

COMENTARIOS EN TORNO A TRES RECUENTOS DEL LÉXICO PUERTORRIQUEÑO

Dwight García

Introducción

Acaba de publicarse —casi diez años después de que se realizara la encuesta en que se basa— el *Léxico disponible de Puerto Rico*, de Humberto López Morales (1999), trabajo que se inscribe en la línea de una serie de estudios que, particularmente en las dos últimas décadas, han abordado el análisis del léxico puertorriqueño a partir de criterios cuantitativos, y entre los que se destacan, junto con el *Léxico disponible*, otras dos investigaciones publicadas en forma de libro: la recopilación del *Léxico básico de Puerto Rico*, de Amparo Morales (1986), y la del *Léxico del habla culta de San Juan de Puerto Rico*, coordinada por López Morales (1987).

Aparte del objetivo fundamental de describir el léxico de una comunidad lingüística dada, los tres estudios se proponen expresamente servir de instrumentos en la planificación escolar y, en el caso de los recuentos del léxico culto y del léxico disponible, se señala también la posibilidad de usar sus hallazgos en estudios comparativos posteriores.

A pesar de que trabajan con muestras, materiales y métodos diferentes, las tres investigaciones enfocan una misma parcela de la realidad lingüística, a saber, el léxico de los hablantes que conforman los estratos de mayor cultura. Tanto el estudio del léxico

culto como el del disponible se basaron en informantes con un grado de escolaridad superior a los doce años: en el primero, los/as encuestados/as debían tener por lo menos el grado de bachillerato y, en el otro, la muestra estuvo integrada por estudiantes que cursaban su segundo semestre de estudios universitarios. Morales, por su parte, trabajó con un corpus de obras publicadas, lo que también delimita un léxico conforme con el de los hablantes cultos.

Se trata, por otro lado, de investigaciones complementarias: mientras el estudio de Morales busca describir el llamado *léxico básico*, los otros dos enfocan el vocabulario denominado *disponible*; se entiende que “Ambos exámenes nos dan el verdadero perfil léxico de una comunidad lingüística”, es decir, su *léxico fundamental* (López Morales, 1986).

Frente a los primeros estudios cuantitativos del vocabulario, basados únicamente en la frecuencia con que se usaban las palabras, los análisis del léxico básico toman en cuenta la dispersión de los términos, es decir, la cantidad de “mundos” —temas y tipos de discurso— en que se utilizan. La frecuencia ponderada con la dispersión nos da los índices de uso, que “calcan las frecuencias de uso real de las distintas clases de palabras” (López Morales, 1986). Así, estos inventarios recogen las voces usadas normalmente en cualquier tipo de acto comunicativo, sin importar su tema, como es el caso de las palabras gramaticales (pronombres, preposiciones, etc.) y los verbos de sentido general.

Los estudios del léxico disponible, en cambio, buscan recopilar los vocablos “temáticos”, es decir, los que sólo se utilizan cuando se habla de un asunto determinado, como la comida, la ropa o la transportación. A pesar de que se trata de términos conocidos y usados, no solían aparecer (o alcanzaban una posición muy baja) en los recuentos basados únicamente en la frecuencia, ya que, al estar muy vinculados a un tema, se utilizan sólo en contextos comunicativos específicos. Para obtenerlos, los investigadores trabajan con pruebas asociativas basadas en unos estímulos, llamados “centros de interés” (v.g. Alimentos, Partes del cuerpo, etc.), en torno a los cuales los/as informantes producen las respuestas.

Las tres publicaciones que ahora comentamos tienen un mismo formato, con una introducción seguida de las listas de

vocablos, cada uno acompañado por el número de respuestas o los índices (de uso, dispersión, disponibilidad, etc.) correspondientes. La sobriedad de las ediciones, junto con la presencia de estas cifras y la ausencia de comentarios o análisis fuera de los que aparecen en la introducción, configuran la imagen de objetividad y rigor científico que persiguen, mas no siempre alcanzan, estos proyectos.

2. El *Léxico básico de Puerto Rico*

El recuento de Amparo Morales recoge un listado de 4,456 palabras seleccionadas de una muestra representativa de textos en prosa escritos por puertorriqueños/as y publicados en el país entre 1948 y 1970. Partiendo del modelo elaborado por A. Juilland (1964), la muestra se clasificó en cinco “mundos”, o subconjuntos discursivos (drama, narrativa, ensayo, literatura técnica y periodismo). El vocabulario se obtuvo mediante la aplicación de fórmulas de frecuencia, dispersión y uso a la muestra.

El trabajo, magno y extremadamente riguroso, representa la adopción —y en algunos casos, un refinamiento— de los criterios y métodos que en la actualidad propician la obtención de información fidedigna. En los diversos aspectos de la investigación, se evaluaron múltiples alternativas críticas y metodológicas, y en casi todos se utilizaron las de mayor coherencia y valor explicativo. Así, por ejemplo, en la segmentación de las unidades léxicas —la determinación de lo que constituye un vocablo diferente— se optó por trabajar con “listas lematizadas”, es decir, las que incluyen en el cómputo principal de la unidad todas sus formas flexivas. De este modo, los índices de un vocablo como diminuto incluyen los de *diminuto*, *diminuta*, *diminutos* y *diminutas*. También se observa gran coherencia en el tratamiento de los homógrafos gramaticales, separados según la clase a la que corresponden (v.g., vacío n y vacío aj). Sin embargo, y como la misma autora reconoce, el procedimiento empleado en los casos de voces polisémicas es menos consistente, y así, mientras algunas figuran por separado (capital n, núcleo urbano y capital n, acumulación de dinero), muchas otras aparecen con una sola entrada (*ascendencia*, *establecimiento*, *patrón*, *pasaje*, *parroquiano*, etc.).

