechazan el término registro y prefieren el término estilo, que ven como un concepto aglutinador para referir a las variedades situacionales en general (1991, p. 39). Para aumentar la confusión entre los términos registro y estilo, hay que recordar que existen varios otros términos que se han usado para referir la variación situacional. En este sentido, Biber y Finegan mencionan, en la introducción a su antología de estudios sobre la variación situacional, algunos términos que se han usado de manera indistinta para designar dicha variación. Entre estos términos están el registro, el género, el tipo de texto y el estilo. Según estos autores, todos estos

términos se han utilizado para referir variedades lingüísticas relacionadas con usos situacionales (Biber y Finegan, 1994, p. 4). Dichos términos, además, se oponen al concepto aglutinador *dialecto*, que designa a las variedades lingüísticas relacionadas con grupos enteros de usuarios (Biber y Finegan, 1994, p. 4; Holmes, 2001, p. 123). Ahora bien, ante la necesidad de utilizar aquí un solo término que represente a la variación lingüística situacional, he decidido utilizar el término *estilo* por dos motivos. En primer lugar, he preferido usar el término *estilo* por el hecho de que el término *registro* se ha llegado también a asociar con la variación sociolectal jergal, como se aprecia en Holmes (2001, p. 246). Como la variación social y la situacional son las principales en este estudio, escoger un término que se ha utilizado en estos dos ámbitos para referir tipos de variedades distintas me parece poco recomendable. Segundo, como este trabajo tiene como finalidad explorar la posibilidad de crear un diccionario, si éste algún día se pudiera elaborar, sería mejor, creo yo, utilizar el término *estilo* en la presentación del mismo, pues el significado de este término sería más entendible y transparente para el lector común. Así, una vez seleccionado el término *estilo* como el término a utilizar en este estudio para designar a la unidad de la variación situacional, resulta pertinente hacer un par de comentarios acerca de la caracterización que hacen del *estilo* algunos de los estudiosos del mismo.

De acuerdo con Labov, la variación de *estilo* como fenómeno lingüístico individual es un rasgo obligatorio más que alternativo. Según este autor no existen hablantes que cuenten con un solo *estilo* o tipo de variedad situacional. Para Labov, las diferencias en la variación estilística de los hablantes tienen que ver más bien con la cantidad de *estilos* y de cambios entre *estilos* con que opera cada hablante, y no con la presencia o ausencia de dicha variación. De igual manera, según este autor, el cambio de *estilo* bien puede diferir de una comunidad a otra de acuerdo con los tipos de cambios de *estilos* que se observan en

ellas, así como de acuerdo con los elementos lingüísticos implicados en los distintos estilos que manejan dichas comunidades, mas no de acuerdo con la existencia misma del cambio de estilo (Labov, 1984, p. 29). En cuanto a los elementos lingüísticos que están implicados en los cambios entre distintos estilos, Bell hace notar que la variación de estilo puede tener lugar en distintos niveles lingüísticos, como el nivel fonológico, es decir, el sistema de sonidos de una lengua; el nivel sintáctico, relacionado con el ordenamiento de las palabras; el nivel semántico o léxico, que tiene que ver con los significados y con las palabras mismas; e incluso el nivel discursivo, o sea de patrones más amplios del habla y de la conversación (1997, p. 240). Ahora bien, en cuanto al propósito de la utilización de diferentes rasgos lingüísticos presentes en el habla de un individuo, Ryan, Giles y Sebastian nos dicen que estas diferencias pueden ser utilizadas por el interlocutor o destinatario para hacer una serie de inferencias sobre el hablante. Entre las inferencias que pueden llevarse a cabo se encuentran inferencias acerca de las características personales del hablante, acerca de su pertenencia a cierto grupo social, así como acerca de los estados psicológicos que caracterizan su mensaje (Ryan, Giles y Sebastian, 1982, p. 2). Por tanto, para estos autores, la selección de un estilo por encima de otros posibles y la interpretación de dicha selección dependerán de la situación específica de comunicación y de los valores asociados con el estilo a seleccionar (Ryan, Giles y Sebastián, 1982, p. 2). Para Finegan y Biber, por su parte, la selección de los estilos depende de las características de los tipos de discursos y las situaciones en que se presentan. Estos tipos de discursos, que ellos llaman registros, varían de acuerdo con tres parámetros que determinan el tipo de estilo a emplear para la comunicación en dichas situaciones. Los parámetros son (a) la oportunidad que proporciona la situación o tipo de discurso al hablante para producir un mensaje cuidado, (b) el propósito de comunicación que tienen el hablante y su mensaje, y (c) el grado de

conocimiento compartido que tienen los interlocutores acerca del contexto o situación de comunicación (Finegan y Biber, 1994, p. 325). Así pues, una vez revisadas las implicaciones de la variación lingüística situacional y de su funcionamiento, habría que decir cuáles son los ejes de categorización de las variedades lingüísticas propios de dicha variación.

Como se mencionó al hablar de las dimensiones o ejes de categorización de la variación social, Holmes señala tres dimensiones situacionales de la variación lingüística. La primera de estas dimensiones o ejes es el de la solidaridad frente a la distancia social. Este eje tiene que ver con el grado de conocimiento y confianza que hay entre los interlocutores, así como con el grado de afinidad existente entre sus actitudes y valores (Holmes, 2001, p. 374). Dependiendo de qué tan cerca o lejos de su interlocutor se sienta el emisor o hablante, variará el estilo a utilizar en la comunicación. En segundo lugar, Holmes sitúa el eje de la formalidad frente a lo casual. En uno de los extremos de esta dimensión se halla la formalidad del lugar o de la situación en que se usa la lengua, opuesta a la ausencia de la misma en la cual se da una conversación casual o relajada; dicha ausencia, claro está, se localiza en el otro extremo del eje. De acuerdo con esta dimensión el estilo seleccionado por el hablante depende de la mayor o menor formalidad que caracteriza a la situación misma en que se da la comunicación. Finalmente para Holmes, se encuentra la dimensión representada por la función de la interacción de los hablantes (2001, p. 375). En este sentido, el eje resultante no tiene dos extremos sino que es un continuo en el cual pueden aparecer diferentes niveles que dependen del fin principal de comunicación que persigue el hablante. Así, el estilo variará según si dicho fin es expresar sentimientos, dar a conocer información, lograr que el interlocutor o destinatario haga algo, comentar algo sobre el mensaje mismo, o alguna otra función posible (Holmes, 2001, pp. 258-260). De estos tres

ejes (el de la solidaridad frente a la distancia social, el de la formalidad frente a lo casual y el de la función de la interacción de los hablantes), los dos primeros son los más comúnmente definidos y trabajados en los estudios de la variación estilística.

El eje de categorización de la variedad lingüística situacional más frecuentado en los estudios sobre la materia, es el segundo eje propuesto por Holmes (2001), el de la formalidad frente a lo casual. Respecto de este eje, trabajado en el campo de la sociolingüística por Labov, este autor nos dice que el extremo representado por la formalidad, o por el contexto formal, es aquel en el que el discurso recibe más atención de la mínima que podría ser puesta en él (1984, p. 29). El extremo opuesto, por tanto, está representado por el discurso producido en un contexto en que tan sólo se requiere una atención mínima en la producción del mismo. Siguiendo este mismo eje, Briz (1996) opone el estilo formal frente al estilo que él llama informal-coloquial. La oposición de Briz es básicamente la misma de Labov, si bien Briz añade una acotación respecto de la composición del eje que merece mención. Para Briz el eje no es plenamente dicotómico, con tan sólo dos opciones, sino que se trata más bien de un continuo con una serie de cortes intermedios (1996, p. 16). Otro comentario de importancia respecto del eje de lo formal frente a lo informal es hecho por Irvine (1984). De acuerdo con esta autora, existen entre los lingüistas dos tendencias distintas en el uso del concepto de formalidad en la definición de los estilos. Así, una primera corriente considera que la formalidad se manifiesta en una mayor estructuración y predeterminación del discurso, que se encuentra entonces sujeto a más reglas, más elaboradas también (Irvine, 1984, p. 212). Este uso del concepto de formalidad está muy cercano al uso del mismo dentro del eje presentado por Labov (1984) y por Briz (1996), en el cual lo formal recibe mayor atención. Sin embargo, de acuerdo con Irvine, otra tendencia en el uso del concepto de formalidad en los estudiosos de la variación

del estilo ve en la formalidad una manera de describir ciertas características de la situación en que se habla. Dichas características, para esta autora, están relacionadas más bien con el tono afectivo de la situación de comunicación, tales como el grado de seriedad e intimidad, el respeto y la cortesía manejada por los hablantes (Irvine, 1984, p. 212). En este sentido el uso del concepto de formalidad parece acercarse más a la primera dimensión de Holmes (2001), de la solidaridad frente a la distancia social. Así pues, tomando en cuenta lo dicho por Irvine, se puede notar que la diferenciación entre los distintos ejes propuestos por los estudiosos de la variación situacional es más complicada de lo que podría parecer en un primer momento.

Otro problema para distinguir los ejes de categorización de la variación situacional surge cuando Labov coloca, por encima del eje de lo formal frente a lo informal o casual, el eje de la atención que es puesta en el discurso (1984, p. 29). Este eje es, igualmente, un eje íntimamente relacionado con los dos mencionados anteriormente, el de lo formal frente a lo informal y el eje, de Holmes (2001), de la solidaridad frente a la distancia social. De hecho para Labov, la dimensión representada por el grado de atención prestada al discurso bien podría constituir el eje único de ordenamiento de los distintos estilos (1983, p. 154; 1984, p. 29). Los extremos resultantes de esta dimensión en el eje de categorización de la variación lingüística son el del estilo cuidado frente al estilo espontáneo, es decir, aquellos que reciben más y menos atención, respectivamente (Labov, 1983, p. 115). Esta última terminología, sin embargo, ha recibido algunas críticas. Dichas críticas se han orientado sobre todo al uso que ha hecho Labov del concepto de *discurso cuidado* al hablar de la variación social correspondiente a lo vernáculo o a la variedad no estándar. En este sentido, hay que recordar que para este autor la variación social no estándar genera un discurso menos cuidado frente a uno más sistemático, que sería el estándar (Labov, 1983, p. 265).

Esta caracterización de la variación social no estándar ha sido criticada por Bell, quien llama la atención sobre el hecho de que la variedad social no estándar también puede manifestarse a través de un estilo en el cual se pone atención y cuidado en la forma de hablar (1997, p. 242).

Como un último comentario sobre los ejes de la variación situacional, quisiera traer a colación la advertencia hecha por Briz (1996) respecto de los diferentes nombres que se han utilizado para referir la variación situacional informal, que él llama informal-coloquial (Briz, 1996, p. 16). De acuerdo con este autor, el estilo informal o coloquial ha sido referido también como conversacional, familiar, popular y vulgar (Briz, 1996, p. 25). Respecto del primer término, el de estilo conversacional, Briz se opone a su uso como equivalente de estilo informal-coloquial, pues como bien hace notar, la conversación y la oralidad no coinciden plenamente con la informalidad del discurso. De igual manera, lo escrito tampoco se corresponde totalmente con lo formal. Es decir, puede haber expresiones orales y escritas que sean más o menos formales y más o menos informales (Briz, 1996, p. 23). En cuanto al segundo término utilizado como equivalente del término *estilo informal-coloquial* que este autor menciona, el término *estilo familiar*, Briz rechaza su empleo debido a que dice que alude de manera excesiva a posibles relaciones de parentesco entre los interlocutores, o bien a la simple interacción comunicativa entre personas allegadas (Briz, 1996, p. 25). A mí, sin embargo, me parece que este término resulta útil para aludir la dimensión en que se opone la solidaridad a la distancia social, y que tiene que ver con el grado de conocimiento y confianza que hay entre los interlocutores (Holmes, 2001, p. 374). El tercer término mencionado por Briz es el de estilo popular. Respecto de este término, coincido con Briz cuando llama la atención sobre la equivocación en el tratamiento de la etiqueta *popular* como un estilo, pues dicha etiqueta designa más bien a una variación

social. Lo popular representa, de hecho, a la variedad social propia de lo vernáculo o del no estándar. Esta variación social, que se comentó en la sección anterior (ver sección 2.01.03), está relacionada con las características socioculturales del individuo (Briz, 1996, p. 25). Según hace notar Seco (como se cita en Briz), el estándar no está casado con el estilo formal, y el no estándar o lo popular no lo está con el estilo informal. Bien puede haber manifestaciones lingüísticas que pertenezcan a la variación social estándar y que utilicen una variación situacional informal, al igual que se pueden encontrar manifestaciones lingüísticas que formen parte de la variación social no estándar o popular y que presenten una variación situacional formal (1996, p. 26). Por último, el único término cuyo uso Briz defiende, pero que separa del término de estilo informal-coloquial, es el de estilo vulgar. De acuerdo con este autor, al igual que el término popular, el término vulgar corresponde a una variación lingüística de orden social, que pertenece a un nivel de lengua bajo, al margen de la norma estándar. En este sentido, el término vulgar para Briz coincide con la variedad no estándar, pero a diferencia de ésta, representa usos incorrectos o fuera de la norma. Así, Briz defiende el uso del término vulgar aun en contra de las opiniones de otros lingüistas, como Vigara (como se cita en Briz, 1996), que rechazan el uso del término por considerarlo peyorativo. Además, se opone a fusionarlo ya sea con la variación situacional informal-coloquial, ya sea con la variación social popular o no estándar (Briz, 1996, p. 26). Así pues, como se puede deducir de su concepción del término vulgar, la defensa de Briz acerca del uso de este término, a pesar de la oposición hecha a dicha utilización por parte de otros lingüistas, obedece a que Briz sostiene una postura más bien normativa respecto de la variación lingüística.

Aunada a la ya mencionada oposición del uso del término *vulgar*, que se considera peyorativo, la definición de Briz para este término también encuentra objeciones en otro

respecto. Dichas objeciones están relacionadas con el tratamiento que se da al léxico tabuizado, malsonante o grosero. Para explicar el conflicto que rodea en este sentido el uso de este término en Briz, hay que hacer primero dos aclaraciones. Por un lado, según Haensch, la parte del léxico que corresponde al léxico tabuizado o malsonante no constituye un nivel de lengua o una variedad lingüística en sí misma, sino que más bien forma parte del léxico vulgar (1982b, p. 144) que, como ya hemos visto, Briz (1996) sitúa como una variación social. Por otro lado, como bien hace la RAE (2003), en el CD ROM de su último diccionario, el léxico malsonante está situado en el árbol de uso lingüístico bajo la etiqueta “Valoración del hablante respecto del receptor”, la cual es sin lugar a dudas una marcación situacional y no social. Cabe aclarar que otros tipos de léxico que la RAE incluye en esta misma rama de su árbol del uso lingüístico son el léxico de la amenaza, el del eufemismo (que pretende suavizar las expresiones malsonantes) y el del insulto (2003). Tomando en cuenta estos dos hechos, la consideración de Briz respecto del léxico vulgar como una variación social resulta conflictiva. Esto se debe a que, si se considera el léxico vulgar exclusivamente como una variación social, el léxico tabuizado, que forma parte del léxico vulgar según Haensch, no podría caracterizarse según la percepción que el hablante tiene de su interlocutor, como lo hace la RAE, ya que dicha percepción representa un hecho pragmático y situacional. Así pues, respecto del problema en la localización del léxico vulgar y el léxico tabuizado, resulta prudente la medida tomada por Cheshire en su estudio del habla vernácula e informal de los adolescentes (1997). En dicho trabajo, Cheshire decide tomar las groserías como una categoría de estudio, pero separa esta categoría de las características de la lengua vernácula y la coloca como una categoría aparte (1997, pp. 186-187).

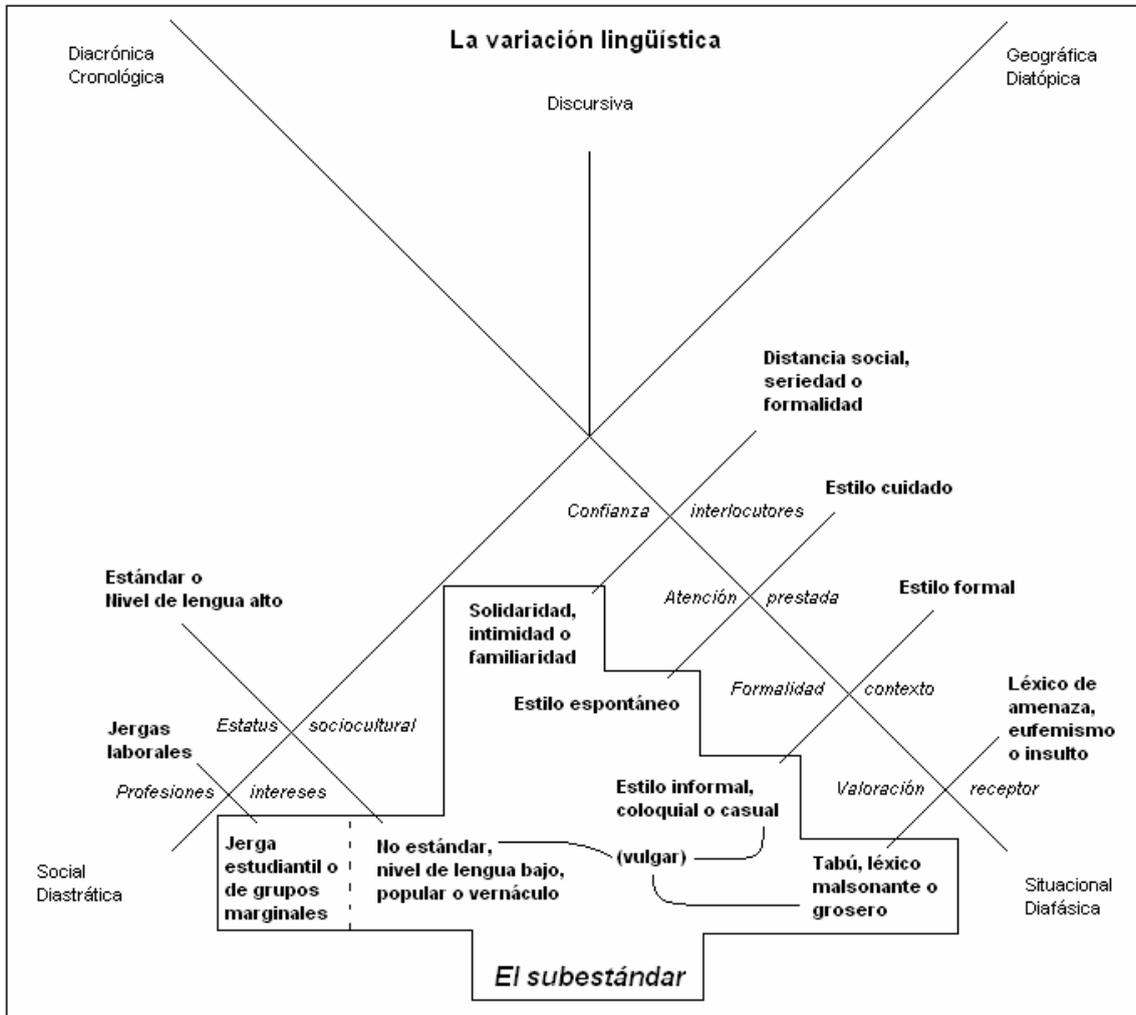
A pesar de los problemas encontrados en el establecimiento de los ejes de categorización de la variación situacional, podemos decir que las variedades lingüísticas situacionales son entidades susceptibles de delimitación. En el caso del estilo informal o coloquial, Briz apunta que en éste se puede encontrar una serie de constantes estructurales o de planificación, retóricas o de recursos expresivos, léxicas o relacionadas con las palabras, y paralingüísticas o del gesto (1996, pp. 34-45). Para ahondar un poco más respecto de estas constantes y al mismo tiempo hablar de la constante que aquí más nos interesa, podemos decir que, según Briz, el léxico informal se distingue por la utilización de un bloque restringido del léxico común, por la existencia de ciertas frecuencias de palabras y por la presencia de lenguajes especiales, principalmente del argot o del léxico jergal (Briz, 1996, pp. 60-62). En este último comentario, resulta interesante ver cómo Briz incluye en el léxico de una variedad lingüística situacional, como lo es el estilo informal, la presencia de un léxico propio de una variedad social, como lo es el de las jergas. Esta concesión de Briz viene en un cierto sentido a corroborar la posibilidad de hacer la inclusión, anteriormente planteada (Haensch, 1982b, p. 144), del léxico tabuizado en esta misma variación situacional. Tal inclusión bien puede tener lugar a pesar de la dificultad de definir de manera exclusiva la pertenencia del léxico tabuizado, ya sea a la variación social del no estándar, ya sea a la variación situacional del estilo informal.

Ya para finalizar con esta sección, creo que es importante hacer notar la importancia intrínseca que tiene la variación situacional respecto de la variación social. De acuerdo con Finegan y Biber, la variación situacional o diafásica subyace a la variación social o diastrática en el sentido de que para que tenga lugar esta última tiene que ocurrir primero la variación situacional. Esta visión se apoya en dos hechos lingüísticos. Por un lado, según estos autores, los patrones de rasgos que motivan la variación en distintas situaciones bien

pueden considerarse motivados por las funciones mismas del acto comunicativo. Por otro lado, estos rasgos situacionales funcionan de manera bastante similar para todos los individuos de una misma comunidad (Finegan y Biber, 1994, p. 315). Tomando en cuenta estos dos hechos, Finegan y Biber realizan un estudio sobre los distintos patrones de la variación lingüística. En dicho estudio, los autores concluyen que los patrones de la variación situacional encuentran una reproducción en la variación social, y que por tanto, los patrones de esta última variación se derivan de los patrones de la primera (Finegan y Biber, 1994, p. 315).

2.01.05 Variación lingüística y subestándar. Ahora, una vez revisados los distintos ejes de la variación social y la variación situacional, hay que aclarar cuáles son los extremos de dichos ejes que corresponden al ámbito del subestándar. Para hacer más fácil el entendimiento de la localización del subestándar, he decidido graficar las diferentes dimensiones de la variación lingüística en general, y colocar los ejes de categorización que se han comentado aquí respecto de la variación social o diastrática y la variación situacional o diafásica. El resultado de esta graficación se muestra en la Figura 2.1, que aparece a continuación. Cabe mencionar que el extremo de la variación social jergal donde se encuentran la jerga estudiantil y las jergas de grupos marginales ha sido separado con una línea punteada del resto del ámbito del subestándar. Esto debido a la exclusión que se hizo de este tipo de léxico del trabajo de documentación con el corpus como resultado de la escasa probabilidad de representación de dicho léxico que se pensó que habría en el *CREA* (ver sección 3.02).

Figura 2.1. Localización del subestándar dentro de la variación lingüística



En esta figura podemos ver que el primer eje de categorización que permea el ámbito del subestándar y que forma parte de la dimensión de la variación social es el que depende del estatus sociocultural del hablante. Este eje tiene en un extremo al estándar o nivel de lengua alto y en el otro extremo al no estándar o nivel de lengua bajo, popular o vernáculo. Este último extremo es el que se halla dentro del ámbito del subestándar. En cuanto a las jergas, el otro elemento de la variación social que es parte del subestándar,

éstas no constituyen un eje dicotómico de categorización. Las jergas, como se dijo antes y como se puede apreciar en la Figura 2.1, son variaciones que comparten grupos humanos cuyos miembros comparten oficios o intereses comunes. Aquí cabe aclarar que por oficio se debería entender la “ocupación habitual” o la “profesión” de dichos miembros (RAE, 2001, p. 1290), mientras que por intereses comunes se debería entender “la inclinación del ánimo hacia un objeto, una persona..., etc.” que comparten quienes conforman el grupo humano en cuestión (RAE, 2001, p. 1611). Las jergas que distinguen a grupos humanos cuyos miembros comparten un oficio, ocupación o profesión no son consideradas como parte del subestándar. Por el contrario, si bien no todas las jergas habladas por grupos humanos con intereses comunes son incluidas en el subestándar, al menos algunas de ellas sí lo son. Estas últimas son comúnmente aquellas jergas que pertenecen a grupos sociales como el de los jóvenes estudiantes, o bien a otros grupos humanos, considerados como marginales, como el de la delincuencia o el de la drogadicción (Haensch, 1997, p. 94). Una vez más aquí es pertinente aclarar que por marginal se debe entender “un grupo: Que vive o actúa, de modo voluntario o forzoso, fuera de las normas sociales comúnmente admitidas” (RAE, 2001, p. 1454). Acerca de las jergas comúnmente incluidas en el subestándar, quiero hacer de nuevo mención del hecho de que el léxico que corresponde a este tipo de jergas no ha sido incluido en la recolección de datos de este trabajo. Esta exclusión se deriva del hecho de que se consideró poco probable la representación de este tipo de léxico en el *CREA*, dada la composición de este corpus (ver sección 3.02).

