

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).

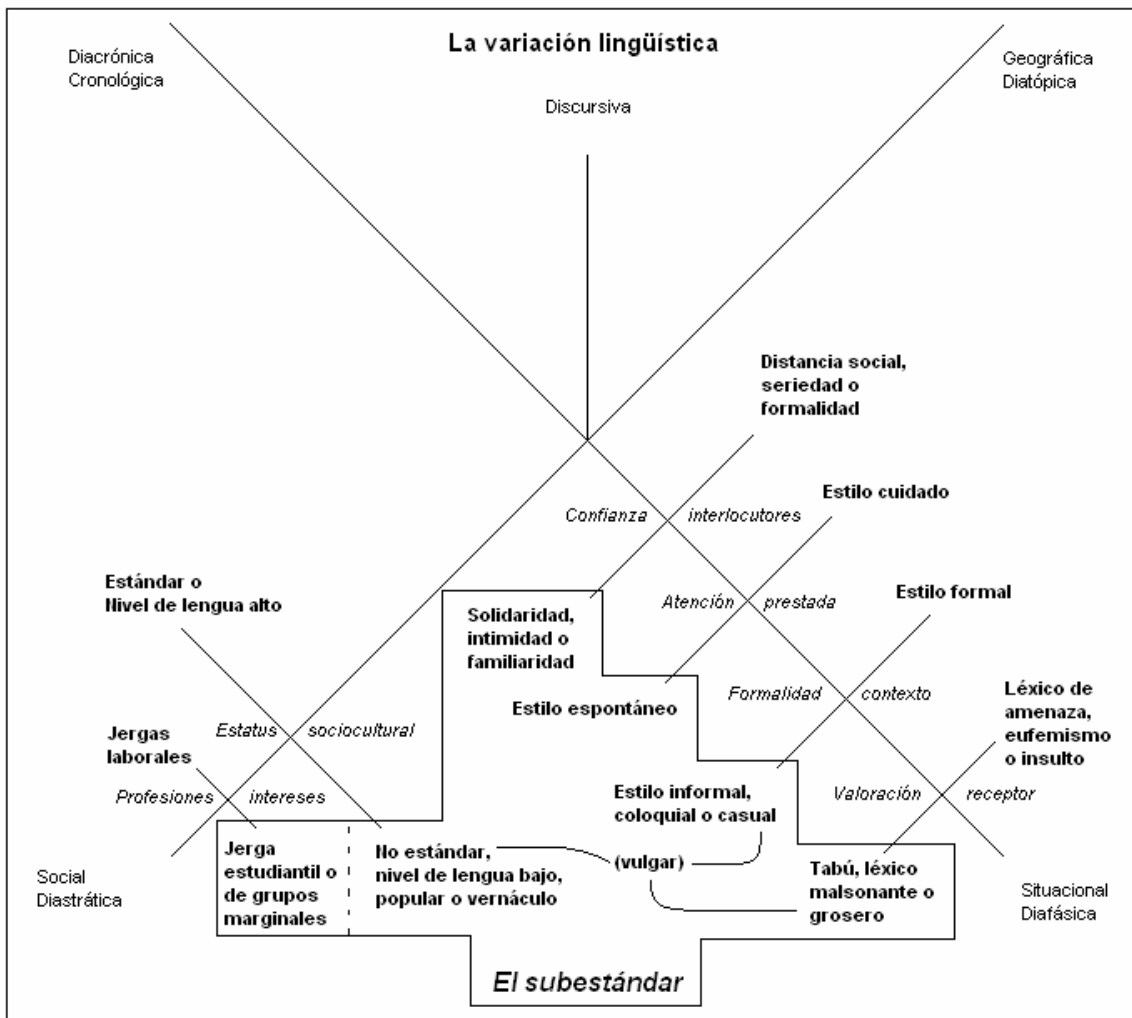
A pesar de los problemas encontrados en el establecimiento de los ejes de categorización de la variación situacional, podemos decir que las variedades lingüísticas situacionales son entidades susceptibles de delimitación. En el caso del estilo informal o coloquial, Briz apunta que en éste se puede encontrar 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 harto 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 lecionario 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.