La muestra además satisface plenamente el criterio de sincronía, ya que los textos, como señalamos, representan un período definido. Por otro lado, aunque la investigación original también cumplía con el requisito de que el período “fuera lo más actual posible”, la publicación tardía del libro, casi una década después de realizada la investigación, desvirtúa ese propósito, tan importante en vista del carácter cambiante del vocabulario. (De hecho, aunque el léxico básico recoge palabras bastante estables en términos diacrónicos, es posible que algunas de las incluidas en el estudio, como *parroquiano*, estén en vías de convertirse en anacronismos.)

Para satisfacer el criterio, igualmente importante, de exhaustividad, se utilizaron varias fuentes bibliográficas que recogen prácticamente todas las obras publicadas en el país durante el período. De cada mundo se seleccionó un número de obras proporcional al de unidades de ese conjunto en la muestra. Las palabras se escogieron luego mediante procesos aleatorios, hasta obtener 100,000 por mundo léxico.

Sin embargo, la muestra no es tan completa (y representativa de los hechos lingüísticos) como hubiera sido posible, ya que los criterios de selección tienden a privilegiar la lengua culta y formal, y los estilos propios de la escritura.

A esto contribuye, para empezar, el que se excluyeran, sin justificación aparente, las obras “que utilizaban lengua dialectal”, con lo que quedan fuera del recuento los regionalismos de uso popular. También se excluyeron “todos los escritos que no hubieran llegado a la imprenta”, tales como composiciones escolares, cartas, etc. A partir de este criterio, el inventario nuevamente recorta un léxico que tiende a equipararse con el de los hablantes cultos y los contextos de mayor formalidad. De hecho, aunque la autora señala que, a diferencia de Juilland, en la elección de las obras de la muestra no utilizó criterios cualitativos, éstos informan los límites del conjunto, ya que las obras publicadas por lo general han pasado por procesos previos de selección (basados precisamente en juicios valorativos) e incluso de edición.

Pero lo que más atenta contra el criterio de exhaustividad de la muestra es la exclusión de todas las manifestaciones del discurso oral, en el que los vocablos de uso familiar, y sobretodo vulgar, suelen aparecer con más frecuencia que en los textos escritos (y particularmente los publicados). Con esta decisión también quedan

fuera (o alcanzan unos índices que no corresponden con su uso real), las formas características del estilo conversacional; de ahí, por ejemplo, que los pronombres tú y usted no figuren entre las cien primeras palabras del *Léxico*, y que los verbos de segunda persona consistentemente alcancen una frecuencia muy baja.

En este contexto, llama la atención el señalamiento de la autora de que “parece haber acuerdo general en asegurar que para los conteos léxicos del sistema, la lengua escrita es la que ofrece más posibilidad de tratamiento adecuado y riguroso, puesto que es la única que ofrece, a priori, unas categorías generales (drama, ensayo, narrativa, etc.), fáciles de identificar y con suficiente autonomía entre ellas” y que “La lengua escrita ofrece la posibilidad de clasificar los textos, ya que permite incluir cada uno de ellos en un determinado género y, por otro lado, facilita la labor de recolectar un universo completo y uniforme”. Resulta cuando menos controvertible la noción de que los textos escritos son los únicos que ofrecen de antemano categorías generales y relativamente autónomas: como se sabe, todo discurso se produce a partir de esquemas en mayor o menor medida convencionales. (Otra cosa es que los textos escritos estén ya clasificados en los recuentos bibliográficos, y aún otra el que los géneros de los textos escritos correspondan con los mundos delimitados por Juilland.) Así mismo, y sin restarles importancia a los criterios de tipo pragmático, consideramos preocupante el que, en su argumentación, Morales subordine la capacidad analítica o descriptiva de un instrumento, a la facilidad de su uso en la investigación.

Por otro lado, la inclusión de obras de teatro y narraciones en la muestra parece obedecer a un deseo de recoger los usos orales, de otro modo excluidos. Sin embargo, es bien sabido que en la literatura el discurso oral se encuentra muy mediatizado, por lo que no constituye una fuente confiable de usos. Además, aun si lo fuera, y si todos los textos dramáticos y narrativos seleccionados representaran el lenguaje oral, estos géneros constituyen sólo el 40% de la muestra, cifra que no refleja la hegemonía de los mensajes orales en el conjunto de las actuaciones lingüísticas.

Si bien es cierto que las formas de uso exclusivamente conversacional, familiar o vulgar no satisfacen plenamente el criterio de dispersión con el que se delimita el léxico básico, lo mismo podría decirse de las que sólo se utilizan en la lengua escrita,

formal y culta, como es el caso de la conjugación *mas*, que sí aparece en el estudio. Además, no se trata únicamente de un problema de exclusión de voces, sino también de la jerarquización de las que se incluyen: como vimos, en el recuento las formas de segunda persona alcanzan unos índices relativamente bajos, debido a su vinculación con el discurso coloquial.