Respecto de los elementos de la variación situacional o diafásica que resultan propios del subestándar, aquí hay que incluir uno de los dos extremos pertenecientes a tres ejes de categorización. Primero, en la gráfica, he colocado al eje dependiente de la confianza o afinidad que hay entre los interlocutores. En uno de los extremos de este eje

está la distancia social, la seriedad o la formalidad, ésta última es entendida aquí, según uno de sus posibles significados o usos en lingüística, como origen de un discurso estructurado. La solidaridad, la intimidad o la familiaridad se encuentran en el otro extremo de este eje. Este último extremo es el que forma parte del ámbito del subestándar. Como segundo eje he puesto aquel determinado por la atención prestada al discurso durante su producción. Los extremos de este eje están representados por el estilo cuidado y el estilo espontáneo. El segundo de estos dos estilos está interceptado por el subestándar. Un último factor que da origen a un eje dicotómico de categorización dentro de esta dimensión de la variación lingüística es la formalidad del contexto donde tiene lugar la comunicación. El extremo de este eje perteneciente al subestándar es el del estilo informal, coloquial o casual. Su otro extremo es el del estilo formal, en el cual la formalidad es entendida según otra de sus definiciones en la lingüística, como parte del tono afectivo de la situación. Finalmente, en esta misma dimensión de la variación nos encontramos con un punto de categorización que no representa un eje dicotómico, el de la valoración social del receptor por parte del hablante. Aquí, el léxico tabuizado, malsonante o grosero, según se ha visto, se llega a considerar comúnmente como parte del subestándar (Haensch, 1982b, p. 144). Otras categorías derivadas de la valoración social del receptor por el hablante son las del léxico de la amenaza, del eufemismo o del insulto. En este último sentido es importante aclarar que si bien estas últimas categorías no son mencionadas como parte explícita del subestándar por ningún autor en particular, su cercanía con el léxico tabuizado o grosero y los pocos elementos que comúnmente aparecen respecto de ellas en los diccionarios han hecho que decida incluirlas en la recolección de los datos de este trabajo. Esta decisión se ve reflejada en los procedimientos de trabajo comentados en la Metodología (ver sección 3.07). Finalmente, cabe hacer notar que en la Figura 2.1 también se ha incluido la categoría

de léxico vulgar. Esta categoría, como se explicó anteriormente, resulta difícil de ubicar de forma exclusiva dentro de uno de los dos tipos de variación que abarca el subestándar, la diastrática o la diafásica. Por ello la he colocado entre paréntesis y he indicado con líneas los extremos con los que se encuentra conectada en los ejes de categorización. Por un lado, el léxico vulgar aparece relacionado con el estilo informal y el tabú lingüístico, propios de la variación situacional. Por otro lado, el léxico vulgar está conectado con el no estándar, que forma parte de la variación social.

El último comentario que quiero hacer en esta sección está relacionado con la existencia de algunos términos cercanos al del subestándar que se pueden encontrar en los estudios de la variación lingüística. De tales términos, el que quizá resulte más cercano al concepto del subestándar es el término *antilinguaje*. El concepto del antilinguaje fue desarrollado por Halliday (como se cita en Hodge y Kress, 1997) para designar la variación lingüística que utilizan los grupos socialmente subordinados que buscan diferenciarse del grupo dominante y que buscan crear una identidad interna (p. 53). El término antilinguaje, de acuerdo con Hodge y Kress, resulta un complemento de dos términos predecesores que elaboró Labov, los *marcadores* y los *estereotipos* (Hodge y Kress, 1997, p. 52). El término marcador, en primer lugar, se refiere a los elementos lingüísticos que diferencian la forma de hablar de una comunidad respecto de otra; el conjunto de todos los marcadores propios de una comunidad constituyen su *acento* (Labov, 1983, p. 387). El término estereotipo, por su parte, incluye tan solo aquellos marcadores que resultan más notorios dentro y fuera de la comunidad que los utiliza (Labov, 1983, p. 387). Según Hodge y Kress, el concepto de Halliday resulta más amplio debido a que Labov trabajó más bien con elementos propios del sonido, mientras que Halliday se enfoca también en el léxico y en la sintaxis (1997, p. 52). Además el término antilinguaje, como ya se dijo, está referido a la variación de los

grupos socialmente marginales, y lingüísticamente opositores. Así pues, el término antilenguaje está relacionado con el del subestándar ya que igualmente incluye los extremos de la variación social o diastrática que abarca el subestándar. Por un lado, abarca el no estándar o nivel de lengua bajo, popular o vernáculo, y por otro lado, abarca las jergas comúnmente consideradas parte del subestándar (ver sección 2.01.03). Sin embargo, el antilenguaje no alude a los extremos de la variación situacional o diafásica que también forman parte del subestándar. Estos extremos, hay que recordar, son el de la solidaridad, el del estilo espontáneo, el del estilo informal y el del léxico malsonante. Por tal razón, el empleo del término subestándar sigue siendo pertinente y justificado para este estudio. Ahora bien, una vez que se han presentado los tipos de variación lingüística relacionados con el subestándar y que se ha delimitado el terreno de éste dentro de dicha variación, es necesario hablar del trabajo lexicográfico que ha abordado, en general, este ámbito de la variación del español, y en particular, el mismo ámbito pero respecto del español en México.

2.02 Antecedentes de la lexicografía hispana del subestándar

La lexicografía hispana tiene una larga tradición en su pudibundez (Haensch, 1997, pp. 33-35) y, como anota Lara sobre la lexicografía académica, en su extravío (2002, p. 10) para el registro de voces pertenecientes al subestándar. La primera carencia de la lexicografía hispana en este respecto está en el registro del léxico tabuizado o grosero en los diccionarios generales. Desde el mismo *Diccionario de autoridades* (1739), primera obra de la RAE, el registro de este tipo de léxico ha sido controvertido. Mientras este diccionario fue realmente innovador en cuanto a su inclusión de voces de origen jergal, la llamada germanía, ‘autorizada’ por los escritores clásicos, la planta de dicho diccionario y

más tarde el primer estatuto de la RAE se declaran en contra del registro del léxico tabuizado (Alvar, 2002, pp. 264-265). En concordancia con este inicio, durante muchos años los diccionarios generales del español posteriores al *Diccionario de autoridades* continuaron siendo excesivamente púdicos en el tratamiento de este tipo de léxico, elemento fundamental entre los materiales que conforman a los diccionarios del subestándar (Haensch, 1982b, p. 144; 1997, p. 94). Así, hasta hace algunos años, los diccionarios españoles aún excluían todas las voces consideradas groseras, y dejaban fuera una gran cantidad de vocabulario de alta frecuencia, que además resultaba necesario para la explicación de otras tantas unidades léxicas, como locuciones, modismos, eufemismos e interdicciones de diferentes áreas geográficas del mundo hispanohablante (Haensch, 1982a, p. 411; 1997, pp. 33-35). Sólo recién, estas carencias de la lexicografía hispana han comenzado a ser superadas gracias a la influencia de la lingüística moderna (Haensch, 1997, pp. 31-33). Por otro lado, gran parte de la actividad lexicográfica del mundo hispano, sobre todo la de la RAE, le ha dado una gran preponderancia al léxico culto (Alvar, 2002, p. 264), y ha subestimado la importancia de las voces coloquiales y populares del mundo hispánico en general, actuando en el registro de estas voces de forma “errática y azarosa” (Lara, 2002, p. 10). Finalmente, la aparición de los llamados diccionarios de insultos, obras lexicográficas específicas que registran vocablos con connotaciones tales como *despectivo*, *insultante*, *eufemístico*, *irónico*, entre otras, ha sido escasa para el español. Esta escasez contrasta con la disponibilidad de este tipo de obras lexicográficas en otras lenguas. Los diccionarios diafásicos, sobre todo los diccionarios de insultos, han sido más ampliamente desarrollados en lenguas tales como el francés (Haensch, 1982b, p. 150), el alemán y el italiano; incluso los hay multilingües, si bien estos últimos tienen aún muchas carencias (Haensch, 1997, p. 96).

Ahora bien, respecto del trabajo lexicográfico relacionado con el español de América, Haensch considera que los problemas principales de la lexicografía hispanoamericana son siete (1984, pp. 344-348): (a) la desigualdad cuantitativa y cualitativa de los distintos materiales lexicográficos existentes; (b) la mezcla y confusión de distintos criterios, como el diacrónico con el sincrónico y el normativo con el descriptivo; (c) el énfasis excesivo en términos folklóricos y etnológicos, como los indigenismos; (d) la copia indiscriminada de unidades léxicas de un diccionario a otro, incluyendo tanto palabras arcaicas y obsoletas, como palabras sin la marcación correcta en cuanto a sus marcas de uso y sus marcas diatópicas; (e) el registro de peninsularismos, o voces propias del español de España, lo cual es una falla en el caso de los diccionarios de regionalismos; (f) la representación insuficiente del vocabulario de la vida diaria moderna, es decir, la falta de representación, de acuerdo con Haensch, del léxico derivado de las innovaciones tecnológicas o de las prácticas y las disciplinas humanas que no existían en el pasado, al menos no de la misma manera en que se presentan ahora; y (g) la presencia de prejuicios, tales como la evaluación de lo regional como anómalo, la sobrevaloración de lo exótico, la pudibundez en la descripción de lo tabuizado y la falta de definición del regionalismo como diferencial o no. Así, revisando los principales problemas de la lexicografía hispanoamericana según Haensch, podemos ver que la mayoría de ellos afectan el ámbito de la lexicografía del subestándar. De hecho, el último problema mencionado por este autor alude a este tipo de lexicografía directamente. Por todo lo anterior, tanto la lexicografía hispana en general como la hispanoamericana en particular presentan todavía muchos vacíos en el ámbito de los diccionarios del subestándar. Esto se puede observar en algunas de las más importantes listas de obras lexicográficas del español, sobre todo respecto del español de América.

2.03 Algunos trabajos de referencia de obras lexicográficas del español

Revisando las listas que ofrece Haensch en su obra más reciente sobre los diccionarios del español (1997) y teniendo en cuenta que dentro de dichos diccionarios se incluyen, junto al subestándar, voces tabuizadas y jergales, nos encontramos con los siguientes resultados. En Haensch (1997) aparecen referidos, sobre el léxico del subestándar en español, cuatro diccionarios, todos publicados en España, dos glosarios en una revista alemana, un par de vocabularios bilingües y una traducción del alemán. Hay también dos diccionarios de insultos, ambos españoles, y nueve diccionarios de vocabulario sexual, seis impresos en España y el resto en Ecuador, EUA y Argentina. En este mismo rubro se mencionan dos diccionarios *sui generis*, también publicados en España. Se encuentran dieciséis diccionarios de sociolectos o jergales, en publicaciones españolas, alemanas y francesas (con excepción de una publicación argentina y otra mexicana). Junto a estos diccionarios jergales se menciona además una obra “hecha con poco profesionalismo lexicográfico, pero rica en materiales” (pp. 95-99).

Respecto de los resultados de la búsqueda en Haensch (1997), hay que tomar en cuenta que para este trabajo sólo interesan aquellas obras que, además de tener una marcación diastrático-diafásica perteneciente al subestándar, presentan una doble o más bien triple marcación diatópica, en este caso del área geográfica de México. Cabe recordar, además, que este tipo de marcación múltiple es plenamente posible (Haensch, 1982b, p. 142). Dicha marcación (del subestándar, por un lado, y del ámbito geográfico de México, por otro) bien puede ser explícita, en cuyo caso puede estar contenida en el título, en la introducción o en las entradas mismas del diccionario, o bien puede estar implícita, de manera que se tenga que inferir del lugar de publicación o de la extensión del material

léxico utilizado en la elaboración de la obra. Teniendo esto en cuenta, de la lista proporcionada por Haensch (1997), sólo existen en ella dos obras lexicográficas plenamente referidas al léxico subestándar del español en México, un diccionario de vocabulario sexual publicado en EUA en 1978, *El tabú lingüístico en México: El lenguaje erótico de los mexicanos*, de Larry M. Grimes, y un diccionario jergal, la segunda edición (1991) de *Así habla la delincuencia y otros más...*, de Guillermo Colín Sánchez, publicado en México por Porrúa. De ambos materiales, el primero resulta irrelevante para este estudio pues es anterior al periodo de vigencia de veinticinco años que me tracé para la búsqueda e inclusión de fuentes secundarias. Cabe aclarar que dicho periodo de vigencia de los diccionarios fuente es equivalente al periodo de procedencia de los materiales del corpus electrónico aquí trabajado. Más detalles sobre el establecimiento de este periodo de vigencia de los diccionarios fuente son proporcionados en la sección inmediata posterior, titulada Otra búsqueda de la lexicografía del subestándar en México (ver sección, 2.04). En cuanto al segundo ejemplo encontrado, de acuerdo con el título de dicho diccionario, éste parece ser un diccionario puramente jergal, tipo de diccionario que se ha excluido de las fuentes secundarias de este trabajo debido a la baja probabilidad de representación de sus materiales en el *CREA* (ver sección 3.02). Sin embargo, revisando el contenido de este diccionario me pude percatar de que se trata más bien de un diccionario del subestándar, según se ha definido este tipo de diccionario con anterioridad. Así pues, este es el único material referido en la lista de Haensch que se pudo incluir en la construcción del leuario tentativo cuya representatividad en el *CREA* es motivo de exploración en este trabajo.

Continuando con la revisión de materiales bibliográficos, nos encontramos con un listado de “Estudios metalexográficos” (pp. 203-214) y otro de “Diccionarios y obras lexicográficas” (pp. 214-252) ofrecidos por Bajo (2000) respecto de la lexicografía del

español. En cuanto a este último, que la autora subtitula como “Bibliografía exhaustiva ordenada por tipos de diccionarios”, la autora ofrece una sección de diccionarios “De determinados grupos sociales (con marcación diastrática). De determinados niveles, estilos y registros lingüísticos (con marcación diafásica).” En esta sección la autora presenta una extensa lista de 42 diccionarios. Desafortunadamente para el propósito de este estudio, Bajo excluye de esta sección, según explica en una nota a pie de página, todas las obras lexicográficas que además de contener cualquier marcación diastrática o diafásica, pertenezcan a “países distintos de España”. Esto obedece a que, según nos dice Bajo, no desea extenderse tanto siendo que “en los cuadernos bibliográficos sobre el español de América publicados por Arco/Libro se encuentran muchas referencias sobre el lunfardo y sobre las restantes variedades hispanoamericanas del léxico subestándar” (p. 241). Respecto a esta referencia de la autora a los cuadernos bibliográficos del español en América hablaremos más adelante.

Otros dos trabajos más que conviene revisar aquí, pues consisten en catálogos bibliográficos de referencia propiamente hablando, son el de González (2002) y el de Barriga y Parodi (1998). El primero, si bien persigue como finalidad crear un catálogo de los estudios sobre lexicografía y lexicología aparecidos entre los años de 1990 y 2002 en la lengua española (González, 2002, pp. 13-14), incluye un par de secciones que me pareció que se deberían inspeccionar en este estudio. Por un lado, el repertorio de González contiene una sección de estudios lexicográficos y lexicológicos pertenecientes a México. En esta sección aparecen las referencias bibliográficas de 38 estudios, de los cuales dos se refieren al léxico subestándar (González, 2002, pp.132-134). Sin embargo, estas dos referencias tratan únicamente aquella parte del léxico subestándar que este estudio se ha propuesto obviar, el léxico jergal (ver sección 3.02). Los dos artículos se titulan: Forma y

función de la jerga estudiantil en la ciudad de Temuco (Contreras y Barra, 1996) e Ideología y lenguaje: El léxico de los chavos banda (Hernández, 1991). Sin embargo, por el hecho de que rebasan el léxico subestándar que aquí interesa, no se ha intentado incluir estos dos estudios entre las fuentes secundarias de este trabajo. También resulta importante examinar la sección de Léxico de argot en González (2002, pp.146-149). En esta parte de su libro, González da 67 entradas bibliográficas, de las cuales doce, de acuerdo a sus títulos parecen ser diccionarios. De todas las entradas sólo hay un estudio referido al léxico juvenil de las drogas en Chile, y el resto de las referencias pertenecen a España. De entre los diccionarios tan sólo uno ostenta en el título su inclusión del léxico erótico de España e Hispanoamérica, pero se trata de una edición española que no circula en México. Finalmente, en la sección de Fuentes bibliográficas de González (2002, pp. 45-51), se puede encontrar la referencia bibliográfica completa que Bajo menciona vagamente, en una nota a pie de página, respecto de los cuadernos bibliográficos del español de América de la editorial Arco/Libros (2000, p. 241). En este sentido, según se puede corroborar en González (2002), el cuaderno dedicado a México por la editorial Arco/Libros como catálogo de obras lexicográficas es su cuaderno número nueve de la serie *El español de América: Cuadernos bibliográficos* por Barriga, Martín y Parodi (1999). Aunque este cuaderno es difícil de conseguir en México, el repertorio bibliográfico de Barriga y Parodi llamado *La lingüística en México: 1980-1996* y que publica El Colegio de México (1998) no lo es tanto.

Este trabajo de Barriga y Parodi (1998), que aborda las distintas áreas de la lingüística que se han estudiado en México durante el periodo que indica su título, es el último repertorio bibliográfico que mencionaré en esta sección. En el trabajo de Barriga y Parodi (1998), cada área de estudio lingüístico que tratan las autoras comienza con un corto

capítulo introductorio, seguido de una Bibliografía, o listado de referencia, propiamente hablando. En la sección de Lexicología y Lexicografía (pp. 119-163), en la parte correspondiente a la Bibliografía (pp. 130-163), aparece una subsección de Diccionarios, vocabularios y terminología (pp. 140-145). En ésta, nos encontramos con dos obras lexicográficas de vocabularios especializados, una es un vocabulario de la publicidad en México (Lara, 1982) y la otra es un vocabulario trilingüe de desechos nucleares (Pozzi, en prensa). Se encuentra también un glosario jergal, que se dedica a la cultura rural (Soler, 1985). Finalmente nos encontramos con tres tesis de licenciatura, dos de léxico especializado, léxico agrícola en ambas (Gutiérrez, 1986; Mata, 1985), y una sobre el léxico jergal de un grupo social con intereses compartidos, el de los homosexuales (Cortés, 1989). Por la delimitación del presente trabajo, que excluye el léxico jergal de la documentación con el corpus (ver sección 3.02), ninguno de estos trabajos pudo ser incluido como fuente secundaria.

2.04 Otra búsqueda de la lexicografía del subestándar en México

Debido a la pobreza de los resultados hallados respecto de la lexicografía hispana del subestándar en los estudios relacionados con ella, se decidió realizar una búsqueda en bibliotecas, librerías y bases de datos de todos los diccionarios que contuvieran material léxico perteneciente al subestándar del español en México. Esto se hizo para ampliar los materiales a incluir en el lecionario tentativo a cotejar contra el corpus electrónico. Como la búsqueda llevada a cabo arrojaba una serie de resultados de calidad muy diversa, se tomaron los siguientes criterios para la consideración de los materiales encontrados como pertinentes para este estudio. Primero, se buscaron diccionarios de todo tipo que contuvieran material léxico del tipo que ocupa este trabajo, es decir, perteneciente al

subestándar con excepción de lo jergal (ver sección 3.02). Como la búsqueda se llevó a cabo en todo tipo de diccionarios se exploraron tanto diccionarios generales, como diccionarios especializados, entre los que se encontraron diccionarios escolares, regionales y del subestándar. Segundo, se consideraron tan sólo diccionarios publicados y elaborados en los últimos veinticinco años. Este criterio se decidió tomar debido a que en algunas bibliotecas aparecían múltiples materiales, sobre todo diccionarios de regionalismos, del siglo pasado y antepasado. Así pues, tomando en cuenta el periodo de tiempo al que pertenecen los materiales del corpus electrónico con que aquí se trabajó, que es de veinticinco años, se decidió utilizar este mismo periodo de tiempo como límite para la inclusión de fuentes secundarias. Tercero, se buscaron materiales impresos, algunos de los cuales también cuentan con versiones electrónicas. En Internet aparecen varios glosarios y listas de origen y permanencia inciertos que se prefirió ignorar¹. Así se pudieron reunir un total de siete diccionarios que de alguna manera se ocupan del léxico del subestándar del español en México. Estos siete diccionarios constituyen las fuentes secundarias, cuya consideración es requisito metodológico en la construcción de cualquier diccionario (Haensch, 1997, pp. 54-55; 1982a, p. 435; Porto, 2002, p. 84). El concepto de fuentes secundarias se tratará con más detalle en la Metodología (ver sección 3.02).

Ahora bien, de los siete materiales encontrados dos son diccionarios generales, uno es un diccionario escolar, otro es un diccionario de regionalismos y tres son diccionarios del subestándar. Todos estos diccionarios fueron no solamente publicados sino que también elaborados durante el período de veinticinco años que este trabajo se propuso como

¹ Dos páginas de Internet con cierta elaboración son: Fitch, R. (1997/2003). *Jergas de habla hispana* [En red]. Disponible en: <http://www.jergasdehablahispana.org/mexico.htm> Grupo Villa Flakozitas. (1999/2000). *Jergas del Habla Hispana* [En red]. Disponible en: <http://www.flakozitas.com.ar/paises/jergas/index.htm>.

restricción para la inclusión de fuentes secundarias. De todos estos diccionarios, se utilizó para este estudio la última edición disponible.

Aquí resulta pertinente explicar que como diccionario general se define a aquel que abarca o pretende abarcar toda la amplitud del léxico de una lengua, es decir, que no se pone limitaciones en el estudio de la misma (Porto, 2002, p. 59). Los dos diccionarios generales encontrados que contienen léxico marcado que pertenece al subestándar del español en México son el *Diccionario de la lengua española*, comúnmente abreviado *DRAE*, de la RAE (2001), y el *Diccionario del español usual en México (DEUM)*, de Lara (1996). Así pues, tomando en cuenta la definición de diccionario general, recién presentada, los otros cinco diccionarios utilizados aquí son todos particulares o especiales, es decir, se ocupan tan sólo de una parte del vocabulario de la lengua (Porto, 2002, p. 59), y lo único que los diferencia entre sí es la parte del léxico que describen. Como ya se dijo antes, entre los otros cinco diccionarios, hay un diccionario escolar, uno de regionalismos y tres del subestándar.

Para introducir el diccionario escolar aquí incluido, hay que decir que este tipo de diccionario pertenece al ámbito de los diccionarios didácticos. Los diccionarios didácticos tienen como fin primordial ayudar al usuario no sólo a descifrar el significado de las palabras en los enunciados en donde se encuentran, sino a producir dichos enunciados de manera óptima (Haensch, 1997. p. 127). Además de esta característica, los diccionarios escolares ponen un énfasis especial en su adecuación respecto de las necesidades de sus usuarios (Bajo, 2002, p. 171). Es decir, los diccionarios escolares, más allá de los didácticos, deben precisar cuál es el grupo de usuarios concretos a que están dirigidos (Haensch, 1997, p. 129), ya sean estudiantes de nivel primaria, secundaria o de cualquier otro nivel. En este sentido, existen varios aspectos principales que debe cuidar este tipo de

diccionarios. Entre estos aspectos están la selección de las voces a incluir; la tipografía y su disposición; la claridad de las instrucciones; la aclaración de abreviaturas y otros signos; el acompañamiento de índices; el nivel de abstracción, el vocabulario y la coherencia de las definiciones; la cantidad de acepciones y su ordenamiento; la presencia o ausencia de ejemplos, de indicaciones normativas, gramaticales, ortográficas y semánticas; y la inclusión de ilustraciones y cuadros (Bajo, 2002, pp. 171-172). Cabe aclarar que, aunque la definición de los diccionarios escolares, con intereses más bien normativos, nos podría hacer pensar que este tipo de diccionarios no representa una fuente abundante de materiales del subestándar, se decidió incluir un diccionario de este tipo durante la identificación de las fuentes secundarias. El diccionario escolar incluido en este estudio es el *Diccionario inicial del español en México*, de Ávila y Aguilar (2003). En cuanto a la inclusión de este diccionario, resulta también pertinente comentar que la identificación de voces propias del subestándar en el mismo resultó, después de todo, harto fructífera (ver sección 3.07.07), además de que su documentación fue la que mayor representatividad alcanzó en el *CREA* de entre todos los diccionarios incluidos (ver secciones 4.01 y 4.03)

Por otra parte, un diccionario de regionalismos o con marcación diatópica recolecta aquellos vocablos peculiares de un área geográfica particular (Haensch, 1982b, p. 92; 1997, pp. 84-94), y por ello, a este tipo de diccionario se le llama también diferencial o contrastivo (Porto, 2002, p. 62; Haensch, 1997, p. 226). El único diccionario que responde a esta definición y que resulta relevante para este estudio es el *Diccionario breve de mexicanismos (DBM)*, de Gómez (2001).

En cuanto a los diccionarios del subestándar, que ya han sido definidos, éstos combinan materiales léxicos del nivel de lengua popular, de origen jergal (de algunos grupos sociales), así como del tabú lingüístico (groserías) (Haensch, 1982b, p. 144; 1997, p.

94). En este último grupo se hallaron tres diccionarios que, que carecen de rigurosidad lexicográfica en su construcción, ya que sus introducciones no contienen instrucciones sobre el uso del diccionario, además no contienen paradigmas de conjugación, no hay utilización de símbolos o siglas, no hay anexos, ni suplementos, no se establecen criterios de lematización, ni de inclusión de variantes ortográficas, no aparecen indicaciones fonéticas, gramaticales, ni sobre el uso de las unidades léxicas, entre otras faltas formales en su construcción (Haensch, 1997, pp. 237-243). Sin embargo, se decidió incluir estos tres diccionarios en este estudio pues, como dice Haensch de algunos diccionarios equivalentes del español europeo, resultan ricos dada la amplitud de los materiales que los componen (1997, p. 99). Los tres diccionarios del subestándar aquí incluidos son el *Tumbaburro de la picardía mexicana: Diccionario de términos vulgares (DTV)*, de Jiménez (1999), *El chingolés: Primer diccionario del lenguaje popular mexicano (PDLPM)*, de Usandizaga (1994), y el ya mencionado por Haensch, *Así habla la delincuencia y otros más... (AHDOM)*, de Colín (1987/2001).

Cabe señalar que de los siete diccionarios mencionados sólo el *DEUM*, que es un diccionario general, fue construido con base en un corpus electrónico. Además, ninguno de los siete diccionarios está ejemplificado exhaustivamente, en el sentido de que no ofrecen citas o ejemplos de uso (Bajo, 2000, p. 49) en todas las acepciones de sus entradas. Teniendo en cuenta esto, una vez ubicados todos los posibles trabajos precedentes a este estudio que le podrían servir como fuentes secundarias, se debería mencionar también cuáles son los referentes que, aunque escasos, son los más representativos de la lexicografía hispana respecto de dos aspectos básicos que ocupan esta investigación: la elaboración de diccionarios basados en corpus electrónicos y la producción de diccionarios ejemplificados.

2.05 Corpus electrónicos en la lexicografía

Para comenzar esta sección quisiera recordar los conceptos, antes presentados de *corpus* y de *corpus electrónico* en lexicografía (ver sección 1.01). Un corpus está compuesto por los textos, tanto orales como escritos, en que aparecen las palabras que se pretenden incluir en un diccionario. Estos textos representan los materiales lexicográficos en que se basa la elaboración de la obra (Porto, 2002, p. 84). Dichos materiales han sido tradicionalmente almacenados en fichas textuales (Sinclair, 1985, p. 83). Sin embargo, con la aparición de la informática, el corpus se ha construido por medio de una cadena de textos electrónicos enlazados de forma secuencial y etiquetados de manera que puedan ser identificados según una serie de datos. Los corpus resultantes son los llamados corpus electrónicos (Porto, 2002, p. 131). Respecto de los avances de la lexicografía hispana en la elaboración de diccionarios basados en corpus electrónicos, éstos han sido limitados. Haciendo una comparación con el inglés, la existencia de corpus electrónicos en español resulta raquítica. Alvar (2002) menciona tan sólo tres corpus para el español actual: el *Corpus de referencia del español actual (CREA)*, de la RAE, el *Corpus lingüístico del español contemporáneo (CUMBRES)*, de la editorial SGEL, y el del proyecto del *Diccionario del español de México (DEM)* (pp. 129-130). Este último corpus es conocido con el nombre de *Corpus del español mexicano contemporáneo* (Lara, 1996, p. 13). Como se dijo desde la Introducción misma, tan sólo el primero de estos tres corpus está disponible al público en Internet. El único proyecto de elaboración de un diccionario general basado completamente en la utilización de un corpus electrónico continúa siendo el del *DEM*, iniciado en 1975 y que ha dado tres resultados previos: el *Diccionario fundamental del español de México* (1ª ed., 1982 y 2ª ed. 1993), el *Diccionario básico del español de México* (1982), y el *DEUM* (1996) (Haensch, 1997, pp. 226-227).

Así, la labor lexicográfica en este ámbito resulta escasa comparada con la anglosajona que ya en 1987 dio a la luz su primer ejemplo de gran envergadura de un diccionario basado en un corpus electrónico, el *Collins Cobuild English Language Dictionary*, con una segunda edición de 1995 (Hockey, 2000, pp.148-149). De igual manera, este tipo de proyectos de elaboración de diccionarios por medio de la utilización de corpus electrónicos ha sido anterior en la lengua inglesa respecto del español; el *Dictionary of Old English*, inició en Toronto en marzo de 1969 (Hockey, 2000, p. 147). Finalmente, para terminar con la comparación entre la lexicografía hispana y la anglosajona en cuanto a la elaboración de diccionarios basados en corpus electrónicos, es pertinente mencionar que, mientras la disponibilidad de corpus especializados en español continúa siendo prácticamente nula, Sinclair (1985) ya comenta la existencia de un par de corpus especializados, uno de los cuales es el *Jiaotong corpus of scientific texts*. Además, desde el año en que escribe su artículo, Sinclair menciona el surgimiento y vigencia de varios proyectos para la creación de este tipo de corpus (1985, p. 83). A este respecto, habría que decir que la conveniencia del trabajo con corpus especializados en la elaboración de diccionarios especiales resulta obvia pues la representación en un corpus de las palabras y sus diferentes sentidos depende de los tipos de temas y registros representados en los textos que conforman el corpus (Biber, Conrad y Reppen, 1998, pp. 30, 33-34). Ahora bien, aunque este último comentario parece contradecir el valor metodológico del presente trabajo, en este caso se debe tomar en cuenta lo que Sinclair (1985) arguye respecto a las deficiencias en la composición de los corpus. Para este autor, la investigación debe mejorar tanto la selección de textos que componen los corpus, como el procesamiento de la evidencia lingüística proporcionada por éstos y su aplicación a la lexicografía. Sin embargo, de acuerdo con Sinclair, mientras dicha mejora en la selección de textos para la

composición de los corpus no tenga lugar (como sucede en el caso del español) el tratamiento que se dé a la evidencia obtenida por medio de los corpus debe ser cuidadoso, mas no por eso nulo. Según lo que nos dice este autor no se debe ignorar la posibilidad de obtener dicha evidencia, sino que por el contrario debe ser buscada y utilizada (Sinclair, 1985, p. 86).

2.06 Papel de la ejemplificación en la lexicografía

En cuanto a la ejemplificación de diccionarios en la lexicografía hispana, hay que señalar que los ejemplos sacados de un corpus han formado parte de la construcción de los diccionarios del español desde el nacimiento mismo de éstos en el siglo XVII. De hecho, se podría decir que la utilización de ejemplos, también llamados ‘autoridades’, antecede al nacimiento de la lexicografía hispana monolingüe, pues le es más bien una herencia de las producciones lexicográficas italianas del siglo precedente (Lara, 1997, pp. 36-42). Aquí cabe aclarar que los primeros ejemplos utilizados en la lexicografía hispana monolingüe eran llamados autoridades pues eran considerados testimonios que ‘autorizaban’ ciertos usos. Así pues, la lexicografía hispana se inaugura con un diccionario ejemplificado, el *Diccionario de la lengua castellana* (1739), mejor conocido como *Diccionario de autoridades*, del cual se hizo mención anteriormente. En este sentido, se debe mencionar que esta primera obra ejemplificada de la lexicografía hispana se caracterizó por una apertura mayor a la de sus precedentes italianos, que se reflejaba en la amplitud cronológica de sus citas que abarcaban toda la producción escrita del español hasta ese momento; el *Vocabolario degli Accademici della Crusca* que antecedió al *Diccionario de autoridades* en su ejemplificación, la restringía al clasicismo italiano del siglo XIV y XV (Lara, 1997, pp. 36-37, 42). Además, la apertura de la ejemplificación del *Diccionario de autoridades* se

reflejó también en la inclusión de palabras regionales de España y de voces de germanía, o jergales (Lara, 1997, p. 43). Así, la ejemplificación inicial en la lexicografía hispana casi se inscribe en el concepto más moderno del ejemplo propio del siglo XIX, que tiene lugar en los diccionarios filológicos. En estos, el ejemplo tiene el valor de registro histórico, aunque sin perder su matiz normativo (Lara, 1997, p. 256). Uno de los diccionarios de este tipo más conocidos es el *Oxford English Dictionary*. En este respecto, sin embargo, hay que recordar que el español aún no cuenta con un diccionario filológico, también llamado histórico.

Por otro lado, se debe hacer mención de que la fase más reciente en cuanto al papel de la ejemplificación en la lexicografía está representada por “la conversión de la autoridad en simple cita de la fuente de un uso”. Este cambio de perspectiva acerca de la ejemplificación tuvo lugar debido a la influencia de la lingüística moderna del siglo XX. La nueva concepción de la ejemplificación marca sus inicios con la publicación del *Webster's Third New International Dictionary of the English Language* (1966) (Lara, 1997, p. 256). Una de las obras más representativas de esta concepción de la ejemplificación en la lexicografía hispana actual es el *Diccionario del Español Actual* de Seco, Andrés y Ramos (1999), que Bajo refiere como “un magnífico ejemplo de diccionario con citas” (2000, p. 49) y que cuenta con alrededor de 200 mil “ejemplos de uso real.”

2.07 Planteamiento de la pregunta de investigación

Por medio de una exploración de los diccionarios fuentes de este estudio, se puede apreciar que dichos diccionarios contienen inventarios léxicos del subestándar con extensiones varias y con distintos niveles de formalidad en su presentación y en sus contenidos. Por tanto, he partido del presupuesto de que, para conocer el grado de representación de las fuentes secundarias de este estudio en un corpus electrónico, en este

caso en el *CREA*, sería necesario llevar a cabo por lo menos la documentación de una muestra aleatoria estratificada de dichas fuentes. Así pues, con la meta a posteriori de elaborar un diccionario ejemplificado del español en México, la pregunta de investigación del presente trabajo es: ¿hasta qué punto resulta factible llevar a cabo la ejemplificación, por medio del único corpus electrónico del español de acceso público, de un leuario exhaustivo del léxico subestándar del español en México, proveniente de las todas las fuentes secundarias pertinentes elaboradas y aparecidas en los últimos veinticinco años de labor lexicográfica? En este sentido, se espera encontrar cuál es el nivel de representatividad del corpus electrónico aquí utilizado, el *CREA*, respecto de una serie de muestras del léxico del subestándar provenientes de cada uno de los diccionarios que conforman las fuentes secundarias. Determinando el nivel de representatividad de cada uno de los diccionarios, a través de la documentación de una muestra aleatoria de los mismos, se pretende estimar el volumen resultante de la documentación total de las fuentes secundarias y, por ende, la extensión del diccionario que contuviese dicha documentación.

3 Metodología

En esta sección se hablará del concepto de *fuentes* dentro del área de la lexicografía, y se definirán los tipos de fuentes que se pueden emplear en la labor lexicográfica. Al mismo tiempo se definirá en qué consiste el concepto de *fuentes primarias* y de *fuentes secundarias*, y se dirá cuáles son las fuentes primarias y secundarias particulares de este trabajo. Además se explicará cuáles han sido los procedimientos que se han seguido para obtener las fuentes secundarias y el tratamiento que se ha dado a las mismas.

3.01 Fuentes lexicográficas

Los diccionarios se basan en un sistema lingüístico que puede tener distintas procedencias, es decir, el material léxico que compone la fuente de información de un diccionario puede tener varios orígenes. En la práctica lexicográfica, la recolección de dicho material léxico puede darse por dos vías principales: a través de fuentes secundarias, recogiendo el conjunto de materiales útiles de otros diccionarios y estudios lexicográficos, o por medio de fuentes primarias, aprovechando un corpus de materiales originales como textos, grabaciones, encuestas, entre otros (Haensch, 1997, pp. 54-55; 1982a, p. 435). Sea cual sea el tipo de fuentes que se utilice o que se prefiera en la labor lexicográfica, la elaboración de un diccionario debe estar siempre precedida por la recopilación de dichas fuentes (Porto, 2002, p. 84).

3.02 Fuentes secundarias

De acuerdo con la teoría de la práctica lexicográfica, el primer paso a llevar a cabo en la elaboración de un diccionario es la revisión de las fuentes secundarias. En lexicografía

se entiende por fuentes secundarias a todos aquellos materiales que ofrecen vocablos potenciales a incluir en un nuevo diccionario y que contienen a la vez algún tipo de explicación metalingüística respecto de dichos vocablos (Haensch, 1982a, p. 436). Entre este tipo de materiales, resultan de gran utilidad los diccionarios existentes pues éstos cubren una gran cantidad de información (Sinclair, 1985, p. 82). Uno de los mayores méritos de las fuentes secundarias está en el hecho de que la información contenida en las mismas está ya organizada. Es así como una síntesis de buenos materiales, con la actualización y corrección necesarias, podría dar lugar a la elaboración de un diccionario confiable en un periodo de tiempo relativamente corto (Sinclair, 1985, p. 81). Claro está que la utilización de materiales lexicográficos previos debe tomar ciertas precauciones para no caer en el simple plagio y en el estatismo de la práctica lexicográfica.

Por ello, se deben tener en cuenta por lo menos tres desventajas inherentes al uso de las fuentes secundarias. Primero, es difícil conocer cuándo un término pierde vigencia, por lo cual es especialmente peligroso utilizar diccionarios ya existentes en la redacción de diccionarios de un estado de lengua contemporáneo. Segundo, los errores cometidos en la inclusión de vocablos de otros diccionarios son difíciles de detectar. Tercero, existen nuevos avances en la descripción lexicográfica actual, como la información pragmática, que no existían en la concepción teórica de trabajos anteriores (Sinclair, 1985, p. 81). Sinclair afirma que ningún material léxico debería incorporarse en la elaboración de un nuevo diccionario hasta que no se confirme su existencia de manera independiente a su registro en el diccionario fuente. Esta es la única manera, asegura Sinclair, en que la lexicografía se habrá de liberar gradualmente de dos tipos de elementos prescindibles: las voces caídas en desuso, definitivamente innecesarias en los diccionarios contemporáneos, y los materiales léxicos que son meros productos de la lexicografía y que no existen más allá

de ella, al menos en el sentido de que no se encuentra ninguna evidencia textual de los mismos (1985, p. 82). Con todo, los diccionarios existentes no dejan de ser una fuente importante en la elaboración de nuevos diccionarios y sus materiales deberían ser aprovechados. Sin embargo, el aprovechamiento de estos materiales deberá estar condicionado por la averiguación de: la vigencia de uso de las unidades léxicas en consideración, los posibles cambios de registro experimentado por éstas, la pertinencia de la inclusión de las mismas según la finalidad del diccionario en proyecto y la aparición de nuevas acepciones para cada una de ellas (Haensch, 1997, p. 55).

Siguiendo estas recomendaciones acerca de la utilización de las fuentes secundarias, el material lexicográfico existente que se tomó en cuenta en este estudio está conformado por los subconjuntos del léxico del subestándar de los diccionarios impresos en los últimos veinticinco años cuya extensión abarca de alguna manera el español en México. En cuanto a la forma en que estos diccionarios abarcan el español en México, aquí se ha partido únicamente de lo que los mismos diccionarios expresan en sus títulos, instrucciones de uso, lugares de publicación, entre otros tantos elementos que hacen patente la mencionada extensión de sus materiales. Tomando todo esto en consideración, los diccionarios de los que se han extraído los subconjuntos del léxico del subestándar son siete. Estos siete diccionarios, enlistados una vez más con su nombre completo, las siglas utilizadas en este trabajo para referirlos, el tipo de diccionario al que pertenecen, el nombre de su autor y el año de su publicación, aparecen en la Tabla 3.1.

Tabla 3.1. Lista de diccionarios fuente del léxico subestándar del español en México

Nombre del diccionario	Siglas utilizadas	Tipo de diccionario	Autor	Año
(a) <i>Diccionario del español usual en México</i>	(DEUM)	general	Lara	(1996)
(b) <i>Diccionario de la lengua española</i>	(DRAE)	general	RAE	(2001)
(c) <i>Diccionario breve de mexicanismos</i>	(DBM)	regional	Gómez de Silva	(2001/ 2003)
(d) <i>Diccionario inicial del español en México</i>	(DIME)	escolar	Ávila y Aguilar	(2003)
(e) <i>Tumbaburro de la picardía mexicana: Diccionario de términos</i>	(DTV)	subestándar	Jiménez	(1999)
(f) <i>El chingolés: Primer diccionario del lenguaje popular mexicano</i>	(PDLPM)	subestándar	Usandizaga	(1994)
(g) <i>Así habla la delincuencia y otros más...</i>	(AHDOM)	subestándar	Colín Sánchez	(1987/ 2001)

Finalmente, falta hacer la aclaración de que en el presente estudio se decidió excluir en lo posible los elementos léxicos pertenecientes al ámbito jergal comúnmente incluido en el subestándar, en el que se encuentra la jerga de los estudiantes o la de los grupos humanos marginales. Esto se debe, principalmente a dos factores. En primer lugar, el corpus electrónico con que se va a trabajar, el *CREA*, es un corpus más bien de lengua general y estándar que, a pesar de tener como finalidad el llegar a ser “un corpus de referencia... lo suficientemente extenso para representar todas las variedades relevantes de la lengua en

cuestión [el español]” (RAE, 2004, noviembre), bien puede ofrecer una cierta representación del subestándar, parece poco probable que brinde resultados dignos de consideración respecto de la parte de la lengua que se denomina jergal. Además, tomando en cuenta que el *CREA* es el único corpus de dominio público en el español y que no existen todavía en esta lengua corpus electrónicos especializados, por lo menos no en el ámbito del subestándar, podríamos decir que las voces jergales pertenecientes a grupos marginales se encuentran lejos de la posibilidad de ser ejemplificadas con citas a través de este tipo de medio tecnológico. Esto parece ser cierto cuando menos en el caso de la lengua española. En segundo lugar, si se toma en cuenta que este proyecto tiene cortos alcances, ya que sus dimensiones están restringidas en tiempo y en recursos, resulta razonable restringir la variación diastrática o social a la propia del no estándar. Dicha restricción es comprensible si se tiene presente que la variación social del no estándar parece tener una mucho mayor posibilidad de representación en los textos que componen el *CREA* frente a la posibilidad de documentar la variación social jergal. Esta decisión de restringir la variación social se deriva de los criterios de construcción y composición de este corpus electrónico, que busca más bien recabar documentación de la lengua general y estándar (RAE, 2003).

3.03 Fuentes primarias

Ahora bien, si recordamos las precauciones que hacen Haensch (1997; 1982a) y Sinclair (1985) respecto de la utilización de las fuentes secundarias, podemos decir que éstas adquieren su validez en la labor lexicográfica sólo a través de las fuentes primarias. Las fuentes secundarias constituyen una reunión de material importante y útil en la elaboración de diccionarios; sin embargo, “los verdaderos progresos de la lexicografía se deben al aprovechamiento de fuentes primarias, es decir, de textos en sentido más amplio,

donde la unidad léxica que interesa aparece, por lo general, en un contexto” (Haensch, 1982a, p. 437).

Sin embargo, antes de hablar de este tipo de textos, se debe recordar que en las fuentes primarias se puede hacer una distinción de dos tipos: la introspección y la observación del lenguaje en uso. La introspección a su vez se subdivide en el examen de informantes y la introspección del lingüista (Sinclair, 1985, p. 81). El examen de informantes puede ser espontáneo, recogiendo unidades léxicas que aparecen en la lengua hablada, o sistemático, por medio de encuestas (Haensch, 1982a, p. 442). La recolección espontánea e informal continúa siendo controversial en cuanto a su validez, pues se trata de una forma de documentación con respaldos únicos y difíciles de corroborar en otras instancias igualmente espontáneas. Por su parte, la recolección sistemática o formal de materiales a partir de informantes, es decir la encuesta, presenta la limitante de que requiere una gran inversión de tiempo, preparación y recursos económicos. Por ello tiene pocas posibilidades de convertirse en una fuente importante de materiales primarios (Sinclair, 1985, p. 82). En cuanto a la introspección del lexicógrafo, Haensch hace notar su vigencia e importancia, pues aun con los grandes corpus electrónicos que han aparecido y siguen apareciendo en la actualidad, la carencia de representatividad de vocablos y la falta de interpretación que caracteriza a dichos corpus mantienen vivo el papel de este tipo de reflexión en la labor lexicográfica (1982a, p. 443). El redactor del diccionario, y por supuesto el lector del mismo, sigue siendo el único que puede interpretar una forma, un sentido o su caracterización. Con todo, en la elaboración de diccionarios ejemplificados, el mayor defecto de la introspección se deriva de que no proporciona evidencia del uso. Además, la introspección de un informante difícilmente proporcionará una distinción entre todos los posibles patrones lingüísticos de un elemento léxico. El uso de los vocablos no es

un elemento consciente en la competencia del hablante, de hecho es más bien un conocimiento implícito, y el tratar de definirlo partiendo de la introspección del lingüista podría dar lugar a que se registren ideas y no hechos acerca de la lengua. Sobre todo cuando se trata de hacer ejemplificación, el auténtico potencial de la introspección está en la evaluación de la evidencia del material léxico y no en su creación (Sinclair, 1985, p. 82). En este sentido, si bien el comentario de Sinclair está dirigido al campo de la lexicografía, vale la pena mencionar que la introspección del lingüista ha jugado un papel fundamental en otras áreas de la lingüística, como por ejemplo en el análisis sintáctico inscrito en la tradición chomskiana (Kochanski, 2003, Diciembre). Por ello, los comentarios de Sinclair sobre la introspección bien podrían aplicarse a otras áreas de la lingüística.

Por todo lo anterior, el tipo de fuente primaria que debería aparecer en primer lugar, según Sinclair, es la observación del lenguaje en uso, es decir, la utilización de un corpus (1985, p. 82). Respecto del corpus, este mismo autor distingue el uso de fichas de referencia, como el método tradicional, y lo separa del uso de concordancias, o listas de ejemplos de uso, proporcionadas por computadoras, o sea, de la utilización de corpus electrónicos. Además se declara a favor de este último método (Sinclair, 1985, p. 83).

El corpus electrónico utilizado en este trabajo es el *CREA*, que contaba con 140 millones de palabras en el momento en que se comenzaron a hacer las primeras pruebas con él, en octubre de 2003, y que alcanzó los 156 millones de palabras para cuando se terminó la documentación del muestreo aleatorio, en julio de 2004. Este corpus fue elegido porque, además de ser de uso público, se pensó que su composición geográfica podría resultar favorable para la exploración de la representatividad de los distintos subconjuntos del léxico subestándar en las fuentes secundarias trabajadas. Respecto de dicha composición, hay que hacer notar que el 50% del total del corpus está dedicado al español de América.

De esta porción el 40% pertenece al español de la “parte mexicana” del mundo hispano, que en el corpus está conformada por el español de México, del suroeste de los Estados Unidos, de Guatemala, de Honduras y del Salvador. Además, se debe mencionar que el periodo de tiempo a que pertenecen los materiales de este corpus va de la actualidad hasta veinticinco años atrás (RAE, 2003); periodo al que se ha restringido la composición de las fuentes secundarias de este estudio. Como una aclaración metodológica respecto del ámbito geográfico, quiero señalar que en este estudio se utilizaron únicamente aquellos documentos del corpus que pertenecen a México y no los que pertenecen a la “parte mexicana”.

3.04 Diseño

Para determinar la factibilidad de ejemplificación de los materiales léxicos recogidos, se exploró el nivel de representatividad que alcanzaban los subconjuntos del léxico subestándar que conforman las fuentes secundarias. Como el número de elementos léxicos que componen cada una de las fuentes secundarias sumaban un número demasiado alto para ser explorado en su totalidad en este trabajo, se hizo un muestreo aleatorio de cada uno de los diccionarios. Como el muestreo se llevó a cabo de forma separada para los distintos diccionarios que componen las fuentes secundarias, dicho muestreo resultó ser, además, un muestreo estratificado. Una vez hecha la exploración de la representatividad en el corpus, los resultados se han reportado en porcentajes. Esto se ha hecho separadamente para cada una de las muestras aleatorias del léxico subestándar de los distintos diccionarios con la finalidad de estimar el resultado que se obtendría si se explorara la representatividad del total de las fuentes. Así pues, el presente trabajo está basado en un diseño cuantitativo

que se apoya en un método estadístico simple que busca inferir el comportamiento de una población a partir del de una muestra aleatoria estratificada de la misma.

3.05 Expectativas y limitaciones

Este trabajo ha tenido como finalidad primordial especular acerca de las posibilidades de construcción de un diccionario ejemplificado del léxico subestándar del español en México por medio de un corpus electrónico de acceso público. Ante todo, se ha buscado predecir el resultado potencial de una empresa de este estilo. Desde un principio, se ha buscado poder mostrar la pertinencia de elaboración de un diccionario como el ya mencionado por medio de este tipo de recursos, y comparar los posibles resultados de su elaboración con los diccionarios existentes en otras variantes dialectales del español que cuentan con diccionarios similares, como es el caso del español peninsular, así como con diccionarios de otras lenguas como el inglés.

Por otro lado, también se ha pensado en que las posibles diferencias en el nivel de representatividad de los distintos diccionarios incluidos en la muestra podrían sugerir la exploración, en un futuro estudio, de algún tipo de influencia entre los procedimientos de presentación formal de los diccionarios y la representatividad del material léxico de los mismos en un corpus electrónico. Más allá, se podría intentar establecer algún tipo de correlación entre algún criterio individual de evaluación formal y de contenido de los diccionarios (Haensch, 1997, pp. 239-242) y el nivel de representatividad de sus materiales léxicos en un corpus electrónico. Es decir, podríamos comenzar a especular acerca de qué tipo de características formales de un diccionario están en correlación con una mayor posibilidad de documentación de sus elementos léxicos por medios informáticos.

Con todo, hay que reconocer, como ya se hizo anteriormente, que el presente trabajo se enfrenta con fuertes limitaciones, encontrándose la principal de ellas en el hecho de que el subconjunto del léxico que se pretende explorar en el *CREA*, el léxico del subestándar, no es un elemento que defina a este corpus en su composición. En este sentido, sin embargo, habría que recordar que el *CREA* tiene como finalidad llegar a ser “un corpus de referencia ha de ser lo suficientemente extenso para representar todas las variedades relevantes de la lengua en cuestión [el español]” (RAE, 2004, noviembre). Así pues, hemos decidido utilizar al *CREA*, puesto que hasta el día de hoy continúa siendo el único corpus de dominio público que existe para el español (Porto, 2002, p. 129) y continúa siendo utilizado de manera primordial para el trabajo de actualización y realización de los diccionarios de la RAE (2004, noviembre). Además, nos parece pertinente seguir la ya referida recomendación de Sinclair acerca de la utilización de los medios disponibles en lexicografía, aun cuando éstos no aparezcan como ideales (1985, p. 86).

3.06 Procedimientos

Una vez seleccionados los siete diccionarios de los cuales se podrían extraer las posibles fuentes secundarias del léxico subestándar del español en México, se procedió a hacer dicha extracción. Ésta varió según el tipo de diccionario. Así en los diccionarios generales (el *DEUM* y el *DRAE*), en el escolar (el *DIME*) y en el de regionalismos (el *DBM*), se buscó únicamente el material perteneciente al subestándar, con excepción de los materiales léxicos propios de las jergas de grupos humanos marginales. Esto último, como ya se dijo, debido a que de acuerdo con la composición del *CREA* parecía ínfima la posibilidad de identificar léxico jergal en el mismo. De manera contraria, en los diccionarios del subestándar (el *DTV*, el *PDLPM* y el *AHDOM*) se intentó identificar todo

el léxico que no perteneciera al subestándar, además de que se buscó eliminar aquello que se considerara propio del léxico jergal.