En todo caso, la exclusión de los vocablos de uso familiar y vulgar sólo sería aceptable si el propósito de este recuento fuera la descripción no ya del léxico (y ni siquiera de la norma, que se determina a partir de los hábitos generales) sino de la norma culta, es decir, de los usos que corresponden a personas y contextos de mayor categoría social. Sin embargo, en la Introducción no se señala que ése sea el caso; al contrario, se indica que el estudio recoge las palabras “que a través de métodos estadísticos han resultado ser las básicas del español de Puerto Rico”, e incluso se afirma que “cualquier criterio de selección que se haga intervenir en el momento de constituir las listas básicas no hace sino filtrar prejuicios subjetivos que deben quedar completamente fuera de estos trabajos”. La contradicción entre este afán científicista y la exclusión de las voces familiares y vulgares de seguro obedece al doble objetivo, descriptivo y pedagógico, de este tipo de estudios: históricamente, los inventarios léxicos se han realizado con fines educativos, y a partir del criterio de que el habla culta es la que provee los modelos lingüísticos que deben enseñarse en la escuela. En este caso, se debió explicar que lo que se describe en esta investigación es el léxico culto, y no el vocabulario básico de la comunidad.

Tampoco parece estar bien definido el criterio de homogeneidad, a partir del cual “se eliminaron del conjunto básico las obras que representaban lengua estereotipada o artificial como lo son las obras de poesía o literatura infantil, devocionarios y textos gramaticales”. A nuestro entender, es muy cuestionable esta distinción entre textos estereotipados o artificiales y los que no lo son, ya que, en mayor o menor medida, todos los mensajes contienen estos aspectos; tampoco es aceptable el que, a partir de este criterio, se excluya “la poesía” —como si toda fuera igual— a la vez que se incluyen obras dramáticas y narrativas, que pueden ser igualmente “artificiales”.

Por último, aun si aceptáramos que la muestra se limitara a textos publicados, no resulta del todo justificada la utilización

de los cinco “mundos” de Juilland, particularmente en vista de que esta clasificación no se basa en criterios uniformes: mientras que las obras narrativas, dramáticas y ensayísticas se definen fundamentalmente a partir de su estructura, los límites de la literatura técnica se establecen a partir de la conjunción de criterios formales y temáticos, y los del periodismo (que puede asumir diversas formas y abarcar multiplicidad de asuntos), de acuerdo con el modo de circulación de los mensajes. Esto podría afectar la deseada diversidad de la muestra, ya que textos de un campo podrían representar el mismo tipo de estructura y tratar los mismos temas que textos de otro. También atenta contra el criterio de diversidad la selección de un número uniforme de 100,000 palabras por mundo, sin considerar la cantidad proporcional de vocablos de cada subconjunto en la muestra.

Estos son algunos de los problemas que investigaciones futuras deben atender. Con todo, y por lo previamente señalado, el *Léxico básico de Puerto Rico* constituye una aportación notabilísima a la descripción de nuestro vocabulario, y a los estudios lexicales en general.

3. Emporio celestial de los conocimientos benévolos

Desgraciadamente, no podemos decir lo mismo del *Léxico del habla culta de San Juan de Puerto Rico*, libro que reúne los recuentos léxicos de cuatro tesis de maestría presentadas en el Recinto de Río Piedras de la Universidad de Puerto Rico por los/as estudiantes Isabel Huyke Freiría, María Pérez del Valle, Héctor Rubén Cardona Santana y Zaír Antonio Rodríguez, y dirigidas por las profesoras Amparo Morales y María Vaquero de Ramírez, y por el profesor López Morales, quien además coordinó el estudio y la publicación de los materiales.

Cada investigación se realizó con una muestra de 12 sujetos —estratificados de acuerdo con el sexo y la edad— que satisficieron, entre otros, los criterios de vivir y haber nacido en San Juan, y, como antes señalamos, de haber completado por lo menos un bachillerato.

Las encuestas se realizaron a partir de un cuestionario, preparado por M. Alvar y L. Flórez (1971), que constituye el tercer

volumen, correspondiente al *Léxico*, del *Cuestionario para el estudio coordinado de la norma lingüística culta de Iberoamérica y de la Península Ibérica*. El instrumento, como indica su título, se utilizaría en un abarcador estudio comparativo de los dialectos del español. Los/as investigadores/as puertorriqueños/as complementaron el cuestionario (que por lo general sólo incluía listas de conceptos agrupados en torno a campos o centros temáticos) con los estímulos verbales o pictográficos necesarios para conseguir las respuestas.

Los 21 campos del cuestionario se derivan de los centros de interés utilizados en los trabajos seminales sobre la disponibilidad léxica realizados por Gougenheim y sus colaboradores (1964), Mackey (1971) y otros. Pero, a diferencia de estas investigaciones, en las que sólo se presentaban centros relativamente amplios ante los cuales los/as encuestados/as producían series de palabras, el cuestionario incluye bajo cada campo listas detalladas de conceptos, para cada uno de los cuales se debe ofrecer una respuesta. El efecto de este modelo es el establecimiento *a priori* de los límites de cada centro de interés, y del léxico que a él corresponde, cosa que, como veremos, mina su objetivo analítico y descriptivo.