Para identificar los lemas que respondiesen a los criterios de búsqueda del subestándar aquí planteados, se analizaron las informaciones metalingüísticas que acompañaban al lema, al encabezado de la definición o a la definición misma. Estas informaciones metalingüísticas pueden aparecer de diversas formas, ya sea como abreviaturas, comúnmente antes de la definición, o como datos desarrollados dentro de ella (Bajo, 2000, pp. 22-23). Este tipo de informaciones metalingüísticas se conocen comúnmente como *marcas*. En la definición de Porto, las marcas son “expresiones, utilizadas a lo largo de toda la obra, constituidas casi siempre por frases estereotipadas, abreviaturas, signos especiales o ciertos recursos gráficos (por ejemplo, un estilo o tamaño de letra especial), cuya misión es ‘marcar’ o destacar una palabra o acepción frente a otras que, por no presentar ninguna característica especial o, por el contrario, la que se considera normal o general, aparecen en el diccionario como elementos ‘no marcados’” (2002, pp. 250-251). Así pues, para la identificación de las marcas que permitiesen extraer las unidades léxicas pertinentes a este estudio, o que permitieran la eliminación de las unidades léxicas no pertinentes, se tuvo que dar un tratamiento distinto a cada diccionario. De hecho, la identificación de marcas, propiamente hablando, se llevó a cabo únicamente en el *DEUM*, en el *DRAE*, en el *DBM*, en el *DTV* y en el *DIME*, pues sólo estos diccionarios contenían este tipo de información. Hay que hacer destacar que el trabajo de identificación y extracción de los materiales léxicos de las fuentes secundarias no lo realicé yo solo, sino que lo llevé a cabo, en dos etapas distintas, con la ayuda de un grupo de becarios¹.

¹ Estos becarios son estudiantes que, debido a que reciben descuentos en las colegiaturas de la escuela en que son alumnos, colaboran con los profesores de la institución en proyectos académicos.

En la primera etapa de dicho trabajo recibí apoyo de un grupo de ocho becarios, que trabajaron una hora diaria, cinco días a la semana, durante doce semanas. Entre estos primeros colaboradores había cuatro estudiantes de preparatoria (Anabel Barreda, Isabel Palafox, Paola Osorio e Iyari Zerón) y cuatro estudiantes de licenciatura (Edgar Esquivel, Daniel Márques, Claudia Montesinos y Gabriel de Santos). La primera etapa consistió en la identificación y creación de listas de papel y listas electrónicas de los seis primeros diccionarios que se describen en el siguiente subcapítulo (ver sección 3.07). En cuanto a los criterios de identificación, éstos fueron determinados por mí de acuerdo con la delimitación del subestándar antes presentada (ver sección 2.01.05). Debido a ello, el trabajo de los becarios consistió únicamente en buscar, subrayar y copiar los lemas con las marcas que les fueron indicadas; encomiendas que, si bien implicaban una gran cantidad de trabajo, no requirieron una extensa capacitación. Además, todas estas actividades fueron realizadas de manera múltiple e independiente, y sus resultados fueron cotejados para asegurar la validez del trabajo. Al mismo tiempo, este primer equipo de trabajo contribuyó a la inclusión del significado de los lemas y acepciones en las listas electrónicas finales del *DEUM*, del *DRAE* y del *DBM*. En cuanto a la inclusión del significado de estos diccionarios en sus respectivas listas, ésta fue la más fácil pues se usaron versiones electrónicas de dichos diccionarios y por tanto el proceso de inclusión consistió en un simple proceso de copiado y pegado de información. En la segunda etapa del trabajo de identificación y creación de las listas de las fuentes secundarias, tuve a mí cargo a un grupo de diez becarios, cuyos nombres son los siguientes: Luis Ángel González, María de Lourdes Cessa, Daniela Sánchez, Edgar Esquivel, Raúl Israel Fernández, Juan Carlos González, Claudia Montesinos, Gabriel de Santos, Ricardo Martínez y Yolanda Ruiz. De estos diez becarios, sólo el primero es estudiante de preparatoria y el resto son estudiantes de licenciatura. Este

segundo grupo de becarios trabajó de forma variada, pues algunos trabajaron tan sólo tres semanas en el proyecto y otros trabajaron hasta diez semanas, a razón de una hora diaria de lunes a viernes. Durante esta segunda etapa se identificaron los lemas pertinentes del *DIME* y se crearon las listas de dicho diccionario, que es el último de los presentados a continuación. Además, se capturaron manualmente los significados en las listas de todos los diccionarios de los que no se copiaron los significados durante la primera etapa.

3.07 Identificación del léxico del subestándar en los diccionarios fuente

En esta sección se presentarán los procedimientos utilizados para identificar la información a extraer de los diccionarios fuente. Cabe aclarar que, debido a las características particulares de estos diccionarios, los procedimientos de la mencionada identificación variaron ampliamente para cada uno de ellos. Así pues, los procedimientos de identificación se presentan en diferentes subsecciones tituladas con las siglas del diccionario al cual se refieren. En cada subsección, además, se menciona la creación de una serie de listas con la información identificada. Estas listas tuvieron como finalidad la construcción de una lista final, para cada diccionario, con la información pertinente al subestándar. Aquí las listas finales serán mencionadas sólo de paso, pues el tratamiento que se les dio a las mismas es el tema de la siguiente sección (ver sección 3.8).

3.07.01 DEUM. De este diccionario se localizaron tres marcas definitorias del tipo de léxico que se buscaba, es decir, del subestándar con exclusión de lo jergal: la marca *Popular*, la marca *Ofensivo* y la marca *Groser* (grosería). Para identificar las entradas del diccionario que incluyeran cualquiera de estas marcas, se comenzó trabajando en la versión impresa del diccionario. Se procedió a revisar uno por uno cada uno de los lemas y los números que indican las distintas acepciones de cada entrada del diccionario, los cuales

aparecen en negritas en el mismo. Durante este proceso, se resaltaron los lemas marcados y las marcas pertenecientes al subestándar. Fue de utilidad la coherencia de este diccionario en el hecho de que tanto el lema como el número que distingue a las entradas están siempre en negritas, además de que las marcas se encuentran siempre después de estos dos elementos. Resaltándolas con marcadores directamente sobre el texto del diccionario, se identificaron los materiales pertinentes para este estudio de entre las 14 mil entradas y 60 mil acepciones de este diccionario (Lara, 1996, p. 14). Después se hicieron independientemente dos listas a mano, que más tarde se cotejaron entre sí. Una vez hecho el cotejo de las dos listas, los datos resultantes del cotejo se capturaron en una hoja de cálculo de Excel. Al mismo tiempo, se utilizó la versión digital de este diccionario de la Biblioteca virtual Miguel de Cervantes, que se encuentra en Internet (Lara, 2000). No se utilizó la versión en Internet del Centro de Estudios Lingüísticos y Literarios del Colegio de México (Lara, 1996), pues en el momento en que se trabajó con la versión de la Biblioteca virtual Miguel de Cervantes, no conocíamos la existencia de esta versión con opciones de búsqueda avanzada. Así pues, utilizando la versión de la Biblioteca virtual Miguel de Cervantes, se revisó dos veces más el lemario completo del diccionario aplicando el buscador de Internet Explorer. Se crearon dos listas más, que se cotejaron entre sí y que después se compararon con la hoja de cálculo creada a partir de las listas del diccionario impreso. El resultado de la identificación de marcas en el *DEUM* dio un total de 633 marcas en 568 lemas. La Tabla 3.2 presenta los resultados, por letra, de la identificación de marcas en este diccionario. En total se encontraron 482 marcas *Popular*, 67 marcas *Ofensivo* y 84 marcas *Groser*.

Tabla 3.2. Conteo por letra de los lemas con marcas del subestándar en el *DEUM*

Letra	Popular	Ofensivo	Grosero	Subtotal
A	94	2	2	98
B	29	-	2	31
C	68	3	16	87
CH	32	13	15	60
D	17	1	-	18
E	16	1	-	17
F	27	9	8	44
G	13	6	-	19
H	13	3	1	17
I	1	4	-	5
J	29	12	10	51
L	13	-	2	15
LL	4	-	-	4
M	16	-	14	30
N	7	2	-	9
Ñ	1	-	-	1
O	7	-	-	7
P	24	3	11	38
Q	3	-	-	3
R	20	4	3	27
S	14	-	-	14
T	18	3	-	21
U	2	-	-	2
V	4	-	-	4
Y	8	-	-	8
Z	2	1	-	3
Subtotal	482	67	84	Total

633

Marcas dobles	33
Marcas triples	16
Total de Lemas	568

3.07.02 DRAE. El procedimiento para este diccionario fue más simple, pues se utilizó desde un principio la versión en CD ROM (RAE, 2003). Así, recurriendo al Árbol de ámbito geográfico, en la subcarpeta de América, en la subcarpeta de Áreas geográficas, en la subcarpeta de México, se tuvieron que revisar de todo el diccionario tan sólo 2,905 marcas *Méx* en 2,444 entradas. De esta manera no se tuvo que hacer la búsqueda sobre las 88,431 entradas y 190,581 acepciones del diccionario impreso. Por ser éste un diccionario que incluye lemas de todo el mundo hispano, se buscó solamente la combinación de la marca geográfica del español en México, con alguna de las marcas que podrían reflejar unidades léxicas propias del subestándar, excluyendo las marcas propias de lo jergal. Así se decidió hacer la extracción de aquellos lemas que combinaran la marca *Méx* (México) con alguna de las siguientes: coloq. (coloquial), malson. (malsonante), vulg. (vulgar), despect. (despectivo), eufem. (eufemismo), e irón. (irónico). Aquí se debe recordar que, según se comentó en la subsección de La variación lingüística y el subestándar, de la sección Componentes teóricos del léxico del subestándar, de la Revisión bibliográfica, en este estudio se decidió incluir algunos elementos especiales de la variación social que depende de la valoración social del receptor hecha por el hablante (ver sección 2.01.05). Estos elementos, incluidos en este diccionario y en el siguiente que se presenta, no son comúnmente considerados parte del subestándar, pero se encuentran íntimamente relacionados con el léxico tabuizado, el cual sí es considerado por algunos teóricos como un elemento del subestándar. Las marcas que responden a esta inclusión en este diccionario son las marcas: despect. (despectivo), eufem. (eufemismo), e irón. (irónico). Para cuidar la calidad de la extracción de los lemas pertinentes, se llevaron a cabo de forma independiente un par de revisiones de la subcarpeta *México* del diccionario electrónico, y se generaron dos listas manuales que más tarde se cotejaron entre sí. La lista con los datos resultantes del

cotejo se comparó con la información contenida en el diccionario impreso y, finalmente, se capturó en una hoja de cálculo de Excel. Los resultados por letra de la identificación de marcas en el *DRAE* han sido presentados en la Tabla 3.3. Como se puede ver en esta tabla, para este diccionario se encontró un total de 215 combinaciones de marcas, en un total de 202 entradas. En dichas combinaciones, cuando menos una marca era una marca diatópica correspondiente al ámbito geográfico de México y otra marca era una marca diastrática-diafásica propia del subestándar. Así, se identificaron 155 combinaciones de la marca *Méx* con la marca *coloq.*, veinticinco combinaciones con la marca *malson.*, veinte combinaciones con la marca *vulg.*, diez combinaciones con la marca *despect.*, cuatro combinaciones con la marca *eufem.*, y una combinación con la marca *irón.*

Tabla 3.3. Conteo por letra de los lemas con marcas del subestándar en el *DRAE*

Letra	coloquial	malsonante	vulgar	despectivo	eufemismo	irónico	Subtotal
A	6	2	-	-	-	-	8
B	13	-	-	-	-	-	13
C	27	2	2	1	-	-	32
CH	17	13	2	2	1	-	35
D	2	-	-	-	-	-	2
E	7	-	-	1	-	1	9
F	8	-	10	-	-	-	18
G	6	-	1	2	-	-	9
H	6	-	-	-	1	-	7
I	1	-	-	-	-	-	1
J	6	-	1	2	-	-	9
L	1	-	-	-	-	-	1
LL	-	-	-	-	-	-	-
M	14	7	1	-	-	-	22
N	-	-	1	-	-	-	1
Ñ	1	-	-	-	-	-	1
O	2	-	-	-	-	-	2
P	16	1	-	1	-	-	18
Q	-	-	-	-	-	-	-
R	5	-	-	1	-	-	6
S	5	-	-	-	-	-	5
T	6	-	2	-	1	-	9
U	-	-	-	-	1	-	1
V	5	-	-	-	-	-	5
Y	-	-	-	-	-	-	-
Z	1	-	-	-	-	-	1
Subtotal	155	25	20	10	4	1	Total
							215

Lemas con marcas dobles	13
Total de Lemas	202

3.07.03 DBM. Las marcas buscadas en este diccionario fueron muy similares a las buscadas en el *DRAE*. Esto resulta entendible si se toma en cuenta que este diccionario tiene como autor a Gómez de Silva (2001/2003), uno de los miembros de la Academia Mexicana, suscrita a la RAE. De este diccionario se buscaron las siguientes marcas: *despect.* (despectivo), *irón.* (irónico), *malsonante*, *vulgar*, *eufemismo*, e *insulto*. Al igual que se comentó respecto del *DRAE*, las marcas *despect.*, *irón.*, *eufemismo* e *insulto* no corresponden a la variación lingüística comúnmente considerada como parte del subestándar; sin embargo, su cercanía con el léxico tabuizado, así como sus escasas instancias en los diccionarios fuente, hizo que se incluyeran en este estudio. Cabe aclarar que algunas de estas marcas eran abreviaturas y otras eran datos desarrollados dentro de la definición. Además, ya que este diccionario registra exclusivamente voces del ámbito geográfico de México, no fue necesario buscar combinaciones de marcas como se hizo en el *DRAE*. Respecto del proceso de identificación de lemas marcados, desde un principio se comenzó a trabajar con la versión de Internet de este diccionario (Gómez de Silva, 2001/2003). Se utilizó el buscador de Explorer para identificar los lemas con marcas pertinentes y crear independientemente dos listas con los mismos. Las dos listas se cotejaron, se compararon con la información del diccionario impreso y se capturó la información revisada en una hoja de cálculo de Excel. Los totales por letra de esta búsqueda se presentan en la Tabla 3.4. Según se aprecia en la tabla, en el DBM se identificaron un total de 142 marcas en 141 entradas. Se encontraron quince marcas *despect.*, dieciséis marcas *irón.*, 83 marcas *malsonante*, dos marcas *vulgar*, veintitrés marcas *eufemismo* y tres marcas *insulto*.

Tabla 3.4. Conteo por letra de los lemas con marcas del subestándar en el *DBM*

Letra	despectivo	irónico	malsonante	vulgar	eufemismo	insulto	Subtotal
A	1	1	-	-	3	1	6
B	-	1	-	-	-	-	1
C	1	2	9	1	3	-	16
CH	1	-	13	-	7	-	21
D	-	-	2	-	-	1	3
E	1	4	1	-	-	-	6
F	-	-	7	-	1	-	8
G	4	-	-	-	-	-	4
H	-	-	6	-	-	-	6
I	-	1	-	-	-	1	2
J	-	-	4	-	1	-	5
L	-	2	2	-	-	-	4
LL	-	-	-	-	-	-	-
M	2	-	24	1	2	-	29
N	-	-	-	-	-	-	-
Ñ	-	-	-	-	-	-	-
O	-	1	1	-	-	-	2
P	-	1	14	-	1	-	16
Q	-	-	-	-	-	-	-
R	2	1	-	-	1	-	4
S	1	1	-	-	-	-	2
T	1	-	-	-	3	-	4
U	-	-	-	-	1	-	1
V	-	1	-	-	-	-	1
Y	-	-	-	-	-	-	-
Z	1	-	-	-	-	-	1
Subtotal	15	16	83	2	23	3	Total
							142

Lemas con marcas dobles

1

Total de Lemas

141

3.07.04 DTV. Para este diccionario y los dos siguientes los procedimientos fueron inversos a los anteriores, pues más bien se buscaba eliminar las marcas jergales, las cuales se decidió excluir de la prueba de representatividad en el corpus por las razones ya antes comentadas (ver sección 3.02). Así, se comenzó contando las entradas. Después, se encontraron seis marcas a eliminar desarrolladas en las definiciones de los lemas de la siguiente manera: (1) jerga de la prostitución, (2) caló del hampa, (3) jerga carcelaria, (4) caló delincuente, (5) caló periodístico y (6) dominó a la mexicana. Las cuatro primeras, constituyen marcas jergales de grupos humanos marginales, mientras que la quinta es una marca jergal en un sentido distinto. Esta última marca puede considerarse una marca jergal laboral perteneciente a un grupo cuyos miembros comparten una profesión u oficio. La sexta marca resultó desconocida, y no se pudo encontrar su significado con ningún informante, por lo que se decidió también eliminarla. El total de entradas con marcas a eliminar fueron: una entrada con marca de *jerga de la prostitución*, 343 entradas con marca de *caló del hampa*, una entrada con marca de *jerga carcelaria*, una entrada con marca de *caló delincuente*, dos entradas con marca de *caló periodístico* y dos entradas con marca de *dominó a la mexicana*. La Tabla 3.5 muestra los totales por letra de entradas en el diccionario (2,307) y de entradas marcadas a eliminar (350), así como los totales de entradas a utilizar. Respecto de este último dato, para este estudio se emplearon 1,957 unidades léxicas o lemas de este diccionario.

Tabla 3.5. Conteo por letra de los lemas a utilizar del DTV

Letra	Entradas		
	totales	eliminadas	utilizadas
A	180	18	162
B	133	28	105
C	284	58	226
CH	136	20	116
D	121	18	103
E	146	13	133
F	64	10	54
G	82	10	72
H	63	9	54
I	25	1	24
J	50	4	46
K	4	1	3
L	71	8	63
LL	9	-	9
M	210	28	182
N	48	4	44
Ñ	7	1	6
O	31	4	27
P	247	38	209
Q	20	1	19
R	78	8	70
S	88	17	71
T	139	27	112
U	3	-	3
V	35	7	28
W	4	-	4
X	9	-	9
Y	6	1	5
Z	14	2	12
Total	2307	336	1971

3.07.05 PDLPM. Este diccionario presentó particularidades que ningún otro de los diccionarios fuente mostró. En la redacción de este diccionario no se lematizó, propiamente hablando, es decir, no se hizo la abstracción de la forma básica de la unidad léxica, que deja

de lado los cambios gramaticales o morfológicos de las misma (Biber, Conrad y Reppen, 1998, p. 29). Ciertamente en este diccionario aparecen lemas o encabezamientos para cada una de las entradas, pero estos son más bien frases en contexto. Así, para el que debería haber sido el lema *chingón, na*, aparecen un total de 286 frases en contexto, a las cuales se les dio una entrada, y cuyas definiciones son más bien enciclopédicas, es decir, definen la realidad no la palabra. Tomando en cuenta esta situación, de las 1,667 entradas que se contaron en este diccionario, se pudieron extraer tan sólo 34 formas lematizables, todas derivadas de la familia lexicogenésica del verbo *chingar*. Esto se debe a que además del caso de *chingón*, la forma *chingada*, incluida en dos locuciones adjetivas, *de la chingada* y *como la chingada*, aparece lematizada con distintos contextos 702 veces en todo el diccionario; la forma *chingadera* aparece lematizada en contexto 180 veces, el verbo *chingar* conjugado y en contexto aparece en 131 lemas o encabezados, el adjetivo *chingado, da* tiene 310 apariciones contextualizadas y lematizadas. La Tabla 3.6 muestra los resultados del conteo de entradas y el conteo de formas lematizables que se llevó a cabo en este diccionario. Como ya se dijo, sólo 34 formas lematizables se identificaron, con las cuales se creó una hoja de cálculo en Excel para proceder posteriormente a su exploración en el corpus. Hay que hacer notar que las formas lematizables son más bien reconstrucciones más de los lemas potenciales que debería haber tenido este diccionario si se hubiesen seguido criterios lexicográficos en su elaboración.

Tabla 3.6. Conteo de entradas y formas lematizables en el *PDLPM*

Formas contadas	Totales
chingón (a)(es)(as)	286
de la chingada	402
como la chingada	300
chingaderas	180
chingar (conjugación)	131
chingo	14
chingado	138
chinga	10
chingonamente	8
chingada	172
hasta la chingada	26
chingadazo	12
historias y comentarios	24
otras formas	28
Total	1731
Formas múltiples que aparecen en una sola entrada	64
Total de entradas	1667
Formas lematizables de la familia lexicogenésica de chingar	34

3.07.06 AHDOM. Este diccionario realiza la lematización independiente de colocaciones, es decir de elementos léxicos que se combinan sin que exista entre ellos la exigencia de aparecer siempre juntos (Porto, 2002, p. 154). Por lo tanto, hay que considerar que muchas de las entradas que aparecen en su inventario léxico son colocaciones que deberían aparecer agrupadas en otras entradas del diccionario. Un ejemplo de lematización más convencional donde las colocaciones no son lematizadas de forma independiente lo

podemos encontrar en la primera definición de la entrada *allanar* de (Alvar, 1987). En esta entrada, después de la definición “Poner llana [una cosa]” es introducida con dos puntos y destacada en letra cursiva una serie de posibles colocaciones del verbo *allanar*, el cual es sustituido con una tilde nasal (~): “~ *unas cercas; ~ una piedra; ~ un monte; el terreno allana o se allana*” (p. 62). Por el contrario en el *AHDOM*, todas las posibles colocaciones de la entrada *aguantar* en su acepción de “ser paciente, tolerante” han sido lematizadas independientemente de ésta, de forma que de esta misma acepción nos encontramos con la serie de entradas siguiente: “**AGUANTAR**”, “**AGUANTAR HORRORES**”, “**AGUANTAR LA ‘BARA’**”, “**AGUANTAR LA RIATA**” y “**AGUANTAR LA VERGA**”. Además de todas estas entradas para la acepción mencionada de *aguantar*, aparecen otras tantas entradas para otras posibles acepciones del mismo lema: “**AGUANTA**”, “**AGUANTA UN PIANO**”, “**AGUANTADOR (SER)**”, “**AGUÁNTAME TANTITO**”, “**AGUANTE (QUÉ!)**” y “**AGUANTE (TENER MUCHO)**” (Colín Sánchez, 1987/2001, p. 24). Debido a este procedimiento de lematización excesivo, el conteo del inventario léxico del *AHDOM* aparenta ser muy prolijo. El conteo del total de entradas en este diccionario, que se tuvo que hacer manualmente, asciende a 9,613 entradas. El conteo por letra se presenta en la Tabla 3.7. No se encontraron marcas de ningún tipo que contribuyeran a la reducción de este inventario léxico, así que simplemente se creó una hoja de cálculo con el total de las entradas del diccionario que se incluyeron en la exploración del *CREA*.

Tabla 3.7. Conteo por letra del lemario del *AHDOM*

Letra	Entradas	Letra	Entradas
A	727	N	302
B	409	Ñ	17
C	1069	O	123
CH	454	P	1143
D	456	Q	106
E	546	R	356
F	233	S	507
G	321	T	647
H	252	U	63
I	96	V	209
J	156	W	4
K	7	X	12
L	335	Y	101
LL	42	Z	57
M	862		
Total	9612		

3.07.07 DIME. Aunque al listar los distintos diccionarios fuente se colocó a este diccionario después de los dos generales incluidos en este estudio, el *DEUM* y el *DRAE*, la extracción de los lemas del *DIME* se comenta hasta este momento debido a que éste fue el último diccionario hallado e integrado en este trabajo. En cuanto a la labor de extracción misma, en este diccionario, dedicado a las “niñas, niños y jóvenes mexicanos” (Ávila y Aguilar, 2003, p. v), se emplearon los lemas con las marcas siguientes: “uso coloquial”, “uso popular”, “uso ofensivo” y “uso grosero”. Revisando el diccionario entrada por entrada, se identificaron dichas marcas y se creó al mismo tiempo una lista electrónica de los lemas marcados. Este procedimiento se llevó a cabo dos veces de manera

independiente. Las dos listas obtenidas se cotejaron entre sí y se verificaron las discrepancias contra el diccionario mismo. En cuanto a la identificación de marcas, al igual que sucedió con el DEUM y con el DBM, en este diccionario la marcación geográfica de “México” es más bien implícita. Esto encuentra su justificación en la acotación contenida en la introducción de este diccionario respecto de los destinatarios del mismo, que son las “niñas, niños y jóvenes mexicanos” (Ávila y Aguilar, 2003, p. v). Los resultados de la búsqueda de marcas se presentan en la Tabla 3.8 que aparece a continuación. En este diccionario se encontraron un total de 651 marcas “uso coloquial”, 126 marcas “uso popular”, 39 marcas “uso ofensivo” y 71 marcas “uso grosero”. Con ello se identificaron un total de 887 marcas distribuidas en 821 acepciones. Aquí cabe señalar que en la composición de la lista de lemas a explorar en el corpus, en este diccionario se decidió hacer una lista más bien de acepciones que de lemas. Esta última decisión se tomó debido a los resultados de la exploración de otros diccionarios, como el *AHDOM* y el *PDLPM*, que muestran que la lematización de éstos es poco confiable. Por ello, se pensó que sería mejor trabajar con acepciones, sobretodo porque, en una posible elaboración posterior del diccionario cuya ejemplificación ha sido proyectada aquí, la mayoría de las entradas del *AHDOM*, que representan alrededor de tres cuartas partes del total de las fuentes secundarias, tendrían que ser lematizadas nuevamente.