Se trata, por otro lado, de un instrumento repleto de “ambigüedades, redundancias y deficiencias”, de inconsistencias tales que nos recuerdan las que un personaje de Borges atribuye a cierta enciclopedia china titulada *Emporio celestial de los conocimientos benévolos*, en cuyas “remotas páginas está escrito que los animales se dividen en a) pertenecientes al Emperador, b) embalsamados, c) amaestrados, d) lechones, e) sirenas, f) fabulosos, g) perros sueltos, h) incluidos en esta clasificación, i) que se agitan como locos, j) innumerables, k) dibujados con un pincel finísimo de pelo de camello, l) etcétera, m) que acaban de romper el jarrón, n) que de lejos parecen moscas”.

Para empezar, resulta problemático –pues dificulta la consulta de estos materiales– el uso arbitrario de títulos y subtítulos para delimitar los campos y subcampos. Así, por ejemplo, bajo el encabezamiento de *Los hijos* figura una lista que comienza con *primogénito* y aparentemente se extiende hasta *comadre*, incluyendo conceptos como *sobrino*, *primo*, *consuegro*, *cuñado*, etc., que claramente no se ajustan a ese centro de interés, sino al del parentesco, que, sin el título correspondiente, le sigue. Bajo *El arroz*

figuran no sólo la *paella* y el *arroz blanco*, sino también la *ensalada* y la *ensalada rusa*. Y, en aparente previsión del moderno follón por las carnes blancas sintéticas, se incluyen bajo *Carne de pavo* no sólo, y como era de esperarse, el *pavo relleno*, sino también la *carne asada*, la *frita*, la *cocida*, la *rellena*, la *mechada*, el *filete*, el *bistec*, las *fiambres*, los *embutidos*, el *salchichón*, el *salame*, la *mortadella*, el *jamón*, el *tocino*, el *tocino frito*, el *tocino de pernil* (seguramente se trataba de un pavo muy grande), la *tocineta con veta*, la *salchicha*, la *morcilla*, el *chorizo*, el *queso de cabeza*, la *longaniza* y la *butifarra*. Del mismo modo, los títulos para las secciones correspondientes a las comidas se limitan a dos, *El desayuno* (bajo el cual se incluyen 16 conceptos) y *El almuerzo* (cuyo campo parece comprender 267 de las 298 entradas de *La alimentación*; nos imaginamos que, tras tan pantagruélico festín temprano en el día, los diseñadores de la lista optaron por no cenar).

También entorpece la consulta de los materiales la inclusión de algunos conceptos en campos a los que no corresponden. El subcampo del tabaco (que incluye conceptos como *cigarrillos*, *cigarros*, *colilla*, etc.) figura bajo *La casa*; en el apartado de *Los postres* se incluyen, junto con otras frutas, la *toronja*, el *aguacate* y —para el gusto se hicieron los sabores— la *fruta verde*, la *pasada* y la *podrida* (todos ellos conceptos que debieron figurar bajo el campo de *Las frutas*, que aparece por separado), así como el *café* y el *café instantáneo* (que debieron estar incluidos en el de *El café*, que se encuentra más adelante, y el cual debió a su vez estar subordinado al de *Las bebidas*, que le precede.)

Estas agrupaciones, por otro lado, apuntan hacia una falta de criterios consistentes en la constitución de los campos. Mientras que muchos obedecen a la lógica de clases (v.g. *Postres-manzana*), en la delimitación y el ordenamiento de otros se privilegian relaciones como la de contigüidad espacial accidental (*La casa-cigarrillos*), o incluso la secuencia temporal (*Postres-El café*). A estas relaciones habría que añadir otras, como la de agente-instrumento (*fabricante de zapatos-cuchillos*, *yunque*, *horma*); la de objeto-material de que está hecho (*La casa-Madera de construcción*); y la de parte-todo (*La casa-Partes de la casa*).

Esta heterogeneidad parece responder —como ha sugerido, al referirse a la constitución de los centros de interés, López Morales (1999)— a la necesidad de representar distintos tipos de

experiencias y relaciones, y de ser lo más exhaustivos posible. Pero, aunque no afecta los resultados de la investigación (cuyo objetivo, después de todo, es representar el léxico culto, y no clasificarlo), esta falta de sistematicidad dificulta el manejo de los materiales, situación que pudo haberse evitado si todos los conceptos se hubieran agrupado bajo los conceptos de clase correspondientes. Así, por ejemplo, en vez de *cabeza* estar subordinada a *El cuerpo humano*, lo estaría a *Partes del cuerpo*. Del mismo modo, los instrumentos de un trabajador no estarían incluidos bajo *Profesiones y oficios*, sino bajo *Instrumentos de trabajo*. En todo caso, la redundancia (v.g., la inclusión de manzana bajo *Postres y Frutas*) es preferible a la ausencia de un concepto (y su lexía) en el apartado correspondiente.

Mucho más serio —pues afecta el resultado de las investigaciones— es el problema de las inconsistencias en la delimitación de la extensión de los campos léxicos. El apartado de *Los instrumentos musicales*, por ejemplo, sólo incluye ocho instrumentos y, para colmo, cuatro de ellos son subclases de dos de la lista (*guitarra, guitarra eléctrica, guitarra pequeña de 3 cuerdas, flauta, flauta con lengüeta, flauta sin lengüeta, armónica y tambor*). Del mismo modo, bajo *Profesiones y oficios* figuran 25 clases de médico y 12 de ingeniero, pero sólo dos clases de músico (el *pianista* y el *violinista*), y no aparece ningún tipo de escritor/a. Bajo *El cuerpo humano*, por su parte, figuran *individuo alto* (pero no el bajo), *pequeño* (pero no grande), *persona muy delgada* (y no la muy gruesa), e *individuo grueso y pequeño* (pero no el delgado y pequeño, el delgado y grande, y el grueso y grande). En ese mismo campo, entre las acciones relacionadas con la boca se incluyen *gemir, gimotear, sollozar, hacer gárgaras, escupir, morder, comer, eructar, silbar, cuchichear, chistar, gritar y bostezar*, pero no otras como lamer, besar, chupar, mamar, masticar y hablar. Y en el campo de *La iglesia* figuran la *Virgen del Pilar*, pero no la del Carmen; *San Pedro*, pero no Jesús; el *Miércoles de Ceniza*, pero no el Viernes Santo; la *sotana*, el *campanario* y el *cáliz*, pero no la Biblia, el pecado y la misa. Y suma y sigue.

Ni en la introducción del libro ni en las tesis en que éste se basa se explican las pautas que se siguieron para determinar la extensión de los campos. Sin embargo, resulta evidente que no fueron uniformes ya que, frente al carácter marcadamente parcial

de los centros antes señalados, otros, como el del esqueleto humano, son exhaustivos. Por otro lado, es estadísticamente muy poco probable que la constitución de campos como el de los instrumentos musicales (que, como vimos, incluye sólo ocho conceptos, cuatro de los cuales son subclases de otros dos) se haya realizado mediante un proceso de selección aleatoria. En todo caso, estas arbitrariedades privan al inventario del rigor y la coherencia que se espera de toda investigación que reclama ser científica y, sobre todo, del valor descriptivo al que aspira, ya que, al no estar incluidos los conceptos correspondientes en el cuestionario, numerosas voces del vocabulario analizado quedan fuera del recuento.

4. Franco's Family Values

Mas estas inconsistencias no son siempre el resultado de la falta de rigor metodológico. Al considerar la constitución de algunos campos léxicos, el carácter sistemático de ciertas exclusiones, y la naturaleza y el ordenamiento de los conceptos que se incluyen, delatan las ideologías de naturaleza social y moral que informan dichas estrategias, y que subvierten aún más las pretensiones científicas del estudio.

Para empezar, tiene un carácter marcadamente sexista la falta de simetría entre ciertos conceptos del campo de *El ciclo de vida*, en el cual, si bien se incluyen tanto *soltera entrada en años como hombre soltero entrado en años*, no hay equivalentes masculinos para los conceptos de *mujer en edad de casarse*, *mujer entrada en años* y *mujer madura*, que apuntan hacia la concepción de la mujer como objeto sexual y agente de la reproducción. El sexismo también informa el uso del masculino genérico en la representación de los conceptos del cuestionario (con excepción de aquellos que denotan oficios considerados femeninos, como *modista*, *bordadora*, *peinadora*, *lavandera*, *niñera*, *camarera*, *cocinera* y *doncella*), así como (aparentemente) también la configuración de los estímulos verbales y pictográficos creados por los/as investigadores/as puertorriqueños/as, cosa que explicaría el que invariablemente, y con excepción de las preguntas sobre los trabajos antes mencionados, las personas encuestadas hayan respondido con sustantivos masculinos.

Esta ideología se revela así mismo en el ordenamiento, jerarquizado en términos patriarcales, de *Los habitantes de la casa*

(*el señor, la señora, los hijos, el propietario de la casa, los huéspedes* y, por último, *la criada*), campo cuya configuración también delata otras unidades ideológicas de tipo social y moral. En este caso, la extensión de los *habitantes de la casa* (con sus numerosas posibilidades) se reduce a la institución de la familia burguesa, la cual se define no sólo a partir de relaciones monógamas y heterosexuales (de ahí, *señor/señora*), sino también en términos de una sexualidad limitada a la función reproductiva (*los hijos*), y en un contexto de relaciones de autoridad, a las que remite, como vimos, el ordenamiento de los conceptos, así como el empleo de los términos *señor* y *señora*, los cuales, como se sabe, connotan jerarquía social.

La concepción de la sexualidad en términos exclusivamente reproductivos también informa la configuración del campo de *El ciclo de vida*, el cual se inicia, como por obra del Espíritu Santo, *in medias res*, con el embarazo, al que siguen, en orden cronológico, el *feto*, el *parto*, y las llamadas edades del hombre, hasta culminar con *la vejez*. A la vez que se suprime toda referencia directa a la sexualidad (en consonancia con la visión cristiana del cuerpo como *locus del pecado*), lo *sexual* se codifica eufemísticamente en los términos sociales del matrimonio (*mujer en edad de casarse, soltera entrada en años* y *hombre soltero entrado en años*) y mediante sinécdoques causales que representan los efectos biológicos del acto carnal (el *embarazo*, el *feto*, etc.).

Esa ideología se hace aún más evidente en la constitución y ordenamiento del campo de las *Relaciones de las personas*. Tras una serie de conceptos de relaciones no sexuales (*vecino, colega, etc.*), aparece *noviazgo*, con el que se inicia una secuencia cronológica —una trama— que culmina con el matrimonio y la procreación. La primera etapa de esta secuencia está constituida por los conceptos *galantear/cortejar, acompañar, declaración, declararse, aceptación, rechazo, hablar los novios, compromiso* y *petición de mano*. Se trata de una serie de intercambios en los que la sexualidad sólo está presente en forma marginal (en los conceptos de *galantear* y *cortejar*), ya que aún no ha sido legitimada mediante el contrato social —*el compromiso* y *la petición de mano*— con el que concluye la serie, y que a la vez autoriza los comportamientos que siguen, en los que se intensifica el elemento erótico (*coquetería, ligar, besuquearse/manosearse la pareja*). Esta segunda secuencia

culmina con otro contrato social: al llegar a ese borde del pecado del besuqueo y el manoseo, inevitablemente aparecen —*precoitus interruptus*— las amonestaciones (que anticipan que se trata de un matrimonio católico), la *boda* y las *fiestas de esponsales*.