Tabla 3.8. Conteo por letra de los lemas con marcas del subestándar en el *DIME*

Letra	Coloquial	Popular	Ofensivo	Grosero	Subtotal
A	45	4	-	-	49
B	26	1	-	-	27
C	85	34	5	17	141
CH	41	13	13	17	84
D	17	1	1	-	19
E	24	1	-	1	26
F	14	7	-	1	22
G	27	7	1	-	35
H	18	4	2	2	26
I	6	-	-	-	6
J	16	7	5	5	33
L	26	5	-	2	33
LL	2	-	-	-	2
M	68	10	2	12	92
N	2	1	-	-	3
Ñ	2	1	-	-	3
O	27	-	2	2	31
P	83	9	6	10	108
Q	9	-	-	-	9
R	24	5	-	1	30
S	28	6	-	-	34
T	33	9	1	-	43
U	4	-	-	-	4
V	18	1	1	1	21
Y	2	-	-	-	2
Z	4	-	-	-	4
Subtotal	651	126	39	71	Total

887

Marcas dobles	56
Marcas triples	5
Total de Lemas	821

Una vez comentados los procedimientos para la extracción de las fuentes secundarias de este estudio, la Tabla 3.9 muestra finalmente el total de entradas o artículos de cada uno de los diccionarios, el total de materiales léxicos extraídos de los mismos (que constituyen las fuentes secundarias de este trabajo) y el tamaño requerido para obtener una muestra aleatoria representativa de cada uno de dichos subconjuntos del léxico del subestándar. Este último dato está basado en lo que dicen Krejcie y Morgan respecto del número de elementos mínimos a incluir en una muestra para que ésta sea representativa de la población de que se le extrae (como se cita en Gay y Airasian, 2002, p. 113).

Tabla 3.9. Fuentes secundarias del léxico subestándar del español en México

Diccionario	Total de		Entradas utilizadas	Tamaño requerido de la muestra
	entradas	acepciones		
DEUM	14,000 (aprox)	60,000 (aprox)	568	226
DIME	13,000 (aprox)	22,000 (aprox)	821	163
DRAE	88,431	190,581	202	132
	Total de entradas			
DBM	6,200 (aprox)		141	103
DTV	2,307		1,971	321
PDLPM	1,667		34	32
AHDOM	9,613		9,612	370
Lista fusionada			13349	1347

En general para Krejcie y Morgan, las muestras para poblaciones con 100 elementos o menos se deberían acercar al total de la población, alrededor de los 500 elementos deberían contener aproximadamente el 50%, en torno a los 1,500 elementos se debería

incluir más o menos el 20%, y por arriba de los 5,000 elementos en la población la muestra escasamente tendrá que rebasar los 400 elementos. Ahora bien, más allá de estas pautas generales para determinar la representatividad de un muestreo, el total de elementos que se incluyó en este trabajo respecto de la muestra de cada diccionario se determinó por medio de la utilización de la tabla, mucho más precisa, que proporcionan Krejcie y Morgan para poblaciones de diversos tamaños (como se cita en Gay y Airasian, 2002, p. 113). Así podemos ver que de todos los diccionarios fuente, de los cuales se extrajo un total de 13,349 entradas, se tuvo que hacer un muestreo total mínimo de 1,347 elementos para explorar la representación de las fuentes secundarias en el corpus. Para obtener el muestreo aleatorio representativo de los diccionarios, se ha aplicado un programa de cómputo de estadística en forma independiente al subconjunto del léxico subestándar de cada uno de los mismos. Debido a que la aplicación se ha hecho de manera separada para cada diccionario, se puede decir que el tipo de muestreo que se ha utilizado en este trabajo es un muestreo aleatorio estratificado. Este tipo de muestreo se ha utilizado con la finalidad de explorar la representatividad del *CREA* respecto de cada uno de los diccionarios de manera independiente. Con ello se espera poder hacer inferencias más certeras sobre el comportamiento potencial de toda la población que conforma las fuentes secundarias del léxico subestándar del español en México.

3.08 Manejo de las listas de las fuentes secundarias

La primera etapa en el manejo de la información extraída de las fuentes secundarias fue la extracción misma de los datos pertinentes para su documentación en el *CREA*. Para esta extracción, comentada de manera particular para cada diccionario en la sección anterior, se decidió crear listas en hojas de cálculo de Excel (ver sección 3.07). Estas

primeras listas que se crearon se nombraron con la palabra *lista* seguida de las siglas del diccionario de procedencia. Aunque los conceptos a manejarse en cada lista variaron de acuerdo al diccionario mismo del cual se extrajo la información, había algunos datos comunes a todas las listas. Éstos fueron: un número progresivo en la lista del lema o acepción que se extrajo del diccionario, el lema o encabezado de la entrada y el número de la página en la cual se encontraba el lema. La Figura 3.1 muestra un ejemplo de la *lista DRAE*, donde se puede ver que el número progresivo del lema extraído corresponde a la primera columna, o columna A, de la hoja de cálculo, el lema o encabezado se encuentra en la segunda columna, o columna B, y el número de página en el cual se encontró el lema dentro del diccionario aparece en la quinta columna, o columna E.

Figura 3.1. La *lista DRAE* en la primera etapa de trabajo con las fuentes secundarias

	A	B	C	D	E	F
1	Núm	Lema	Marca	No. acep	No. pág.	Colocación o variante
2	1	acordeón	coloq.	2	35	-
3	2	albarda	coloq.	1	88	albarda sobre aparejo
4	3	alipús	coloq.	2	111	-
5	4	anca	coloq.	1	147	dar ancas vueltas
6	5	andar	coloq.	1	149	ándale
7	6	andar	coloq.	1	149	ándele
8	7	apendejar	malson.	2	179	-
9	8	apendejar	malson.	3	179	-
10	9	babosada	coloq.	1	268	-
11	10	babosada	coloq.	2	268	-

Además, a las listas de la mayoría de los diccionarios, de hecho a todas ellas salvo a la del *DEUM* que fue el primer diccionario en trabajarse, se les añadió una columna llamada de “Colocación o Variante”, donde se incluyeron varias de las palabras que acompañaban al lema. Esto se hizo sobre todo en el caso de las unidades pluriverbales más extensas. En la Figura 3.1 se puede ver que esta columna corresponde a la sexta de la *lista DRAE*. En esta columna se capturó también, en el caso de la *lista DBM*, algunas variantes no lematizadas que ofrecía el mismo diccionario inmediatamente después del encabezado de la entrada propiamente dicho. A la *lista DBM*, a la *lista DEUM*, a la *lista DRAE* y a la *lista DIME* se les añadió una columna con la marca que dichos diccionarios colocaban a sus lemas. En este respecto, las listas del *DEUM*, del *DRAE* y del *DIME* contienen también una columna con el número de acepción o acepciones afectadas por dichas marcas. En la Figura 3.1 se observa que la tercera columna de la *lista DRAE*, o columna C, contiene las marcas de los lemas que se encuentra en la segunda columna, o columna B. Por su parte, en la cuarta columna, o columna D, aparecen las acepciones afectadas por las marcas de la tercera columna. Así por ejemplo, si vemos el lema *alipús*, con el número progresivo tres en la *lista DRAE*, podemos saber que está considerado como una voz coloquial cuando se le utiliza con el segundo significado de los que lista el *DRAE*, que en este caso corresponde al de “bebida alcohólica (RAE, 2001, p.111). De manera distintiva, la *lista DTV* fue construida con una columna exclusiva para sus variantes, independiente a la columna de colocación. Esto último se debió sobre todo al gran número de variantes no lematizadas, mucho más numerosas que aquellas presentadas por el *DBM*, que el *DTV* ofrece respecto de algunas de sus entradas. Finalmente, la *lista PDLPM* tiene una columna con el número de apariciones de los lemas. Hay que recordar que este diccionario no ofrecía en muchos casos unidades léxicas lematizadas, propiamente hablando, sino una serie arbitraria de

concordancias o apariciones en varios contextos de un mismo lema. Así pues, esta columna muestra el número de veces que un lema, que más bien fue reconstruido por el autor de este trabajo para aparecer tal cual en la lista, se presentó en todo el diccionario. Esta columna buscaba también justificar la falta de información en la columna del número de página del lema, que en el caso de los lemas reconstruidos no podía ser llenada con ningún contenido. Todas estas listas fueron posteriormente ampliadas con una columna para el significado del lema o encabezado.

En una segunda etapa del manejo de la información extraída de las fuentes secundarias, se decidió añadirle a todas las listas el significado de las entradas que las conformaron. Las listas resultantes tenían el mismo nombre de las anteriores, pero fueron ubicadas en una nueva carpeta. La Figura 3.2 es un ejemplo de la *lista DRAE* con el significado incluido en la séptima columna, o columna G.

Figura 3.2. La *lista DRAE* en la segunda etapa de trabajo con las fuentes secundarias

	B	C	D	E	F	G
	Lema	Marca	No. acep	No. pág.	Colocación o variante	Significado
1	acordeón	coloq.	2	35	-	m. coloq. Méx. chuleta (apunte para usarlo dis exámenes).
2						
3	albarda	coloq.	1	88	albarada sobre aparejo	expr. coloq. Méx. albarda sobre albarda
4	alipús	coloq.	2	111	-	m. coloq. Méx. Bebida alcohólica.
5	anca	coloq.	1	147	dar ancas vueltas	fr. coloq. Méx. Conceder una ventaja en cualqu
6	andar	coloq.	1	149	ándale	expr. coloq. Méx. U. para animar a alguien a h
7	andar	coloq.	1	149	ándele	expr. coloq. Méx. ándale
8	apendejar	malson.	2	179	-	prnl. Malson. Méx. acobardarse.
9	apendejar	malson.	3	179	-	prnl. malson. Méx. Hacerse bobo, estúpido.
10	babosada	coloq.	1	268	-	f. coloq. Méx. Necedad, tontería.

La finalidad del trabajo de captura del significado fue ante todo el hacer de las listas materiales más independientes y manejables en el momento de la documentación. El tener prácticamente toda la información en la misma lista a documentar permitía trabajar más rápidamente con las hojas de cálculo e Internet, sin necesidad de recurrir a los diccionarios de papel de mucho más lento acceso. La única modificación mayor que implicó la captura del significado en las listas originales fue la partición de la *lista AHDOM* en dos listas, la *lista AHDOM 01* y la *lista AHDOM 02*. Esta división se llevó a cabo debido a que el tamaño de una sola lista con significados para las 9,612 entradas utilizadas de este diccionario volvía al archivo electrónico demasiado pesado y difícil de desplegar y manipular. Cabe aclarar que la ayuda proporcionada por los becarios en la construcción de las listas de las fuentes secundarias (ver sección 3.07) sólo tuvo lugar hasta aquí. El resto del trabajo relacionado con el manejo de las listas, y del que se hablará a continuación, corrió a cargo únicamente del autor de esta tesis.

3.09 Organización del muestreo aleatorio estratificado

La tercera etapa en el manejo de las listas de las fuentes secundarias constituye el último paso de preparación de dichas listas antes de comenzar con el trabajo de documentación en el corpus electrónico. En esta etapa se llevó a cabo la creación de las listas del muestreo aleatorio. Al nombre de estas nuevas listas se le agregó únicamente la palabra *sample* para diferenciarlas de las anteriores. Los nombres de las nuevas listas quedaron, por tanto, de la siguiente manera: *lista DEUM sample*, *lista DRAE sample*, *lista DBM sample*, y así sucesivamente. Cabe aclarar de antemano que estas nuevas listas requirieron una vez más dividir la información de la *lista AHDOM*, ahora en tres listas: la *lista AHDOM 01 sample*, la *lista AHDOM 02 sample* y la *lista AHDOM 03 sample*. Todo

ello con la finalidad de manejar más fácilmente la información de este diccionario. También se debe mencionar que durante la confección de las listas *sample*, anterior al trabajo de documentación, se estandarizó el formato de las mismas. Así, se introdujeron en ellas columnas para todos los posibles datos que presentaban los distintos diccionarios, quedando las listas *sample* de la manera que se describe a continuación. Primero se colocó una columna titulada “*sample*”, donde se indicó si la fila en cuestión había sido seleccionada o excluida por el programa de cómputo que llevó a cabo el muestreo aleatorio. Esta columna corresponde a la columna A de la *lista DRAE sample*, en la Figura 3.3. Esta primera columna estaba seguida de una columna donde se anotó un número progresivo para los lemas seleccionados durante el muestreo aleatorio, que en la Figura 3.3 se observa en la columna B. A estas dos columnas se le sumaron los siguientes datos en sendas columnas y en el orden aquí enunciado: las siglas del diccionario a que pertenecieran los lemas, el número progresivo del lema o encabezado en la lista general de lemas extraídos de dicho diccionario, el lema o encabezado en sí, el número de acepción del lema utilizado, la marca correspondiente, el número de página de procedencia del lema, el significado del lema o entrada y la colocación o variante. Todos estos datos aparecen, siguiendo el mismo orden, en las columnas C a la J de la Figura 3.3. Así por ejemplo, si tomamos el lema *bañar* de la *lista DRAE sample* en dicha figura, nos encontramos que en su columna A contiene un número uno, lo cual quiere decir que este lema fue elegido para ser parte de la muestra. Vemos también que su número progresivo en la muestra es el número nueve (columna B), que el diccionario al que pertenece es el *DRAE* (columna C), que su número progresivo en la extracción de lemas del diccionario fue el número diez (columna D), que su lema es *bañar* (columna E), que la acepción de este lema que es parte del subestándar es la primera del diccionario (columna F), que dicha acepción es parte del subestándar porque está

considerada como coloquial (columna G), que el lema se encuentra en la página 284 del *DRAE* (columna H), que de acuerdo con este diccionario significa lo mismo que “a paseo” (columna I), y que comúnmente aparece en el contexto “a bañar”, como cuando se dice “vete a bañar” (columna J).

Figura 3.3. La lista *DRAE* sample en la tercera etapa de trabajo con las fuentes secundarias

A	B	C	D	E	F	G	H	I	J
row	row	Dic	Nu	Lema	Ace	Marca	Pág	Significado	Calificación o v
1	1	DRAE	1	soardeón	2	coloq.	35	m. coloq. Méx. chuleta (apunte para usarlo disimuladamente en los exámenes)	-
1	2	DRAE	2	albarda	1	coloq.	88	expr. coloq. Méx. albarda sobre albarda	albarda sobre
1	3	DRAE	3	slipus	2	coloq.	111	m. coloq. Méx. Bebida alcohólica	-
1	4	DRAE	4	anica	1	coloq.	147	f. coloq. Méx. Conceder una ventaja en cualquier juego, sobresalir en él.	dar antes vu
1	5	DRAE	5	andar	1	coloq.	149	expr. coloq. Méx. U. para animar a alguien a hacer algo	ándale
1	6	DRAE	7	apendejar	2	malson.	179	prnl. Malson. Méx. acobardarse.	-
1	7	DRAE	8	apendejar	3	malson.	179	prnl. malson. Méx. Hacerse bobo, estúpido.	-
1	8	DRAE	9	baborada	1	coloq.	268	f. coloq. Méx. Necedad, tontería.	-
1	9	DRAE	10	baborada	2	coloq.	268	f. coloq. Méx. Cosa intrascendente, sin valor.	-
1	10	DRAE	11	bañar	1	coloq.	284	loc. adv. coloq. Méx. a paseo. ¡Vete a bañar! ¡Ándá a bañarte!	a bañar

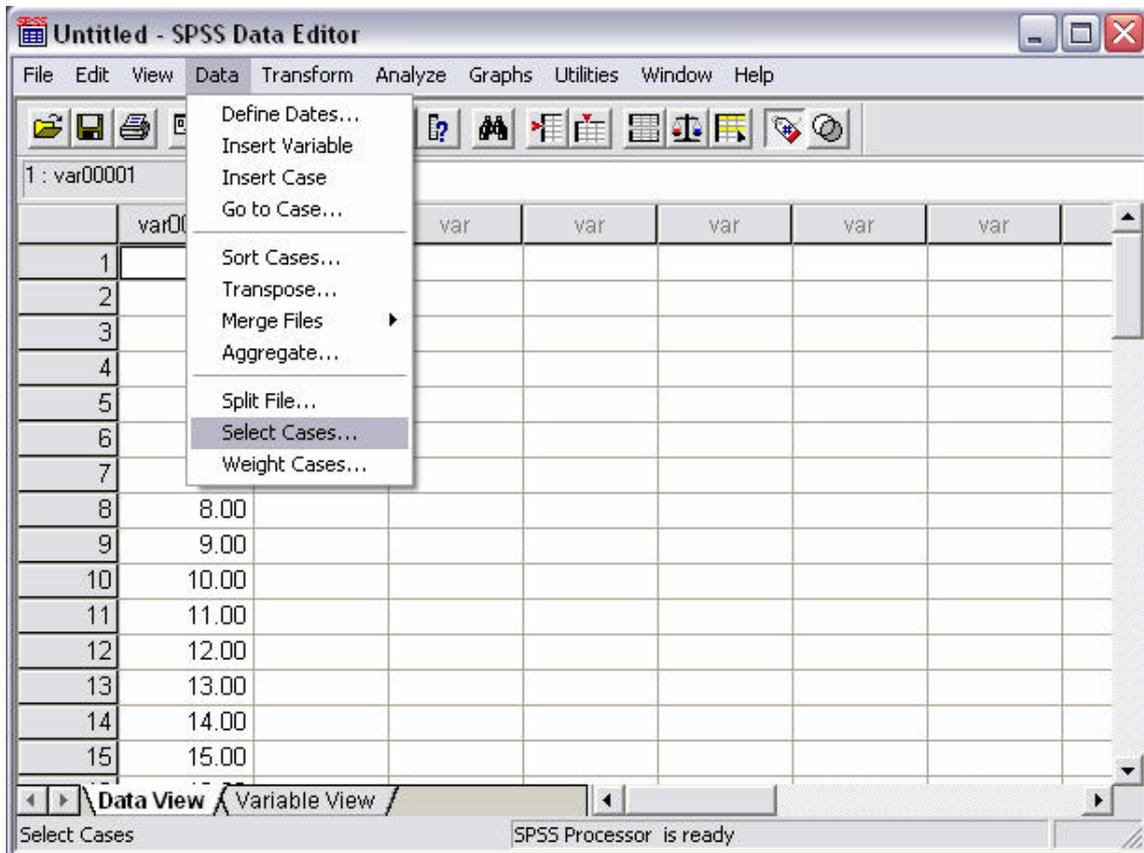
Además, se debe comentar que en las listas sample se añadieron cinco columnas, a llenar durante el proceso de documentación: una columna para anotar la formula de búsqueda en el *CREA*, una columna para anotar el resultado de dicha búsqueda, y tres columnas para indicar si, al analizar las concordancias arrojadas por el corpus, el lema quedaba sin ejemplo, con dos ejemplos o con un sólo ejemplo. Estas columnas aparecen en las listas sample de la columna K a la columna O; sin embargo, en la Figura 3.3 no se aprecian pues no tendría sentido mostrarlas aquí ya que durante esta etapa estas columnas

no fueron llenadas con ninguna información. Finalmente, hay que decir que las distintas columnas fueron igualadas en su ancho y en el tipo y tamaño de letra de su contenido en las listas de todos los diccionarios.

Ahora bien, con la finalidad de llevar a cabo un muestreo aleatorio estratificado del lecionario tentativo obtenido de todas las fuentes existentes del léxico subestándar del español en México, se tuvo que hacer una selección aleatoria de los elementos mínimos para la obtención de muestras representativas de los distintos diccionarios. El tamaño de las muestras representativas se determinó, según se indicó arriba, siguiendo las recomendaciones de Krejcie y Morgan, como se citan en Gay y Airasian (2002). La selección aleatoria en sí se hizo con la ayuda del programa de cómputo para estadística SPSS para Windows, en su versión 10.0.1 Estándar (1999).

Para hacer la selección aleatoria se hizo lo siguiente con todas las listas. Se comenzó corriendo el programa SPSS para Windows. En la ventana de bienvenida se inició seleccionando la opción “Type in data” ante la pregunta “What would you like to do?”. Esta primera instrucción abre una hoja de datos, como la que se muestra en la Figura 3.4. Para comenzar a introducir datos en la hoja nueva de SPSS para Windows, se hizo un copiado de la columna en la hoja de Excel que contenía el número progresivo de los lemas extraídos en la lista general del diccionario en cuestión. Esta columna se pegó en la primera columna de variantes “var” de la hoja de datos de SPSS para Windows, la cual después del copiado recibe automáticamente el nombre de “var00001”. A continuación se seleccionó la opción “Select cases” mostrada al hacer clic sobre la opción “Data” de la barra del menú principal. Las operaciones hasta aquí descritas se pueden ver también en la misma Figura 3.4.

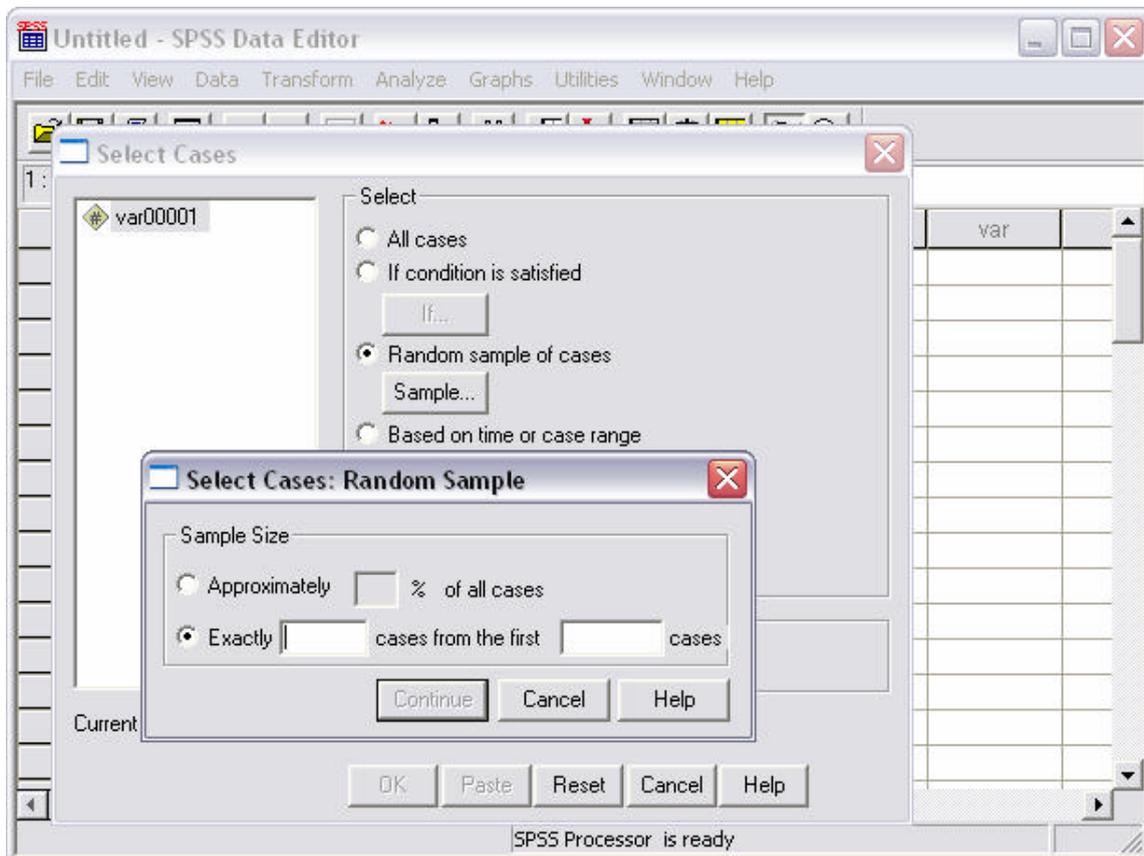
Figura 3.4. Hoja de datos del programa SPSS para Windows



Después de seleccionar la opción “Select Cases” en el menú “Data” de la barra de menú principal, el programa arroja una ventana llamada igualmente “Select cases”, la cual mostraba ya seleccionada la columna “var00001”. Esta nueva ventana aparece en la Figura 3.5. En esta ventana, se seleccionó la opción “Random sample of cases” para obtener un muestreo aleatorio de dicha columna, exportada desde la lista de los diccionarios. Antes de llevar a cabo el muestreo aleatorio en sí, se hizo clic en el botón “Sample” habilitado por la selección de la opción “Random sample of cases”, para desplegar la ventana “Select cases: Random Sample”. Esta ventana ofrece la posibilidad de obtener un muestreo con un

porcentaje preciso del total de la población, en su opción “Aproximately”, o un muestreo con un número específico de casos de entre un número determinado de casos iniciales de la columna seleccionada, en su opción “Exactly...”. Para el muestreo requerido en este trabajo, se optó por la segunda opción. Todas estas posibles opciones de SPSS para Windows en la obtención de una muestra aleatoria se pueden visualizar también en la Figura 3.5. Respecto del tamaño de la muestra solicitado en la opción “Exactly...”, se indicó primero el número de casos necesarios para obtener una muestra representativa de la lista del diccionario en cuestión, siguiendo las recomendaciones de Krejcie y Morgan (como se citan en Gay y Airasian, 2002), y se pidió su extracción del total de casos de la lista del diccionario. Una vez introducidos estos dos datos se hizo clic en el botón “Continue” que cierra la última ventana “Select cases: Random Sample”, mostrada igualmente en la Figura 3.5. Después se hizo clic en el botón “OK” de la ventana “Select cases”. Como resultado de esta operación el programa SPSS para Windows generaba una columna después de la columna “var00001”, llamada “filter_\$”, donde asignaba un número uno a los casos seleccionados y un cero a los casos eliminados del muestreo. La columna “filter_\$” del programa SPSS para Windows se copió a la columna titulada “sample” de las listas del muestreo aleatorio en Excel.