A las fiestas sigue una digresión —la unión sexual queda nuevamente diferida— en la que se enumeran (imposible tapar el cielo con la mano, aun en la España de Franco, en la que se publica el *Cuestionario*) otros tipos de relaciones socio-sexuales frente a las cuales se define el matrimonio católico: el *matrimonio civil*, el *adulterio*, el *concubinato* y la *unión libre*. Aunque constituyen márgenes prohibidos por la ortodoxia católica, la inclusión de las tres primeras relaciones cumple la función de afirmar el orden (hetero)sexual. El adulterio, según el *Diccionario de la Real Academia Española de la Lengua*, se define como “ayuntamiento carnal voluntario entre persona casada y otra de *distinto sexo* que no sea su cónyuge” (énfasis nuestro). De ahí que el concepto para las personas que así se ayuntan sea *adúltero, a*, uno de los poquísimos casos en que en el *Léxico* no se usa el masculino genérico. Lo mismo ocurre con *concubinato*, ya desde su etimología un concepto de relaciones heterosexuales: la palabra se deriva de concubina, “manceba o *mujer* que vive y cohabita con *un hombre* como si este fuera su marido” (DRAEL). De hecho, el sema de heterosexualidad contenido en el concepto hace posible que en el cuestionario se use la conjunción disyuntiva *o* en la representación de las personas que así cohabitan: el *o la que vive en concubinato*; el que así vive necesariamente lo hace con una mujer, y vice versa. En este contexto, el concepto de la *unión libre*, al estar desprovisto de semas indicadores del sexo de los participantes, constituye un concepto potencialmente ambiguo y peligroso. De ahí que el texto, para afirmar la ortodoxia sexual, recurra al empleo de la conjunción copulativa y en el concepto de las personas: el *y la que vive en unión libre* y no, como en el caso del concubinato, *el o la que vive*.

Ahora bien: aunque estos conceptos reafirman el orden sexual establecido, representan una desviación moral —no es fortuito que también constituyan una digresión textual— que debe ser corregida, y de ahí que el texto retome el hilo discursivo —la secuencia cronológica interrumpida tras las fiestas de esponsales— con la enumeración de aniversarios de boda (¿está de más decir que no se menciona para nada el divorcio?), *happy ending* que afirma el triunfo

de la moral católica: *aniversario de matrimonio, al año, a los cinco años, a los diez años, a los veinticinco años, a los cincuenta años y —una proeza— a los setenta y cinco años*. Tras este último es que aparece el *viaje de bodas* y la *luna de miel*, representaciones eufemísticas del acto sexual hasta ahora diferido, y aquí ubicadas porque conducen de forma inmediata a *día onomástico, aniversario del nacimiento*, y el campo de *Los hijos*, con los que se representa la procreación, fin último (y único) del trato carnal.

La negación de la sexualidad se hace aún más patente en la figuración del cuerpo humano que se presenta en el campo homónimo. Por un lado, el cuerpo se somete a un proceso de higienización, mediante el cual se excluyen elementos escatológicos (en la lista figuran eructo y eructar, pero no los gases que se expelen por el ano o la acción correspondiente; tampoco figura la excreta, a pesar de que se incluyen secreciones y excrementos como sudor, *legaña, moco, destilación continua producida por un resfriado* y orina). Paralela a este proceso de limpieza, se da una supresión de elementos de carácter sexual, y así, mientras que a algunas partes del cuerpo corresponden listas extensas de conceptos (para el esqueleto hay 23 apartados, que incluyen conceptos relativamente técnicos como el *radio, el cúbito* y la *tibia*; para la boca hay 17, entre los que figuran dientes, *incisivos, incisivos muy grandes, encías, paladar, saliva, garaganta, lengua y labios*), los órganos sexuales se reducen a dos entradas, el órgano masculino y el femenino, sin detallar sus partes. (De milagro no se redujo todo a “las partes pubendas”.) A esta “castración” textual corresponde la sustitución del semen y su eyaculación por el concepto, cargado ideológicamente, de *poluciones nocturnas*, que arrastra el sentido de condena moral de *poluciones*, término que, a partir del significado de “mancha” de la voz latina primitiva, denota, además de “efusión de semen”, “corrupción moral” y “acto carnal deshonesto”. Se trata, por otro lado, de una unidad ideológica del discurso (de estirpe escolástica) de la iglesia católica en torno a la sexualidad: estas efusiones, a pesar de que constituyen un desperdicio (es decir, son contrarias al fin reproductivo), son involuntarias (“nocturnas”), por lo que no constituyen pecado.

No es de extrañar, en este contexto, que las experiencias y prácticas de tipo espiritual o religioso, para las cuales algunos investigadores proponen un centro de interés amplio —v.g., el

Mundo espiritual, de Valencia (1994)—, queden en este inventario reducidas al campo de *La Iglesia*, concepto en cuya representación el artículo definido y la mayúscula ya indican claramente el presupuesto ideológico de universalidad de la iglesia católica que informa la configuración del campo, cuyos ochenta conceptos invariablemente se refieren a objetos, sujetos y prácticas de esa institución.