Figura 3.5. Opciones para la obtención de un muestreo aleatorio en SPSS para Windows



Finalmente, en las listas sample que contenían la columna con la selección aleatoria del programa SPSS para Windows, se aplicó la función “Autofiltro” de Excel. Esta función se activó eligiendo la opción “Filtro” mostrada en el menú “Datos” de la barra de menú principal. Con esta opción se pudo desplegar una lista que contuviera únicamente aquellas filas que en su columna “sample” contuvieran un número uno, es decir, aquellas que hubiesen sido seleccionadas por el programa de estadística. El resultado de esta operación

se puede ya apreciar desde la Figura 3.3, donde los números progresivos de la muestra (columna B) no coinciden con los números progresivos de la extracción de lemas (columna D), debido a que de los lemas extraídos sólo están desplegados aquellos seleccionados por el programa SPSS para Windows. En este respecto, cabe hacer notar que en la columna B, donde se anotó el número del lema seleccionado para la muestra, éste se introdujo manualmente, puesto así lo requería la lista una vez que se le aplicó la función “Autofiltro”. Esta última numeración se llevó a cabo para corroborar que el tamaño de la muestra representativa arrojada en la columna “filter_\$\$” de la hoja de datos de SPSS Windows coincidiese efectivamente con el número de filas mostradas por el “Autofiltro” de Excel. En este sentido, no hubo ninguna discrepancia. Todo este procedimiento, como se dijo anteriormente, se realizó de manera individual para las distintas listas de los diccionarios fuente. Con esto se terminó la preparación de las listas de las fuentes secundarias.

3.10 Documentación de las listas del muestreo

Una vez preparadas las listas definitivas para llevar a cabo la documentación, me preocupó la posibilidad de tener casos en que algunos lemas con las mismas acepciones se presentaran en diferentes diccionarios, y que por ello, la documentación de dichos lemas se llevara a cabo de forma múltiple y repetitiva. Para evitar esto, se construyó una lista con los lemas seleccionados durante el muestreo aleatorio procedentes de todos los diccionarios. Esta lista fue sorteada por orden alfabético para poder detectar las posibles coincidencias de entradas entre las distintas muestras. En esta lista, llamada *lista fusionada sample* y presentada en la Figura 3.6, se incluyeron sólo los datos necesarios para la identificación de los lemas repetidos en la muestra. Así las columnas de esta lista contenían los siguientes datos: una numeración progresiva general de los lemas que compusieron la lista fusionada

misma (columna A), el número del lema seleccionado durante el muestreo aleatorio de acuerdo a cada diccionario (columna B), el nombre del diccionario del que procediese el lema en cuestión (columna C), el número del lema o encabezado en la lista general de lemas extraídos de dicho diccionario (columna D), el lema (columna E), el número de acepción del lema (columna F), la marca del mismo (columna G), y el número de página en que se halló el lema dentro del diccionario (columna H). Cuando se conformó la *lista fusionada sample* con estos datos, se sombrearon las celdas en la columna de los lemas donde había coincidencias entre los mismos. Así por ejemplo, en la parte de la *lista fusionada simple* que se muestra en la Figura 3.6, se puede ver que el lema *chafirete*, que se encuentra sombreado, está presente en varios diccionarios, el *DBM*, el *DRAE* y el *DTV*. De la *lista fusionada simple* se imprimió una copia física para consultar durante la documentación individual de las listas sample.

Figura 3.6. La *lista fusionada sample* utilizada durante la documentación

	A	B	C	D	E	F	G	H
1	Folio	# sample	Dicc	Núm	Lema	Acep	Marca	Pág
279	278	18	DBM	21	chafirete		despect	41
280	279	29	DRAE	39	chafirete	1	despect.	512
281	280	93	DTV	499	chafirete			63
282	281	30	DRAE	40	chahuisclé	1	coloq.	513
283	282	86	DEUM	208	chale	-	Popular Ofensivo	309
284	283	87	DEUM	209	chale!	-	Popular	309
285	284	31	DRAE	41	chamba ⁴	2	coloq.	514
286	285	76	AHDOM 1	2243	chamois			80
287	286	94	DTV	508	champerico			64
288	287	31	DIME	126	chanchullo		uso popular	120
289	288	77	AHDOM 1	2254	chancla			80
290	289	88	DEUM	210	chancla	4	Popular	310
291	290	89	DEUM	212	chango	3	Popular	310
292	291	95	DTV	519	chapuza			65

Para llevar a cabo la documentación en sí, se siguieron los pasos enunciados a continuación. Primero se consultó la *lista fusionada sample* y se identificaron las coincidencias de lemas entre distintas listas. Esto se llevó a cabo con la finalidad de hacer una sola búsqueda en el corpus electrónico para los lemas comunes y generar posibles fichas de ejemplo compartidas. Así, si se encontraban coincidencias entre lemas, se procedía a verificar si la coincidencia se hallaba también a nivel del significado. Si la coincidencia se encontraba sólo a nivel del lema y no del significado se podía generar por lo menos una fórmula de búsqueda común y revisar las concordancias del corpus teniendo

en cuenta los varios significados presentes en las distintas listas. Si la coincidencia se daba también a nivel del significado se podía incluso generar fichas de ejemplos comunes para varios casos en distintas listas. La coincidencia de significado se corroboró con las listas *sample* completas o directamente con los ejemplares de los diccionarios. En los casos en que se detectaron coincidencias de lemas y significados, se procedió a escribir una definición simplificada al margen de la *lista fusionada sample*, para tener en cuenta dicha definición durante el análisis de las concordancias. Una vez que se llevó a cabo esta revisión de las coincidencias entre distintas listas *sample* se podía llevar a cabo una primera búsqueda de los lemas similares, con fórmulas y significados comunes.

Tanto para los casos similares como para los que no lo eran, lo primero que se hizo, en el proceso mismo de documentación, fue desplegar la lista *sample* en la que se estuviera trabajando y salvarla con un nuevo nombre para guardar la información reunida durante este proceso. De esta manera, las listas pasaron a ser renombradas con la palabra *lista*, más el nombre del diccionario al que pertenecieran, y la palabra *documentada*. Con la nueva nominación, las listas adquirieron nombres tales como *lista DEUM documentada*, *lista DRAE documentada*, *lista DBM documentada*, y así sucesivamente. En estas listas se identificaba en primer lugar el lema del que se buscarían ejemplos de uso en el corpus. Después, se anotaba la fórmula o las fórmulas de búsqueda que se consideraba que pudiesen desplegar más eficientemente las concordancias pertinentes para las distintas variaciones morfológicas del lema buscado. Aquí hay que recordar que las concordancias son las listas de ejemplos de uso que proporcionan las computadoras que manejan las grandes bases de datos llamadas corpus electrónicos (Sinclair, 1985, p. 83). Para algunos lemas se utilizó más de una fórmula de búsqueda, y algunos de ellos alcanzaron a tener más de cinco fórmulas distintas. La Figura 3.7 muestra un ejemplo de las últimas columnas de la

lista DRAE documentada. En esta figura se puede ver cómo del lema *andar* y del lema *alipús* se utilizaron dos fórmulas de búsqueda distintas, mientras que del resto de los lemas que aparecen en esta figura se utilizó solamente una fórmula.

Figura 3.7. Llenado de la lista DRAE documentada

	E	F	G	H	I	J	K	L	M	N	O
	Lema	Ac	Marc	Pág	Significado	Colocación o variante	Fórmula de consult	Resultado	s / -s	2 r -s	1 -s
1	acordeón	2	coloq.	35	m. coloq. Méx. chuleta (apunte para usarlo disimuladamente en los exámenes).	-	acordeón	29/14			1
2	albarda	1	coloq.	88	expr. coloq. Méx. albarda sobre albarda	albarda sobre aparejo	albarda sobre aparejo	0		1	
3	alipús	2	coloq.	111	m. coloq. Méx. Bebida alcohólica.	-	alipus-alipús	0-0		1	
4	anca	1	coloq.	147	fr. coloq. Méx. Conceder una ventaja en cualquier juego, sobresalir en él.	dar ancas vueltas	ancas vueltas	0		1	
5	andar	1	coloq.	149	expr. coloq. Méx. U. para animar a alguien a hacer algo.	ándale	ándale-- ándele	45/17-- 14/18			1
6	apendejar	2	malson.	179	prnl. Malson. Méx. acobardarse.	-	apendej*	9/7		1	
7	apendejar	3	malson.	179	prnl. malson. Méx. Hacerse bobo,	-	apendej*	9/7			1
8	babosada	1	coloq.	268	f. coloq. Méx. Necedad, tontería.	-	babosada*	8/7			1
9	babosada	2	coloq.	268	f. coloq. Méx. Cosa intrascendente, sin	-	babosada*	8/7			1
10	bañar	1	coloq.	284	loc. adv. coloq. Méx. a paseo. ¡Vete a	a bañar	a bañar	23/16		1	

Después, de generar un fórmula de búsqueda para el lema en cuestión, ésta se anotaba en la lista documentada y se introducía en el buscador de la pantalla principal del CREA (esta pantalla aparece más adelante en el segundo elemento de la Figura 3.8). Cabe aclarar que al corpus electrónico aquí utilizado, el CREA, se puede acceder a través de Internet en la página principal de la RAE (disponible en <http://www.rae.es>). Para acceder al CREA se debe hacer clic en el hipervínculo del lado izquierdo de la página llamado “Consulta banco de datos”, que despliega en su parte inferior dos opciones: “Corpus actual”

y “Corpus histórico”. La primera de estas opciones es la que da acceso directo a la pantalla de consulta del *CREA*. La página principal de la RAE, con los hipervínculos mencionados, se muestra en el primer elemento de la Figura 3.8. Como se mencionó recién, en un segundo elemento de esta misma figura se ha incluido la pantalla del buscador del *CREA*.

Figura 3.8. Página principal de la RAE y del CREA

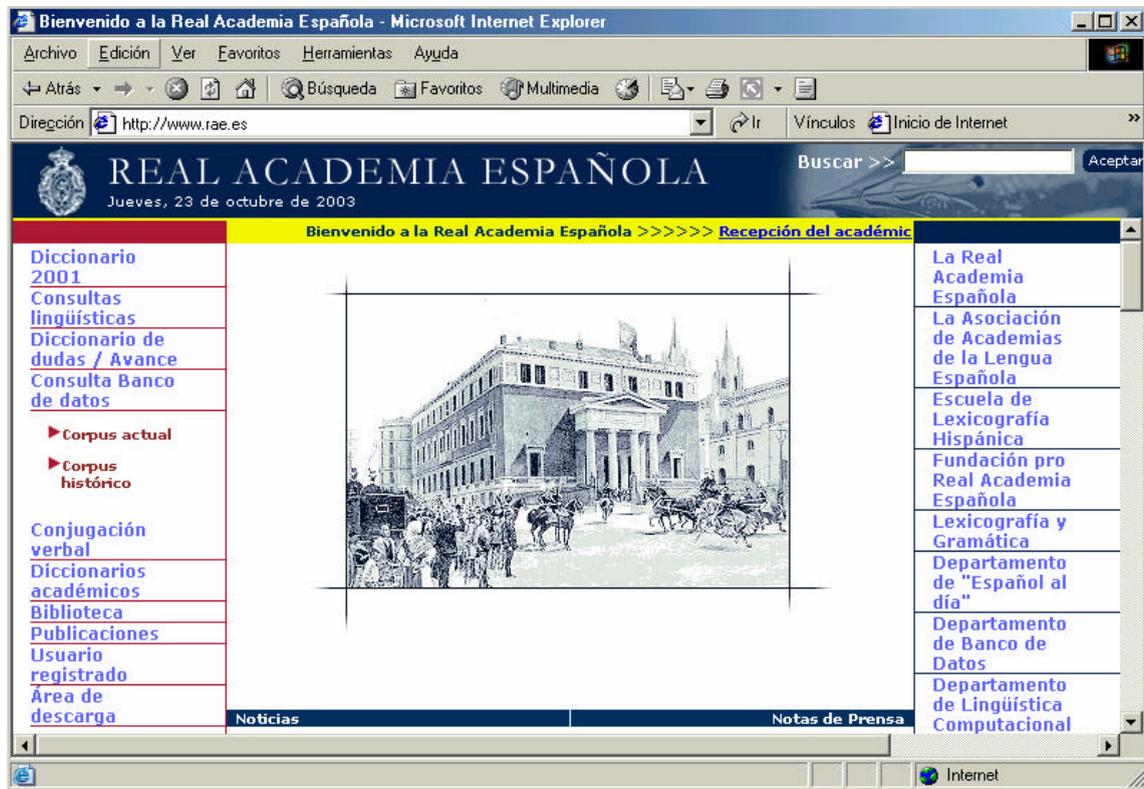
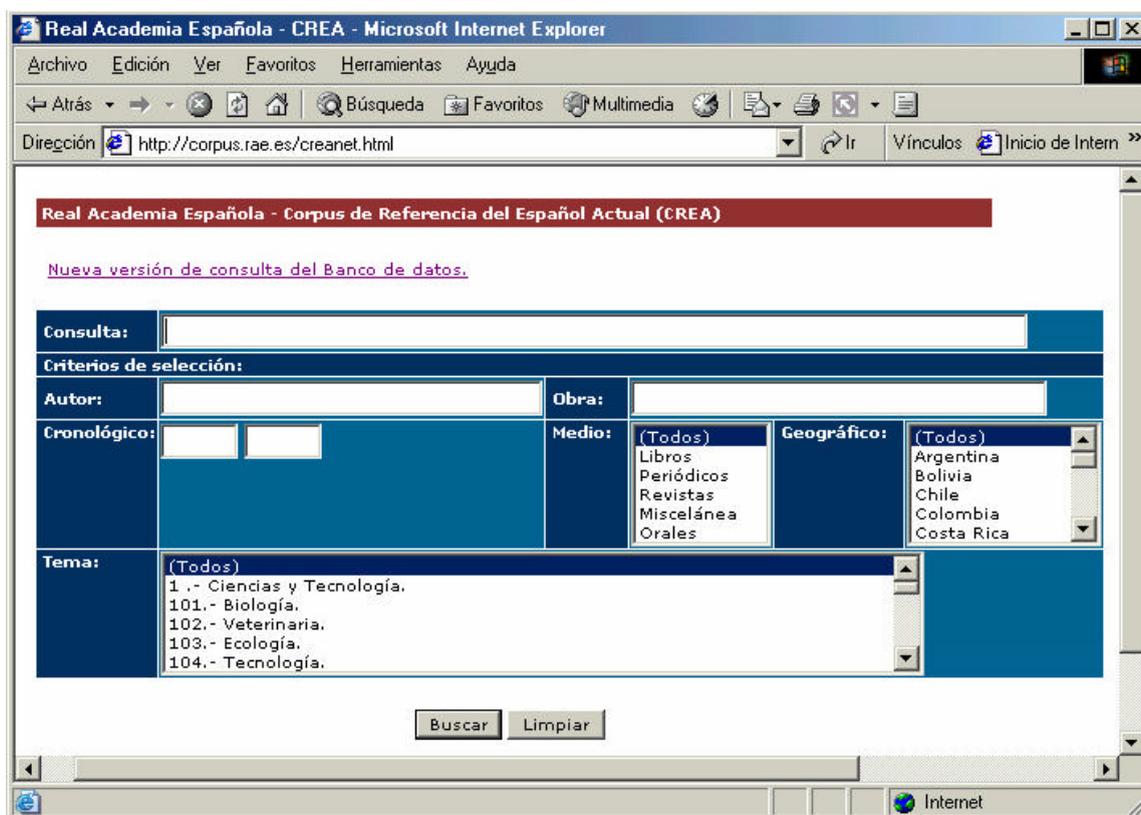


Figura 3.8. (continuación)



Para definir las fórmulas de consulta más eficientes respecto a la recuperación de concordancias, se tomaron en cuenta las anotaciones sobre la “Sintaxis del lenguaje de consulta” que ofrece el *Manual de consulta* del CREA (RAE, 2003, octubre). La búsqueda, además, se llevó a cabo estableciendo un criterio restrictivo de selección del ámbito geográfico que recuperase tan sólo aquellas concordancias que perteneciesen a documentos de origen mexicano. Este criterio de selección se puede indicar seleccionando la opción “México” con la barra de desplazamiento del criterio ‘Geográfico’, mostrado en la parte central derecha del segundo elemento de la Figura 3.8. Una vez solicitada una búsqueda, el corpus muestra el total de casos o concordancias, en la base de datos del CREA, que coinciden con la fórmula. Además, indica el total de documentos en los cuales se

encuentran estos casos. Todos estos datos son desplegados en una ventana como la que se muestra en el primer elemento de la Figura 3.9. Las concordancias en sí se pueden desplegar haciendo clic sobre el botón “Recuperar” de la sección “OBTENCIÓN DE EJEMPLOS”, en la parte inferior de la mencionada ventana. En el ejemplo de búsqueda cuyos resultados se encuentran en el primer elemento de la Figura 3.9, vemos que la fórmula de búsqueda “ching*” (que permitiría recuperar, entre otras cosas, todas las formas conjugadas del lema *chingar*) arroja un total de 462 concordancias, en un total de 72 documentos distintos. El despliegue de estas concordancias se puede ver en el segundo elemento de la misma Figura 3.9. A su vez, si desea ver un contexto más amplio de una concordancia en concreto, se puede hacer clic sobre la fórmula resaltada de la concordancia en cuestión, la cual despliega una pantalla como la mostrada en el tercer elemento de la Figura 3.9. En el caso del ejemplo incluido en este último elemento de la figura, se puede ver que en éste se ha desplegado un contexto más extenso respecto de la concordancia número dos de las que aparecen listadas en el segundo elemento de la misma figura.

Figura 3.9. Resultados de consulta en el CREA

Bienvenido a la Real Academia Española - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

← Atrás → Búsqueda Favoritos Multimedia

Dirección <http://www.rae.es> Ir Vínculos Inicio de Internet

REAL ACADEMIA ESPAÑOLA Buscar >>

Lunes, 27 de octubre de 2003

Resultado de la consulta al banco de datos

Consulta:	<i>ching*</i> , en todos los medios, en CREA , en MÉXICO
Resultado:	462 casos en 72 documentos.

Ver estadística

Filtros: Casos

Ratio: 10

Mantener documentos (Solo para filtro sobre casos).

Filtrar

OBTENCIÓN DE EJEMPLOS

Recuperar Concordancias Normal.

Clasificación:

Agrupación: Marcas:

Listo Internet

Figura 3.9. (continuación)

Bienvenido a la Real Academia Española - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

← Atrás → Búsqueda Favoritos Multimedia

Dirección <http://www.rae.es> Ir Vínculos Inicio de Internet

REAL ACADEMIA ESPAÑOLA Buscar >> Aceptar

Lunes, 27 de octubre de 2003

Concordancias (RAE)

Consulta:	<i>ching*</i> , en todos los medios, en CREA , en MÉXICO
Resultado:	462 casos en 72 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar Concordancias: Normal: Clasificación:

Agrupación: Marcas:

Concordancias.

Pantalla: 1 de 19. [Siguiente](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [Ver párrafos](#)

Nº CONCORDANCIA

1 tendrás que esperarte. - Pos se va a morir, a la **chingada** tu hijo. - Pos ni modo

2 y que la empresa me diga: párale y mándalos a la **chingada**, pero ojalá eso no suc

3 hermano que me platicaba que el gobierno me iba a **chingar** y que lo que más correc

4 tengo con la gente, dicen que ya están "hasta la **chingada** del gobierno", que si

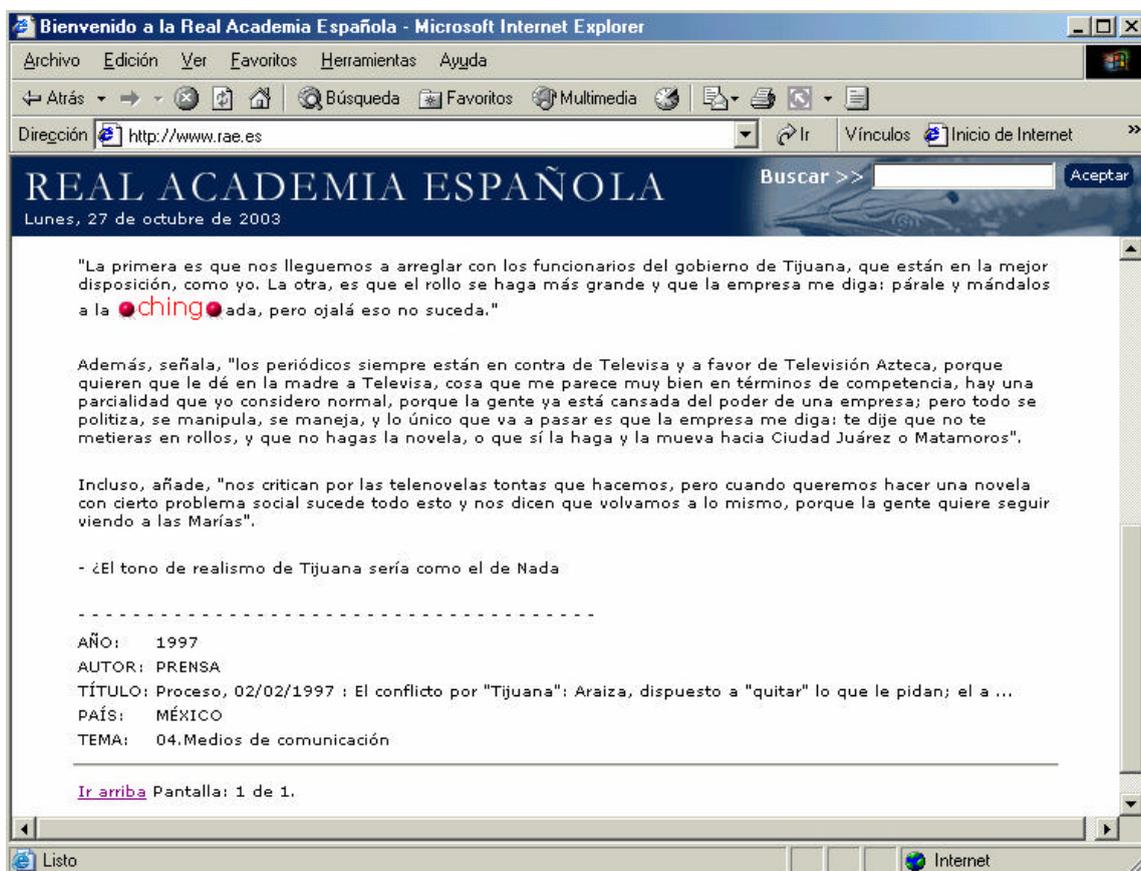
5 empezaremos por ese par de hijas de puta que cómo **chingan** la madre. Nos referimos

6 lemento importante o no? (Ruido) -¡Ay, hijo de la **chingada**! ¿Y este cabrón del Pi

7 qué? (Ruido) -¡Ah, pues a ese güey también hay que **chingarlo**... Es falsa Entrevist

Internet

Figura 3.9. (continuación)



Después de obtener las concordancias de una fórmula en particular, se procedió a hacer el análisis dichas concordancias una por una hasta encontrar un máximo de dos ejemplos, por lema y por significado o acepción. Terminada la revisión de todas las concordancias, se llenaron los datos del lema trabajado, anotando cuántas concordancias o ejemplos contenía el corpus, en cuántos documentos distintos se hallaban dichas concordancias y cuántos de dichos ejemplos habían correspondido al lema y acepción buscada. En este último dato se reportó únicamente si se habían encontrado dos ejemplos, un ejemplo o ninguno. Si regresamos a ver la *lista DRAE documentada* en la Figura 3.8, podemos ver cómo en la columna M, la columna N y la columna O, se reporta que de los

lemas *andar*, *apendejar* (en su tercera acepción) y *babosada* (en su primera y segunda acepción) se encontraron dos ejemplos en el corpus; del lema *acordeón* se encontró un solo ejemplo; y de los lemas *albarda*, *alipús*, *anca*, *apendejar* (en su segunda acepción) y *bañar* no se encontró ningún ejemplo.

En cuanto al hallazgo de ejemplos pertinentes al lema y la acepción buscados, decidí que sería mejor crear fichas para una posible utilización posterior en caso de que se llevase a cabo la elaboración del diccionario. Estas fichas fueron creadas con base en los siguientes criterios. Se elaboraron fichas sólo para aquellos lemas de los que se hubiesen encontrado ejemplos de uso entre las concordancias del corpus. Una vez que se detectaba algún ejemplo de uso que pareciera reflejar el mismo significado o alguno relacionado con el presentado en el diccionario fuente, se procedía a desplegar todo su contexto en el corpus (ver el tercer elemento de la Figura 3.9). Esto se llevaba a cabo haciendo clic sobre la fórmula de búsqueda resaltada con un hipervínculo en la concordancia (ver el segundo elemento de la Figura 3.9). Si el contexto completo corroboraba la coincidencia entre el uso del lema y el significado plasmado por el diccionario fuente, se hacía un copiado de todo el contexto desplegado por el corpus, junto con los datos del documento de procedencia, a un documento nuevo de Word. Esto se intentaba llevar a cabo en un máximo de dos ocasiones por lema o acepción. Sin embargo, en algunos casos en que las concordancias pertinentes al lema trabajado resultaban abundantes, llegué a copiar más de dos ejemplos, sobre todo cuando me parecía encontrar ejemplos especialmente interesantes por los distintos matices de significación que reflejaban. Con todo, el máximo de ejemplos reportados por lema en el conteo final de la ejemplificación nunca fue más de dos. Hay que anotar también que, aun en los casos en que se halló un solo ejemplo de uso, también se generó una ficha electrónica. Ahora bien, una vez copiados los ejemplos recuperados por el corpus en el

documento de Word, se le ponía un encabezado a este último con los datos que permitieran identificar a qué lema le correspondían dichos ejemplos. Para llevar a cabo esto, se llenaban, en la lista documentada, las últimas cuatro columnas de la fila del lema trabajado. En la primera columna se anotaban dos datos, cuántas concordancias o ejemplos contenía el corpus respecto de la fórmula de búsqueda y en cuántos documentos distintos se hallaban dichas concordancias. En las tres columnas restantes se anotaban cuántos ejemplos, entre dichas concordancias, se habían encontrado para el lema en cuestión. Como se dijo anteriormente, respecto de este último dato sólo se indicó si se habían obtenido uno o dos ejemplos para aquellos lemas de los que se elaboró una ficha, y cero ejemplos en los casos en que no se elaboró ninguna.