5. España, o sea, el Universo

Así constituidos los campos, resulta inevitable el que la imagen del léxico culto de San Juan que ofrece este recuento resulte trunca y distorsionada. Pero el problema mayor es que, a partir del presupuesto, de corte decididamente imperialista, de la universalidad de los referentes españoles (o castellanos), el cuestionario recorta un universo de objetos, saberes y experiencias ajenas a las realidades de la comunidad lingüística que se estudia.

Por un lado, esta falta de correspondencia entre el instrumento y el objeto de estudio explica el elevado número de ítems de la encuesta —578 de 4,452— que no obtuvieron respuesta. No es necesario realizar (como propone el coordinador del trabajo) “un estudio monográfico” para descubrir las causas fundamentales de tan alto porcentaje de entradas sin contestación, ya que él mismo lo sugiere antes, al señalar que los investigadores preguntaron todos los puntos “aun a sabiendas de que algunos de ellos no obtendrían respuesta por tratarse de realidades desconocidas en el país”, como sucede con conceptos como *bofe*, *sangre encebollada* y *queso de cabeza* (del campo de *La alimentación*); *lucernario*, *gárgola* y *buhardilla* (de *Partes de la casa*); *tetera*, *tartera* y *puchero* (de *Utensilios de la cocina*); *granizada*, *carámbanos* y *cellisca* (de *Meteorología*); *espliego*, *cáñamo* y *mora de morera* (de *Vegetales/ Agricultura*); *langosta*, *tábano* y *ciervo volátil* (de *Insectos*); y *raya*, *jurel* y *sepia* (de *Pescados y mariscos*). Tampoco recibieron respuesta, y por razones evidentes, los preguntas sobre las voces para *llamar*, *espantar* o *animar al burro* (¿¡Plateeeero!?). Y, por los mismos motivos, recibieron muy pocas contestaciones conceptos como *las pastas para el té*, *la sopa de tapioca*, *la escarola*, *el cogollo*, *el dulce de membrillo* y *los orejones*, del campo de la

alimentación, y la gran mayoría de las 102 entradas correspondientes a *El ferrocarril*, entre muchos otros ejemplos.

Además, las respuestas a algunos puntos parecen obedecer a la competencia lingüística general de los/as encuestados/as, más que a su conocimiento de los conceptos (o vocablos) propiamente dichos. Este sería el caso de ciertos nombres colectivos de árboles incluidos en el campo de *El terreno*, como *naranjal*, *pinar*, *cedral*, *robledal* y *encinar*, para los que los sujetos que ofrecieron las respuestas “correctas” aparentemente recurrieron a su conocimiento de la derivación; algo similar ocurre con la respuesta *huevos crudos con limón*, en la que el encuestado meramente describe el estímulo verbal o pictográfico de un referente culinario que nos resulta desconocido: lo mismo hubiera pasado si entre los postres se hubiera incluido un concepto como “ojos de gato tuerto y triste en escabeche”.

En términos metodológicos, la inclusión de todos estos referentes ajenos a las realidades del grupo cuyo léxico se pretende analizar quizá sólo constituya una pérdida de energías y tiempo. Es más: no ignoramos que la ausencia o el bajo número de respuestas sirvan para definir los contornos de ese léxico. Sin embargo, resulta inadmisibles que se pregunte sobre esos conceptos a la vez que se excluye un conjunto, cuando menos igualmente numeroso, de otros que sí corresponden a nuestra realidad.

Así, por ejemplo, mientras que en el campo de los deportes se incluyen 37 conceptos relacionados con los toros y 48 con el fútbol, sólo figuran 5 bajo el campo del baloncesto, 1 —el concepto mismo— relacionado con la lucha libre, y ninguno con el volibol. Del mismo modo, entre los condimentos figura el *azafrán*, pero no el *achiote*; entre los arroces se encuentra la *paella*, pero no el *asopao*; entre los postres no figuran el *majarete*, el *tembleque* o el dulce de *papaya*; los conjuntos de árboles excluyen el *palmar*; entre las siembras, que incluyen las de *cebada*, *avena* y *papas*, no aparecen las de *yautías*, *ñames* o *batatas*; las flores incluyen el *gladiolo* y el *pensamiento*, pero no la *cruz de malta* o el *canario*; entre los insectos se encuentra la *libélula*, mas no la *changa* o el *gongolón*; los pescados incluyen la *mojarra*, pero no el *chillo*, la *sierra* o el *capitán*.

El problema de esta asombrosa falta de correspondencia entre el cuestionario y el objeto de estudio se agrava por el hecho de que, al estar integrado por listas extensas de conceptos específicos, el

instrumento establece *a priori* los límites del vocabulario que se va a recopilar. En este contexto, resulta a todas luces insostenible el señalamiento de López Morales de que este *Léxico* “es un *corpus* representativo y fidedigno, que recoge el vocabulario activo, de disponibilidad espontánea e inmediata, de los hablantes sanjuaneros”: más que el léxico del habla culta de San Juan, lo que este estudio configura es el léxico de España (o de Castilla) que conoce (o desconoce) un grupo de hablantes cultos de la ciudad.