Una vez llenada la fila correspondiente a un lema, se intentó incluir todos sus datos en el encabezado de los ejemplos contenidos en la ficha. Para hacer esto de manera ágil decidí crear un documento de Excel, llamado *Transposición* donde en una primera columna se colocase el nombre de los contenidos de cada una de las columnas de la lista documentada, pero en forma vertical en vez de horizontal. Hecho esto, lo único que se tenía que hacer era un copiado de la fila del lema trabajado y “trasponer” sus elementos de una fila horizontal a una columna vertical, en este caso en la columna subsiguiente a la primera del documento llamado *Transposición*. Esta operación se realizó utilizando la opción “Trasponer” de la ventana desplegada por la opción de “Pegado especial” del menú “Edición”, en la barra del menú principal de Excel. Después de crear este documento con todos los datos del lema al que le correspondían los ejemplos de una ficha, se copiaba como encabezado de la misma. Los cambios hechos sobre el documento *Transposición* no se guardaban para poder seguir utilizándolo como una plantilla para nuevas fichas. Las fichas se iban salvando con un nombre conformado por las siglas del nombre del diccionario al

que pertenecieran, un guión bajo y el número progresivo, a cuatro cifras, del lema correspondiente en la lista general del diccionario. La nominación resultante dio ejemplos como los de las fichas *DEUM_0004*, *DRAE_0001*, *DBM_0003*, *AHDOM_0113*, entre otras tantas. Finalmente, resulta pertinente comentar un par de casos especiales que se presentaron durante la creación de las fichas y que requirieron nombrarlas de forma distintiva.

La primera situación se mencionó con anterioridad y tiene que ver con el hecho de que, como el muestreo se llevó a cabo de manera estratificada, era común que un lema se presentara en varios diccionarios con el mismo significado. El problema más notorio derivado de esta situación era una posible documentación múltiple que inflara los resultados de la representatividad del corpus. Además, este trabajo repetitivo implicaba una pérdida de tiempo y esfuerzo. Para evitar esta situación, como ya comentamos, se crearon fichas compartidas entre dos o más diccionarios. Las fichas resultantes fueron nombradas igual que las convencionales, pero se les añadió un segundo guión bajo seguido de la partícula “comp”, tal como se puede apreciar en las fichas *DEUM_0117_comp*, *DRAE_0008_comp*, *DBM_0021_comp*, entre otras. Así pues, si tomamos como ejemplo a las fichas *DRAE_0039_comp*, *DBM_0021_comp*, y *DTV_0499_comp*, comentadas al presentar la *lista fusionada sample* en la Figura 3.6, podemos ver que estas tres fichas comparten los mismos ejemplos. Esto se debe a que estas fichas se refieren al lema común *chafirete*, el cual tiene un significado aproximado en las tres fichas, “mal chofer, mal conductor de vehículo automóvil” en el *DBM*, “chófer”, en el *DRAE*, y “conductor de taxi, camión carguero o autobús” en el *DTV*. Los ejemplos de uso hallados en el *CREA* fueron un total de tres, de los cuales se copiaron todos, si bien tan sólo se reportaron dos. Estos ejemplos, en una versión condensada son:

(a) COZUMEL, Quintana Roo, 31 de agosto.- El veterano Efraín Payán disparó tres incogibles, incluso un cuadrangular, para encabezar el ataque con que los Rockies vencieron 4-3 a los Taxistas, para tomar ventaja en la serie final de la Liga Municipal Asterio Tejero.

Jonrón de Payán, ex jugador de los Leones de Yucatán en la Liga Mexicana, puso en ventaja a los Rockies en la primera entrada, pero jit de Alfonso Martín empató la pizarra.

El propio Martín, con sencillo en la sexta entrada, puso el pizarrón 3-1 a favor de los **chafiret**es, pero en la novena los Rockies atacaron con tres para llevarse el valioso triunfo.

AÑO: 1996

AUTOR: PRENSA

TÍTULO: Diario de Yucatán, 01/09/1996 : Truena el fusil de Payán en Cozumel

PAÍS: MÉXICO

TEMA: 05.Deportes

(b) Llegamos a La Perla Negra, y apenas entramos mi papá me murmuró al oído: "está igualito a como lo dejé". En la pista había un conjunto de música tropical que se parecía al Combo de Lobo y Melón; frente a él, dos rumberas que bien hubieran podido ser las Dolly Sisters, se tropezaban en lo redondo de una pequeña pista; un mulato de cabello afro cantaba acompañado del resto del combo: "yo soy el ruletero (que sí, señor, el ruletero). Yo soy el **chafiret**e (que sí, señor, el chafirete). Yo soy el macalacachimba (que sí, señor, el macalacachimba). Yo soy el icuiricui (que sí, señor, el icuiricui)".

AÑO: 1985

AUTOR: Alatraste, Sealtiel

TÍTULO: Por vivir en quinto patio

PAÍS: MÉXICO

TEMA: 07.Novela

(c) Cosa rara: Cuca, la cuñada del Cachorro, no le respondió el saludo ni se detuvo, pasó de largo metiéndose a su casa, el nueve. Pero más tardó en entrar, que en salir disparada por un cachetadón del cuñadito. Maximina intervino en su defensa, para impedir que su marido le rompiera la crisma.

- Oye, si porque nos mantienes crees tener derecho de tranquear a mi hermana, te equivocas.

- Ningún mantienes, bien que me saca raja, ¿no me pone a que les ayude en todo?

Él, dirigiéndose a su esposa a modo de justificación:

- No vino a dormir, es una puta.

- Eso quisieras, güey. Me pegas porque soy mujer. Ganas me dan de traerte a mi novio para que te rompa...

- Anda, tráemelo. Me gustaría medirme con ese **chafiret**e de la mudanza foránea, te apuesto a que se la pasó con él. Claro, con ese camión cerrado cualquiera no. Si trae su hotel sobre ruedas.

AÑO: 1993

AUTOR: Hayen, Jenny E.

TÍTULO: Por la calle de los anhelos

PAÍS: MÉXICO

TEMA: 07.Novela

Ahora bien, el conteo de estas fichas con ejemplificación compartida se tuvo que tomar en cuenta para hacer las inferencias correspondientes acerca de los posibles resultados finales del comportamiento de la población. Es decir, por un lado, si se generaron un cierto número de fichas compartidas, había que considerar que el número de lemas ejemplificados será inferior al total del número de fichas obtenidas. Esto se tomó en cuenta en el momento que se hizo el análisis de los datos. Por otro lado, las fichas comunes se tuvieron que incluir de alguna manera en el conteo final de la representatividad del corpus para no minar los resultados en este respecto.

El segundo problema que se presentó al crear las fichas tiene que ver con la diferenciación entre lemas y acepciones en los distintos diccionarios fuente. Anteriormente se llamó la atención sobre el hecho de que del *DIME*, que fue el último diccionario del que se extrajeron lemas, se creó una lista más bien de acepciones y no de lemas. Todo esto, como se explicó, fue hecho al momento de percatarse de las deficiencias en la lematización de varios diccionarios. Sin embargo, esta creación de listas de acepciones y no de lemas no se hizo desde un principio. Por ello, a la hora de construir las fichas resultó necesario dar cuenta del hecho de que algunas de ellas, que eran el resultado de trabajar un solo lema, ejemplificaban más de una acepción. En consideración a esto se decidió afectar también el nombre de las fichas que presentaran esta ejemplificación múltiple para su posterior contabilización, pues ignorar estas fichas implicaría subestimar la representatividad del corpus. La nominación resultante dio ejemplos como el de la ficha *DEUM_0510_triple*, el de la ficha *DBM_0018_doble*, entre otros. Un ejemplo de una ficha con más de una acepción, y con varios ejemplos para dichas acepciones, es precisamente el de la ficha *DBM_0018_doble*, donde el lema *carajo* presenta dos significados distintos (“interj. [interjección] que denota gran enfado o disgusto. | **del carajo**. 1. loc. adj. [locución

adjetiva] Malo, difícil, complicado) ejemplificados de manera múltiple. En el caso del primer significado se encontraron varios ejemplos, dos de los cuales son:

(a) Dice vehemente que "si en 1988 nos hubiéramos decidido por una alianza estratégica, por una defensa del voto mancomunada; si no hubiéramos lanzado aquel llamado a la legitimidad, me pregunto yo si a partir de ahí no hubiera cambiado el destino del país. Nos hubiéramos evitado el paso de seis años por la historia de México de (Carlos) Salinas. ¡No hubiera estado en nuestra historia Salinas, **carajo**! Y ese solo hecho hubiera cambiado muchas de las cosas que hoy son tan dolorosas en este país".

Luego de recordar que a lo largo de la historia del PAN ha habido corrientes a favor de la participación en alianzas electorales con otras fuerzas, Fox dice no entender por qué Carlos Castillo Peraza "cambió repentinamente" su posición al respecto y hoy se opone a las alianzas.

AÑO: 1997

AUTOR: PRENSA

TÍTULO: Proceso, 19/01/1997: Fox insiste en la alianza por la Cámara "para echar a esos barbajanes de Los P...

PAÍS: MÉXICO

TEMA: 03.Política

(b) En lo que se refiere a su pretensión de que nunca soborna a los periodistas porque no asiste a banquetes y francachelas con ellos, parece un poco contradictorio cuando párrafos abajo leemos: "Está bien ir a comer, emborracharte con ellos en una feria, prestarles un carro para que se vayan si están muy borrachos y conseguirles boletos para el palenque; eso no es deshonestidad, no estoy comprando a nadie con eso. No es cierto que haya corrupción, **carajo**".

Agregaré a lo anterior que el mismo Herrerías me confesó en una ocasión, cuando decía que éramos amigos (?), que a los periodistas de un diario fotograbado les seguiría dando su sueldo usual.

AÑO: 1996

AUTOR: PRENSA

TÍTULO: Proceso, 15/12/1996 : De Enrique Guarnier

PAÍS: MÉXICO

TEMA: 05.Tauromaquia

Dos ejemplos del segundo significado de este mismo lema son:

(a) CARGADOR 2 Pero que ya sigue mejor, me dijo el Guacho.

Página 63

CARGADOR-JEFE Quién sabe. Lo que sí es que yo no sé de qué la va a hacer el pobre cuate.

CARGADOR 2 De albañil, ésa era su chamba.

CARGADOR-JEFE Ya ni para eso. (Transición.) Bueno, pues ese mismo día, todavía luego del accidente, se armó un escándalo con el pinche dueño de la casa.

Empistolado el cabrón... A plumazos nos quería enfriar a todos el hijo de su rechingada. A ver nomás.

CARGADOR Bueno, ¿y por qué?

CARGADOR-JEFE Por sus pinches güevos... No, si te digo que fue un sábado **del carajo**.

Cargador-jefe se interrumpe. Jorge y Sara están entrando. Jorge va adelante, llevando libros que coloca junto a los que trajo anteriormente. Detrás, Sara llevando la base de una lámpara y algunos otros objetos.

AÑO: 1979

AUTOR: Leñero, Vicente

TÍTULO: La mudanza

PAÍS: MÉXICO

TEMA: 07.Teatro

(b) No Sergio, no es por ahí, el Mauro. Si quieren conseguir la base, eso cuesta, y no dinero, ciertamente, sino esfuerzo, sacrificio. Te mandé llamar porque, aunque no lo creas, me preocupa su situación. Podemos llegar a un arreglo, si quieres, tú y yo. Te conseguimos una buena liquidación y te vas por donde viniste sin problema alguno. Claro, yo entiendo que la situación está **del carajo** y que cada quien quisiera tener seguridad para su familia. Uno lucha por lo mejor para ella. El asunto es que no siempre es posible lograr, de momento, lo que uno pretende. Pongo por ejemplo mi caso. Ve esa fotografía. Quizá no te diga nada, pero para mí es conmovedor, francamente conmovedor ver desde dónde empezamos.

AÑO: 1990

AUTOR: Montaña Hurtado, Alfredo

TÍTULO: Las cenizas de los sueños

PAÍS: MÉXICO

TEMA: 07.Relatos

Así pues, debido a la ejemplificación múltiple que se presentó en algunos lemas, el resultado de la exploración del corpus, que se reportará en la siguiente sección de este trabajo (ver sección 4.01), se hará en acepciones y no en lemas, ya que para estimar el número total de estos últimos se necesitaría hacer una relematización de los materiales pertenecientes a las fuentes secundarias. Sin embargo, esta relematización, que sería parte necesaria del trabajo de elaboración del diccionario, rebasa los alcances del problema de investigación que se ha presentado aquí.

4 Análisis de los datos

4.01 Representación de las listas del muestreo en el corpus

Para hacer un reporte de los resultados de la búsqueda de ejemplos en el *CREA* respecto de las fuentes secundarias, hay que establecer antes cuáles son los datos resultantes de dicha búsqueda que se deberían incluir en el mismo. En este sentido, el dato primordial por reportar se ha considerado que debería ser el total de ejemplos hallados en los documentos del corpus electrónico. Este dato aparece en las dos últimas celdas del extremo inferior derecho de la Tabla 4.1. De igual forma se pensó que, de los ejemplos hallados, se debería decir cuántos de ellos cumplen con la proporción de dos por lema o acepción, y cuántos de ellos son ejemplos individuales, al mismo tiempo que se debería indicar cuántos lemas quedaron sin ejemplificación. Estos datos aparecen en la fila de “Subtotales” bajo el grupo de columnas de “Ejemplificación”. Además, se incluyó en la Tabla 4.1 el reporte de los dos casos especiales de la ejemplificación, el de la ejemplificación compartida y el de la ejemplificación múltiple, recién mencionadas (ver sección 3.10), en los grupos de columnas de “Casos comunes” y de “Casos extras”, respectivamente. El primer grupo de columnas da cuenta de cuántos de los lemas trabajados dieron origen a fichas compartidas entre diccionarios y menciona cuántos ejemplos se encontraron para cada uno de estos lemas. Este grupo de datos, claro está, debe aminorar el total de acepciones ejemplificadas. El grupo de columnas que aparece bajo el título de “Casos extras”, por su parte, muestra la cantidad de fichas que contienen ejemplos para varias acepciones de un mismo lema. Igualmente, aquí se ha especificado cuántos ejemplos se localizaron por lema. Reportando de forma separada la información de estos tres grupos de columnas (“Ejemplificación”,

“Casos comunes” y “Casos extras”), se puede intentar dar un resultado más certero acerca del total de acepciones ejemplificadas obtenidas por diccionario. Para estimar este número, se tuvo que hacer una suma por diccionario de los lemas ejemplificados de forma doble y de forma simple, restarle al total la cantidad de lemas con acepciones en común con otros diccionarios, y añadirle los casos de acepciones extra por lema. El resultado de esta operación se muestra en la columna “Acepciones”. Para estimar el total de ejemplos diferentes obtenidos, se debe hacer un cálculo matemático similar. De manera independiente para los ejemplos dobles y para los ejemplos simples obtenidos para un diccionario en concreto, se debe tomar el resultado global de su ejemplificación, restarle los casos comunes y sumarle los casos extras. Los dos datos resultantes aparecen en las columnas “ej. dobles” y “ej. simples”. Todo esto dentro del grupo de columnas “Resultados finales”. Una vez realizados los cálculos comentados, el total de ejemplos distintos obtenidos para toda la muestra, contenido en las dos últimas celdas del extremo inferior derecho de la Tabla 4.1 antes mencionadas, se puede deducir de la suma de ejemplos dobles diferentes por diccionario, su multiplicación por dos, y su adición de la suma de ejemplos simples por diccionario. Cabe aclarar que para no dejar fuera del conteo de los resultados los casos comunes con fichas compartidas entre distintos diccionarios, éstos se restaron completamente de los diccionarios en que aparecían y se contabilizaron aparte. Además, estos casos requirieron ser revisados uno por uno para determinar cuántas acepciones representaban exactamente. Todo esto fue hecho y reportado en la antepenúltima fila de la tabla, de manera que no se afectaran los totales de ningún diccionario en concreto.

Tabla 4.1. Resultados de los ejemplos obtenidos en el CREA

Lista	búsqueda	Ejemplificación			Casos comunes			Casos extras			Resultados finales			
		nula	doble	simple	total	dobles	simples	total	dobles	simples	acepciones	ej. dobles	ej. simples	
AHDOM 1	109	74	21	14	2	1	1	1	1	-	34	21	13	
AHDOM 2	138	89	34	15	8	8	-	1	-	1	42	26	16	
AHDOM 3	123	82	27	14	6	4	3	-	-	-	35	23	11	
Total AHDOM	370	245	82	43	16	13	4	2	1	1	111	70	40	
DBM	103	29	64	10	29	27	2	20	14	6	65	51	14	
DEUM	226	89	108	29	23	20	3	9	3	6	123	91	32	
DIME	163	28	122	13	19	17	2	-	-	-	116	105	11	
DRAE	132	50	66	16	30	29	1	-	-	-	52	37	15	
DTV	321	206	92	23	22	21	1	7	1	6	100	72	28	
PDLPM	32	17	12	3	7	7	-	5	4	1	13	9	4	
Subtotales	1347	664	546	137	146	134	13	43	23	20	580	435	144	
											Acepciones comunes	65	59	6
											Totales	645	494	150
											Total de ejemplos	1138		

Como se puede ver en la Tabla 4.1, del *AHDOM*, por mencionar un caso, se obtuvieron un total de 82 lemas con ejemplos dobles y 43 lemas con ejemplos simples, de entre los 370 casos trabajados como mínimo para la muestra representativa de este diccionario. Como consecuencia de estos resultados, 245 entradas del *AHDOM* quedaron sin ejemplificación. Además, se presentaron dieciséis casos compartidos con otros diccionarios, trece de los cuales dieron origen a fichas compartidas con ejemplos dobles y

cuatro a fichas con ejemplos simples. También se dieron dos casos de entradas con acepciones múltiples que produjeron dos acepciones extras, una con una ejemplificación doble y otra con una ejemplificación simple. Por todo esto, podemos ver que del *AHDOM* se pudo ejemplificar un mínimo de 111 acepciones distintas y se pudieron obtener un mínimo de setenta ejemplos dobles y cuarenta ejemplos simples. Una lectura similar se puede hacer en la Tabla 4.1 respecto de cada uno de los diccionarios fuente. Hay que recordar que las acepciones compartidas no han sido agrupadas por diccionario, pues sería imposible contabilizarlas en más de un diccionario al mismo tiempo sin afectar los resultados totales. Por otro lado, estas acepciones compartidas no podrían ser contabilizadas solamente como parte de ciertos diccionarios en concreto, pues eso ignoraría su presencia en otros y afectaría negativamente la representatividad reportada para estos últimos. Finalmente, los totales respecto de toda la muestra son los siguientes. De 1,347 entradas exploradas procedentes de todas las fuentes secundarias, 546 generaron ejemplos dobles y 137 ejemplos simples. 146 de los lemas ejemplificados con el corpus eran lemas compartidos entre los distintos diccionarios. 133 de estos lemas compartidos presentaban ejemplos dobles y catorce ejemplos simples. Todos estos lemas compartidos correspondían a 65 acepciones comunes, con 59 ejemplos dobles y seis ejemplos simples. Además, se encontraron 43 acepciones extra en el total de lemas ejemplificados, veintitrés de estas acepciones tenían ejemplos dobles y veinte acepciones tenían ejemplos simples. En el total de las fuentes secundarias que tuvieron representación en el *CREA*, se hallaron un mínimo de 645 acepciones distintas, con 494 ejemplos dobles y 150 ejemplos simples. Así, los ejemplos correspondientes a acepciones distintas alcanzaron el número de 1,138 para toda la muestra.

4.02 Limitaciones en el conteo de resultados

Ahora bien, la contabilización que se acaba de presentar respecto de los lemas y acepciones ejemplificadas tiene ciertas limitaciones. Por un lado, la situación de los lemas y acepciones comunes entre diccionarios, me ha puesto en la necesidad de subestimar los resultados acerca de la representatividad del corpus para cada diccionario. Esto se debió al hecho de que la coincidencia de lemas entre las distintas fuentes hizo necesaria la elaboración de fichas compartidas, las cuales debían restarse del total de lemas ejemplificados por diccionario. El descuento de estas fichas del total de lemas ejemplificados de cada diccionario resultaba necesario para no sobrestimar la representatividad del corpus respecto del muestreo global. Sin embargo, al restar las fichas compartidas entre diferentes diccionarios, la resta se hizo simultáneamente en los diferentes diccionarios con fichas comunes. Esto implicó descontar todos los lemas ejemplificados con fichas comunes y no sólo la duplicación que se hizo con estas fichas. La manera en que he intentado subsanar esta resta excesiva en los resultados de la representatividad del corpus ha sido por medio de la creación de un rubro especial para los ejemplos o acepciones comunes. Esto se ha hecho, sobre todo, con la finalidad de no subestimar el total representado por el corpus respecto de toda la muestra, más allá de la estratificación de ésta. Este rubro ha tenido que ser considerado de forma independiente a los subconjuntos léxicos de los diccionarios para obviar la controversia de decidir a qué diccionario se le sumarían los lemas representados y a qué diccionarios se le deberían restar los lemas repetidos. Así pues, la creación de este rubro común permitió evitar una toma de decisiones, relacionadas con la suma y resta de los lemas comunes ejemplificados, que tendrían que basarse en criterios más bien personales y no científicos. Con todo, estoy consciente de que la solución que he tomado no deja de tener sus desventajas pues de

alguna manera pone en cuestionamiento los resultados ofrecidos por la estratificación de la muestra, es decir, la creación del rubro de fichas comunes tiene por fuerza que afectar los resultados de la representatividad de los distintos diccionarios cuando se les considera a éstos de manera individual. A pesar de ello, el rubro de los ejemplos y acepciones comunes me ha permitido acercarme más a los verdaderos resultados obtenidos durante el proceso de ejemplificación. Finalmente, los datos originales, sin la inclusión de este concepto de las fichas comunes, permanecen en la misma Tabla 4.1, en el grupo de columnas de “Ejemplificación”, y permiten ver los resultados antes de haber aplicado la solución parcial de descontar las fichas comunes y agruparlas independientemente.

4.3 Posibilidades de ejemplificación de las fuentes secundarias

Tomando en cuenta los resultados presentados, éstos podrían también ser calculados en porcentajes, tanto para la ejemplificación doble como para la ejemplificación simple de los distintos subconjuntos de la muestra. Hecho esto, se podría aplicar a la población estos porcentajes de ejemplificación obtenidos en la muestra. De esta manera, se puede intentar estimar el comportamiento de la población en caso de que se intentara ejemplificarla toda. Esta operación se puede observar en la Tabla 4.2.

Tabla 4.2. Resultados de la muestra y comportamiento potencial de la población

Lista	% resultantes en la muestra			Población total	Resultados probables	
	Ejemplos dobles	Ejemplos simples	Representatividad del corpus		Ejemplos dobles	Ejemplos simples
AHDOM	18.9	10.8	29.7	9612	1818	1039
DBM	49.5	13.6	63.1	141	70	19
DEUM	40.3	14.2	54.4	568	229	80
DIME	64.4	6.7	71.2	821	529	55
DRAE	28.0	11.4	39.4	202	57	23
DTV	22.4	8.7	31.2	1971	442	172
PDLPM	28.1	12.5	40.6	34	10	4
Ejemplos comunes	4.4	0.45	4.8	13349	585	59
Totales					3739	1453
Total de casos ejemplificados						5192
Total de ejemplos						8931

En el grupo de columnas “% resultantes en la muestra” de la Tabla 4.2, aparecen los porcentajes de ejemplificación doble y ejemplificación simple de la muestra de cada diccionario. En estos porcentajes se han restado los casos de fichas comunes, es decir, de acepciones coincidentes entre distintos diccionarios. Todos estos casos se muestran en la fila de “Ejemplos comunes”, en la columna “Lista” de este mismo grupo de columnas. Los porcentajes mostrados en esta última fila han sido calculados de acuerdo a toda la muestra. En la columna posterior derecha, con el título de “Representatividad del corpus”, se encuentra el total de toda la ejemplificación conseguida por los distintos diccionarios y por

el rubro de acepciones comunes. A continuación, en la columna “Población total”, se muestra el total de elementos que conforman toda la población en sus distintos subconjuntos. En esta columna, en la fila de “Ejemplos comunes”, que se ha calculado respecto del total de la muestra, se ha colocado el total de elementos de toda la muestra. Aplicando los porcentajes mostrados en el grupo de columnas “% resultantes en la muestra”, se han obtenido los “Resultados probables” de ejemplificación mostrados en las dos últimas columnas. Así, en caso de que la población se comportara como lo ha hecho el muestreo aleatorio estratificado obtenido de ella, su exploración total en el corpus podría obtener un total de 3,739 acepciones con ejemplos dobles y 1,453 acepciones con ejemplos simples. Por ende, se podrían preparar un total de 5,192 acepciones con 8,931 ejemplos de uso para la elaboración de un diccionario del léxico subestándar del español en México.

4.4 Discusión

Para finalizar con este trabajo, quiero recordar que, después de haber hecho una búsqueda exhaustiva de un diccionario ejemplificado del español en México, no se encontró ningún resultado. La expectativa de contar con un diccionario de este tipo, planteada en la Relevancia del proyecto (ver sección 1.02), por medio de la elaboración del mismo intentando la documentación completa del lecionario aquí reunido (ver sección 1.03), parece asequible dados los números recién reportados (ver sección 4.3). Sin embargo, la inversión de tiempo, esfuerzo y dinero que implicaría la compleción de esta documentación rebasa los límites de esta tesis. Además, hay que tener presente que para llevar a cabo la elaboración definitiva de un diccionario, que incluyera los ejemplos resultantes de dicha documentación, habría que cubrir varios otros pasos de la labor lexicográfica de gran volumen también (Haensch, 1982a). En cuanto al resultado potencial de un diccionario

ejemplificado, obtenible por los medios aquí usados, su valor bien podría calcularse respecto de la extensión de otros diccionarios equivalentes.

En cuanto a los diccionarios del subestándar que cubren el español de España es posible encontrar varias referencias en las distintas obras de referencia citadas en la Revisión bibliográfica (ver sección 2.03). Por ejemplo, en Haensch (1997) nos encontramos con cuatro diccionarios del subestándar, con inclusión del tabú léxico y de voces jergales, el de León (1995), el de Martín (1979), el de Oliver (1987) y el de Villarín (1979). De estos cuatro, el primero es “una de las mejores obras del género”, a criterio de Haensch (1997, p. 95). Además de estos diccionarios, el mismo Haensch menciona dos diccionarios de insultos, el de Montero (1990) y el de Celdrán (1995), y varios diccionarios de jergas que, por abordar principalmente la variación diastrática jergal aquí obviada, no tienen una relación directa con este trabajo (1997, pp. 96, 98-99). El problema de todas estas referencias es que ninguna de estas obras se puede conseguir en las grandes librerías de la Ciudad de México, mucho menos en Puebla. Además, ninguno de los cuatro diccionarios del subestándar ni ninguno de los dos diccionarios de insultos que Haensch menciona parece ser ejemplificado, al menos no si se toman en consideración los títulos dichos diccionarios. En cuanto a Bajo (2000), otra de las obras de referencia citada en la Revisión bibliográfica (ver sección 2.03), en ella nos encontramos con una extensa bibliografía en la cual la autora incluye una sección donde aparecen diccionarios con marcación diastrática o diafásica (pp. 241-243). En la lista de diccionarios que esta autora proporciona no existe una mayor distinción o clasificación entre las distintas obras. Por tanto, aquí haré mención únicamente de aquellas que parecen estar relacionadas directamente con este trabajo. Para ello, me basaré en los títulos de las obras referenciadas. En primer lugar, hay que decir que esta autora menciona a los mismos seis autores que Haensch referencia, es decir, a Celdrán

(1995), a León (1992) --del que Haensch refiere una edición más nueva (1995)--, a Martín (1979), a Montero (1990), a Oliver (1987) y a Villarín (1979). Además, Bajo da algunos otros títulos, como Carbonell (2000), Criado (1981), Díez (1987), López (1998), Martín Sánchez (1997) y Sanmartín (1998). Así la lista de textos relacionados directamente con este trabajo aumenta de seis a doce. Sin embargo, al igual que sucede con los títulos comentados por Haensch (1997), de los diccionarios agregados por Bajo (2000, pp. 241-243) ninguno es adquirible en México. Además, juzgando una vez más los títulos de las obras, ninguno de dichos títulos hace referencia al hecho de que los diccionarios sean o no ejemplificados. Finalmente, si revisamos el apartado llamado Léxico de argot de la sección Lexicología de González (2002, pp. 146-149), podemos encontrarnos con algunos diccionarios relacionados con este estudio. En este sentido, esta autora refiere, una vez más, siete de los títulos antes mencionados. Alude a Carbonell (2000), a Celdrán (1995), a León (de quien refiere una edición más nueva, 1998), López (1998), a Montero (1990), a Oliver (del que presenta otra edición más reciente, 1991), y a Sanmartín (del que da también una nueva edición, 1999). Al mismo tiempo, González (2002) ofrece dos títulos no referidos ni por Haensch (1997) ni por Bajo (2000), el de Luque, Pamies y Manjón (2000) y el de Zamora (2000), los cuales tampoco circulan en las librerías mexicanas. De cualquier forma, la lista de trabajos referidos por los tres autores de referencia contiene, entonces, catorce diccionarios del subestándar elaborados en España, de cuyos títulos ninguno hace alusión a la ejemplificación de los mismos. Así pues, el único diccionario ejemplificado del subestándar del que he podido tener noticia, a través del subdirector del Instituto de Lexicografía de la RAE (R. Rodríguez Marín, comunicación personal, 8 de octubre de 2003), es el de Ruiz (2001). Con todo, hay que hacer énfasis en el hecho de que de los quince diccionarios del subestándar que he mencionado, incluyendo el referido por

Rodríguez Marín, ninguno es adquirible en las librerías de Ciudad de México, ni en las de Puebla. De hecho, el único diccionario del subestándar del español elaborado en España que he podido encontrar en las librerías de México no está mencionado en los listados de referencia aquí utilizados debido a su muy reciente impresión. Dicho diccionario, hay que aclarar, no es un diccionario ejemplificado. Este diccionario es el de Iglesias (2003). Por todo ello, resulta muy difícil hacer una comparación entre los resultados potenciales de un diccionario como aquel cuya ejemplificación se ha explorado aquí y los varios diccionarios similares que existen respecto del español de España.

De cualquier forma, para estimar qué tan significativo se antoja el resultado potencial de la exploración en el *CREA* del lemario aquí reunido, llevé a cabo el conteo del *Diccionario de argot español (DAE)* de Iglesias (2003). Para esto, me apoyaron dos becarios más que trabajaron alrededor de unas quince horas cada uno para llevar a cabo este conteo y el de dos diccionarios análogos de la lengua inglesa, que presentaré a continuación. Los nombres de estos becarios son Gabriel de Santos, que participó también en las dos etapas de la identificación de lemas pertinentes al subestándar en las fuentes secundarias, y Daniela Sánchez, que participó en la segunda etapa (ver sección 3.06). El resultado del conteo de entradas y acepciones por letra del *DAE* se presenta en la Tabla 4.3. Para entender el conteo presentado en dicha tabla, se debe mencionar que este diccionario cuenta con dos partes. En la primera presenta los lemas propios del subestándar, seguidos de sus definiciones. En la segunda parte aparece una sección, poco ortodoxa en la lexicografía de la variación lingüística de una misma lengua, en donde se incluyen algunos lemas del lenguaje estándar y una serie de equivalentes en el subestándar. Por ello, en esta segunda no hay significados propiamente hablando, sino tan sólo equivalentes. Así, esta sección está hecha más bien al estilo de los diccionarios bilingües (Haensch, 1997, p. 187).

Debido a esto, si bien he incluido el conteo que llevamos a cabo de todo el diccionario, aquí sólo comentaré los resultados de la primera parte del diccionario, presentados en el grupo de columnas “argot-lengua estándar” de la Tabla 4.3. En esta parte del diccionario, donde aparecen los lemas del subestándar seguidos de sus definiciones, el *DAE* ofrece un total de 3,427 entradas o lemas. En estas entradas aparecían algunas que ofrecían más de un significado. Para tales casos contamos las acepciones extra, que sumaron de 2,235. De esta manera, como se puede ver en la columna “Total de acepciones” del grupo de columnas “argot-lengua estándar”, el *DAE* ofrece un total 5,662 acepciones, sin ejemplificación. Este número, comparado con las 5,192 acepciones y los 8,931 ejemplos de uso potencialmente obtenibles de la documentación total del leuario aquí reunido, hacen aparecer dicha documentación como una empresa lexicográfica de dimensiones razonables.

Tabla 4.3. Conteo por letra del leuario del *DAE*

Letra	argot -- lengua estándar			lengua estándar -- argot			Totales por letra	
	entradas	acepciones extra	Total de acepciones	entradas	acepciones extra	Total de acepciones	entradas	acepciones
A	191	83	274	53	2	55	244	329
B	193	97	290	20	5	25	213	315
C	604	447	1051	72	15	87	676	1138
D	98	50	148	46	11	57	144	205
E	148	51	199	42	8	50	190	249
F	135	63	198	17	2	19	152	217
G	160	67	227	16	9	25	176	252
H	57	80	137	27	9	36	84	173
I	25	19	44	27	9	36	52	80
J	92	45	137	5	2	7	97	144
K	11	2	13	-	-	-	11	13
L	130	67	197	13	3	16	143	213
M	314	269	583	38	8	46	352	629
N	55	37	92	14	1	15	69	107
Ñ	5	1	6	-	-	-	5	6
O	33	32	65	11	3	14	44	79
P	517	442	959	55	8	63	572	1022
Q	32	25	57	4	2	6	36	63
R	117	77	194	15	-	15	132	209
S	148	76	224	29	13	42	177	266
T	242	132	374	21	4	25	263	399
U	6	3	9	-	-	-	6	9
V	67	55	122	19	4	23	86	145
W	5	1	6	-	-	-	5	6
Y	15	3	18	1	-	1	16	19
Z	27	11	38	2	-	2	29	40
Total	3427	2235	5662	547	118	665	3974	6327

Ahora bien, para complementar aun más la consideración del resultado potencial obtenible de la documentación del leuario de este trabajo, decidí compararlo con los totales de otros dos diccionarios del subestándar asequibles en México. Estos dos diccionarios, sin embargo, pertenecen a la lengua inglesa. El primero que presentaré es el de Spears (1990),

comparable al aquí propuesto por ser igualmente diatópico en el sentido de que abarca tan sólo el inglés de los Estados Unidos. El *Forbidden American English (FAE)* tiene, como se puede observar en la Tabla 4.4, un total de 1,395 entrada o lemas, y 363 acepciones extra, que suman un total de 1,758 acepciones. Además, el *FAE* cuenta con un total de 2,232 ejemplos. Así pues, comparado con el contenido del *FAE*, el resultado de 5,192 acepciones y 8,931 ejemplos de uso que potencialmente se podría obtener por medio de la documentación del leuario aquí presentado se antoja incluso más atractivo.

Tabla 4.4. Conteo por letra del leuario del *FAE*

Letra	Entradas	Acepciones extra	Total de acepciones	Ejemplos
A	25	16	41	72
B	149	65	214	313
C	119	40	159	215
D	89	28	117	176
E	12	1	13	22
F	117	33	150	214
G	80	16	96	129
H	98	12	110	155
I	12	-	12	20
J	42	4	46	41
K	30	3	33	44
L	27	8	35	50
M	39	4	43	44
N	28	7	35	40
O	12	-	12	16
P	143	44	187	226
Q	6	3	9	10
R	36	1	37	43
S	169	48	217	307
T	74	17	91	8
U	12	-	12	10
V	4	-	4	2
W	62	13	75	69
Y	7	-	7	4
Z	3	-	3	2
Totales	1395	363	1758	2232

Finalmente, quiero comentar el contenido de otro diccionario del inglés, el cual pretende abarcar el subestándar del inglés general, me refiero al *Oxford Dictionary of Modern Slang (ODMS)* de Ayto y Simpson (1992). El conteo de este diccionario, como se puede apreciar en la Tabla 4.5, es mayor dado su extensión geográfica. Este diccionario

presenta un total de 5,124 y 2,502 acepciones extra, que suman un total de 7,626 acepciones. A estas acepciones, el *ODMS* añade un total de 6,093 ejemplos. Tomando en cuenta estos números, se puede concluir que, incluso frente al *ODMS*, el resultado potencial de la documentación total del leuario aquí presentado no resulta deleznable. Este último comentario parece justificable pues si bien el número de acepciones del *ODMS* es mayor a las 5,192 acepciones potencialmente obtenibles en este proyecto, el número de ejemplos de este diccionario de la lengua inglesa resulta igualmente inferior a los 8,931 ejemplos de uso que se estima se podrían generar a través de la exploración en el corpus de todo el leuario aquí reunido.

Tabla 4.5. Conteo por letra del lemario del ODMS

Letra	Entradas	Acepciones extra	Total de acepciones	Total de ejemplos
A	119	40	159	130
B	400	194	594	490
C	288	107	395	376
D	251	131	382	338
E	29	13	42	50
F	203	100	303	265
G	262	145	407	342
H	218	112	330	293
I	37	21	58	37
J	137	65	202	161
K	70	53	123	108
L	146	55	201	175
M	270	95	365	304
N	148	78	226	200
O	94	36	130	67
P	466	212	678	343
Q	27	8	35	21
R	265	137	402	229
S	957	568	1525	1262
T	327	161	488	422
U	35	16	51	45
V	32	6	38	32
W	252	110	362	295
Y	58	19	77	60
Z	33	20	53	48
Totales	5124	2502	7626	6093

Así pues, tomando en consideración la extensión de los diccionarios contabilizados, y recordando la pregunta de investigación acerca de la factibilidad de llevar a cabo la ejemplificación, por medio del *CREA*, de un lemario exhaustivo del léxico subestándar del español en México (ver sección 3.07), podríamos concluir que las expectativas que muestra

esta tesis respecto del resultado de dicha ejemplificación, con un volumen potencial de 5,192 acepciones y 8,931 ejemplos de uso, señalan un trabajo harto viable y digno de llevar a cabo, para aquellos que cuente con el rubro económico para realizar esta empresa. Además, esta posibilidad se ha explorado y se ha corroborado a pesar de que para ello se ha utilizado un corpus electrónico que no ha sido especialmente diseñado para este trabajo. Con esto se confirma lo atinado del comentario de Sinclair acerca de la pertinencia en la utilización de los recursos disponibles en lexicografía, aunque éstos no parezcan idóneos (1985, p. 86). Así pues, como se ha podido demostrar aquí, con las herramientas lexicográficas que actualmente existen en español se puede contribuir a llenar un hueco en la lexicografía hispanoamericana, tradicionalmente pudibunda (Alvar, 2002, pp. 264-265; Haensch, 1984, pp. 344-348), y se puede construir un diccionario que seguramente resultaría de interés para los aprendices del español como segunda lengua con miras a utilizar dicha lengua en México, e incluso para aquellos hablantes del español de otras variedades geográficas que estuvieran interesados por los aspectos del subestándar del español en este país.

5 Bibliografía

- Alvar Ezquerro, M. (Ed.). (1987). *Diccionario general ilustrado de la lengua española*. Barcelona: Vox Bibliograf.
- Alvar Ezquerro, M. (2002). *De antiguos y nuevos diccionarios del español*. Madrid: Arco/Libros.
- Ávila, R., & Aguilar, G. (2003). *Diccionario inicial del español en México*. México: Trillas.
- Ayto, J., & Simpson, J. (1992). *The Oxford Dictionary of Modern Slang*. Gran Bretaña: Oxford University.
- Bajo Pérez, E. (2002). *Los diccionarios: Introducción a la lexicografía del español*. Gijón, España: Trea.
- Barriga Villanueva, R., Martín Butragueño, P., & Parodi, C. (1999). *El español de América: Cuadernos bibliográficos, 9, México*. Madrid: Arco/Libros.
- Barriga Villanueva, R., & Parodi, C. (1998). *La lingüística en México: 1980-1996*. México: El Colegio de México.
- Bell, A. (1997). Language style as audience design. En N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A reader and coursebook*. (pp. 240-250). Chippenham, Inglaterra: Palgrave.
- Biber, D. (1994). An analytical framework for register studies. En D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register*. (pp. 31-56). Nueva York: Oxford University.
- Biber, D., Conrad, S., & Reppen, R. (1998). Lexicography. En *Corpus linguistics: Investigating language structure and use*. (pp. 21-54). Cambridge, Inglaterra: Cambridge University.
- Biber, D., & Finegan, E. (1994). Introduction: Situating register in sociolinguistics. En D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register*. (pp. 3-12). Nueva York: Oxford University.
- Briz, A. (1996). *El español coloquial: Situación y uso*. Madrid: Arco/Libros.
- Carbonell Basset, D. (2000). *Gran diccionario del argot: El Sohez*. Barcelona: Larousse.
- Celdrán, P. (1995). *Inventario general de insultos*. Madrid: Ediciones del Prado.
- Cheshire, J. (1997). Linguistic variation and social function. En N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A reader and coursebook*. (pp. 185-198). Chippenham, Inglaterra: Palgrave.
- Colín Sánchez, G. (1987/2001). *Así habla la delincuencia y otros más...* México: Porrúa.
- Contreras, C., & Barra, L. (1996). Forma y función de la jerga estudiantil en la ciudad de Temuco. *Estudios filológicos*, 31, 177-190.
- Cortés Lara, J. C. (1989). *El vocabulario de los homosexuales*. Tesis de licenciatura no publicada, UNAM, México.
- Criado de Val, M. (1981). *Diccionario del español equívoco*. Madrid: SGEL.
- Díez Jiménez, L. (1987). *Diccionario del español eurogilipuetas*. Barcelona: Planeta.
- Ferguson, C. A. (1994). Dialect, register, and genre: Working assumptions about conventionalization. En D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register*. (pp. 15-30). Nueva York: Oxford University.
- Finegan, E., & Biber, D. (1994). Register and social dialect variation: An integrated

- approach. En D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register*. (pp. 315-347). Nueva York: Oxford University.
- Gay, L.R., & Airasian, P.W. (2002). *Educational research: Competencies for analysis and application* (7^a ed.). Englewood Cliffs, NJ: Prentice Hall.
- Gómez de Silva, G. (2001/2003). *Diccionario breve de mexicanismos*. México: Fondo de cultura económica. También disponible en:
<http://www.academia.org.mx/dbm/principal.htm>
- González Aranda, Y. (2002). *Lexicología y lexicografía del español: Repertorio bibliográfico (1990-2002)*. Almería: Universidad de Almería.
- Gutiérrez Ávila, M. A. (1986). *Léxico del trabajo agrícola en San Nicolás, municipio de Cuajimalpa, Costa Chica de Guerrero*. Tesis de licenciatura no publicada, Escuela Nacional de Antropología e Historia, México.
- Haensch, G. (1982a). Aspectos prácticos de la elaboración de diccionarios. En G. Haensch, L. Wolf, S Ettinger, & R. Werner. *La lexicografía: De la lingüística teórica a la lexicografía práctica*. (pp. 395-534). Madrid: Gredos.
- Haensch, G. (1982b). Tipología de las obras lexicográficas. En G. Haensch, L. Wolf, S Ettinger, & R. Werner. *La lexicografía: De la lingüística teórica a la lexicografía práctica*. (pp. 95-187). Madrid: Gredos.
- Haensch, G. (1984). Miseria y esplendor de la lexicografía hispanoamericana. En *Actas del VII congreso de la Asociación de lingüística y filología de América Latina (ALFAL)*, (pp. 333-370). Santo Domingo.
- Haensch, G. (1997). *Los diccionarios del español en el umbral del siglo XX*. Salamanca, España: Universidad de Salamanca.
- Hernández, L. (1991). Ideología y lenguaje: El léxico de los chavos banda. *Signos*, 5, 49-60.
- Hodge, R. & Kress, G. (1997). Social semiotics, style and ideology. En N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A reader and coursebook*. (pp. 49-54). Chippenham, Inglaterra: Palgrave.
- Holmes, J. (2001). *An Introduction to Sociolinguistics*. Harlow, England: Pearson Educational.
- Hymes, D. (1996). *Ethnography, linguistics, narrative inequality: Towards an understanding of voice*. Taylor & Francis.
- Iglesias, J. M. (2003). *Diccionario de argot español*. Madrid: Alianza.
- Irvine, J. T. (1984). Formality and informality in communicative events. En J. Baugh & J. Sherzer (Eds.), *Language in use: Reading in sociolinguistics*. (pp. 211-228). Englewood Cliffs, Nueva Jersey: Prentice-Hall.
- Jockey, S. (2000). *Electronic texts in the humanities*. Oxford: Oxford University.
- Jiménez, A. (1999). *Tumbaburro de la picardía mexicana: Diccionario de términos vulgares*. (52^a ed.). México: Diana.
- Kochanski, G. (2003, Diciembre). *Linguistic isolation*. [En red]. Disponible en:
<http://kochanski.org/gpk/misc/2003stateling.html>.
- Labov, W. (1983). *Modelos sociolingüísticos*. (J. M. Marinas Herreras, Trad.). Madrid: Cátedra.
- Labov, W. (1984). Field methods of the project on linguistic change and variation. En J. Baugh & J. Sherzer (Eds.), *Language in use: Reading in sociolinguistics*. (pp. 28-53). Englewood Cliffs, Nueva Jersey: Prentice-Hall.

- Lara, L. F. (1982). *Vocabulario especializado de la publicidad en México*. México: Comisión Nacional para la Defensa del Idioma Español.
- Lara, L. F. (1990). El objeto diccionario. En *Dimensiones de la lexicografía: A propósito del Diccionario del español actual de México*. (pp. 21-38). México: El Colegio de México.
- Lara, L. F. (Ed.). (1996). *Diccionario del español usual en México*. México: El Colegio de México. También disponible en:
<http://mezcal.colmex.mx/Scripts/Dem/principal.htm>
- Lara, L. F. (1997). *Teoría del diccionario monolingüe*. México: El Colegio de México.
- Lara, L. F. (Ed.). (2000). *Diccionario del español usual en México*. (Edición digital). [En red]. Alicante: Biblioteca virtual Miguel de Cervantes. Disponible en:
<http://www.cervantesvirtual.com/servlet/SirveObras/12471730873498395375679/index.htm>
- Lara, L. F. (2002). Los diccionarios contemporáneos del español y la normatividad. En *II Congreso internacional de la lengua española: El español en la sociedad de la información*. Valladolid, España: Instituto Cervantes.
- León, V. (1992). *Diccionario de argot español y lenguaje popular*. (4ª ed. amp.). Madrid: Alianza.
- León Núñez, V. (1995). *Diccionario de argot español y lenguaje popular*. Madrid: Alianza.
- León, V. (1998). *Diccionario de argot español y lenguaje popular*. (13ª ed.). Madrid: Alianza.
- López, A. (1998). *Diccionario de blasfemias, irreverencias y reniegos*. Madrid: Aldebarán.
- Luque, J. D., Pamies, A., & Manjón, F. J. (2000). *El diccionario del insulto*. Barcelona: Península.
- Martín, J. (1979). *Diccionario de expresiones malsonantes del español: Léxico descriptivo*. (2ª ed.). Madrid: Istmo.
- Martín Sánchez, M. (1997). *Diccionario del español coloquial*. Madrid: Tellus.
- Mata Navarrete, N. (1985). *Léxico agrícola de San José de Ramos*. Tesis de licenciatura no publicada, UNAM, México.
- Mey, J. L. (1993/2001). *Pragmatics: An introduction*. Inglaterra: Blackwell.
- Milroy, L., & Gordon, M. (2003). *Sociolinguistics: Method and interpretation*. Oxford, Inglaterra: Blackwell.
- Moliner, M. (1998). *Diccionario de uso del español*. Madrid: Gredos.
- Montero, R. (1990). *Diccionario de nuevos insultos, maledicencias y expresiones soeces*. Madrid: Librerías Sánchez.
- Oliver, J. M. (1987). *Diccionario de argot*. (2ª ed.). Madrid: Nova.
- Oliver, J. M. (1991). *Diccionario de argot*. Madrid: Sena.
- Porto Dapena, J. –A. (2002). *Manual de técnica lexicográfica*. Madrid: Arco libros.
- Pozzi Pardo, M. (en prensa). Vocabulary of nuclear waste/vocabulaire des déchets nucléaires/ vocabulario de desechos nucleares. Toronto.
- Real Academia Española. (2001). *Diccionario de la Lengua Española*. (22ª ed.). Madrid: Espasa Calpe.
- Real Academia Española. (2003). *Diccionario de la Lengua Española (Versión 1.0)* [CD ROM]. (22ª ed.). Madrid: Espasa Calpe.
- Real Academia Española. (2003, octubre). *Corpus de referencia del español actual: Manual de consulta*. Ver. Html 2.0 [En red]. Disponible en: <http://www.rae.es>
- Real Academia Española. (2004, noviembre). *Departamento de banco de datos* [En red].

- Disponible en: <http://www.rae.es>
- Ruiz, C. (2001). *Diccionario ejemplificado de argot*. Barcelona: Península.
- Ryan, E. B., Giles, H., & Sebastian, J. (1982). An integrative perspective for the study of attitudes toward language variation. En E. B. Ryan & H. Giles (Eds.), *Attitudes towards language variation: Social and applied contexts*. (pp. 1-19). Londres: Edward Arnold.
- Sanmartín Sáez, J. (1998). *Diccionario de argot*. Madrid: Espasa Calpe.
- Sinclair, J. (1985). Lexicographic evidence. En Robert Ilson (Ed.). *Dictionaries, lexicography and language learning*. (pp. 81-94). Gran Bretaña: Pergamon.
- Sherzer, J., & Darnell, R. (2000). Guía resumida para el estudio etnográfico del uso del habla (M. L. Escalante Correa & Y. Lastra, Trads.). En Y. Lastra (Ed.), *Estudios de sociolingüística*. (pp. 63-73). México: UNAM. (Trabajo original publicado en 1978).
- Soler Arechalde, M. A. (1985). Glosario. Voces y construcciones regionales y de cultura rural. *Cancionero folklórico de México*. (pp. 55-83). México: El Colegio de México.
- Spears, R. A. (1990). *Forbidden American English: A serious compilation of taboo American English*. Illinois, EUA: Passport Books.
- SPSS for Windows (Versión 10.0.1 Estándar) [Programa de computadora]. (1999). EUA: SPSS.
- Usandizaga, P. M. (1994). *El chingolés: Primer diccionario del lenguaje popular mexicano*. (8ª ed.). México: Costa-Amic.
- Villarín, J. (1979). *Diccionario de argot*. Madrid: Nova.
- Werner, R. (1982). Léxico y teoría general del lenguaje. En G. Haensch, L. Wolf, S. Ettinger, & R. Werner. *La lexicografía: De la lingüística teórica a la lexicografía práctica*. (pp. 21-94). Madrid: Gredos.
- Zamora, D. (2000). *Diccionario caca-chondo del 3er milenio*. Madrid: Nuer.