El Apéndice 1 del libro trata de remediar en algo el problema, al incluir listas de palabras recopiladas mediante pruebas que se realizaron con la técnica de serie abierta, a partir de centros de interés relativamente amplios. (Así, por ejemplo, entre los subcampos de *El vestido* se incluye *Pantalones* solo, sin la enumeración detallada de conceptos relacionados —corto, largo, a media pierna, etc.— que figura en el cuestionario.) Sin embargo, el Apéndice no incluye siete de los veintidós campos generales del inventario anterior. Por otro lado, al usarse un método diferente, que por su naturaleza genera un número menor de respuestas, se hace imposible establecer una relación cuantitativa válida entre los vocablos así recogidos y los recopilados a partir del cuestionario. Además, con pocas excepciones, las palabras del Apéndice fueron producidas por sólo uno o dos encuestados, por lo que resultan estadísticamente insignificantes.

Por último, cabe señalar que la actitud de imperialismo cultural que anima las inclusiones y exclusiones del cuestionario tiene su contrapartida en la representación de los conceptos, en la que se privilegian los usos castellanos: el mentón (por la barbilla), el tocólogo (el/la obstetra), la cena (la comida), tapas (entremeses), tortilla de *patatas* (de papas), *grill* (parrilla), *solomillo* (lomillo), *asaduras* (mondongo o gandinga), ensalada *rusa* (de papas), *oca* (ganso), *judías* (habichuelas), café *exprés* (expreso), etc., etc.

Posiblemente, los autores del cuestionario (y los/as investigadores/as que tan acriticamente lo adoptaron) recurrieron a estas formas en su afán de “dar uniformidad al considerable volumen de datos que arrojaría la investigación”, y así hacer posibles las comparaciones posteriores, preocupación comprensible en vista de que el número de vocablos panhispánicos de los léxicos disponibles —con excepción del que corresponde a ciertos centros de interés, como el cuerpo humano (Alba, 1998)— es relativamente bajo

(López Chávez, 1995). Sin embargo, el problema muy bien se pudo haber resuelto con el uso de sinónimos en los casos en que no hubiera un vocablo general. Incluso —¿por qué no?— se pudo haber usado las voces puertorriqueñas: después de todo, el estudio se realiza y se publica en Puerto Rico, y tiene por lo menos un lector implícito puertorriqueño: los educadores del país, para quienes se propone como instrumento de la planificación escolar. En todo caso, resulta insostenible el que se pretenda que los/as puertorriqueños/as conozcamos los usos castellanos, mientras que no se les exige a los hablantes de ese dialecto igual conocimiento de los nuestros.

6. El *Léxico disponible* de Puerto Rico

El estudio sobre el léxico disponible realizado posteriormente por López Morales supera en forma significativa muchos de los problemas del recuento anterior. Para empezar, se trabajó con una muestra muy amplia y diversa: de un universo de estudiantes de segundo semestre universitario, se seleccionaron al azar 25 grupos (para un total de 558 individuos), que representaban igual número de centros universitarios del país. Además, se administró la misma prueba a todos/as los/as encuestados/as, a diferencia de las investigaciones del *Léxico del habla culta*, en las que las distintas partes del cuestionario se trabajaron con grupos diferentes. Por otro lado, mientras que en dicho estudio sólo se consideró la frecuencia alcanzada por las unidades léxicas, en el que ahora nos ocupa se ponderó la frecuencia con el grado de “espontaneidad”, u orden de aparición de las palabras, de gran importancia en vista de que aquellas “que acuden primero a la memoria como reacción a un estímulo determinado son realmente las más disponibles” (López Morales, 1999). Para ello, el investigador utilizó sofisticadas fórmulas matemáticas capaces de discriminar por rangos a las unidades que coincidieran en frecuencia.

Contribuye así mismo al valor descriptivo del estudio el uso de la técnica de serie abierta, y el de centros de interés generales, que no establecen de antemano (como sí hacía el cuestionario, excesivamente detallado, de Alvar y Flórez) una extensión definida para el léxico que se va a recopilar. En este caso, López Morales adoptó, con pequeñas modificaciones, los 16 centros diseñados por Gougenheim (1964), los cuales presentan la ventaja de no ser ni

demasiado específicos (lo que generaría la posibilidad de “vacío léxico”, o ausencia de respuestas) ni muy amplios (lo que daría lugar a listas excesivamente heterogéneas, y en las que pocos términos alcanzarían una frecuencia significativa).

También cabe señalar que la consulta de los materiales se facilita con la inclusión de dos listas para cada campo: una ordenada por rango, y otra, alfabéticamente.

Sin embargo, el estudio deja sin resolver importantes cuestiones teóricas y metodológicas, entre las que figura la de la constitución de los centros que se utilizan como estímulos. López Morales señala que recurrió al empleo de los diseñados por Gougenheim en vista de que, habiendo sido ya usados en otras investigaciones, hacen posible la realización de estudios comparativos interdialectales e interlingüísticos posteriores: la decisión “apenas necesita explicaciones, pues el solo hecho de permitir la comparación de materiales hispánicos americanos y europeos, en primer lugar, y también entre resultados logrados entre nosotros y el mundo francohablante, es ya de por sí de importancia sobresaliente”. Sin embargo, la investigación se aleja de este objetivo al sustituir, por razones evidentes, el concepto de la *calefacción*, que figura en el estudio original, por el de *aire acondicionado*, y el de la escuela, de otro centro, por el de la *universidad*. Estos cambios no plantean mayores obstáculos a las investigaciones comparativas, ya que se mantienen inalterados los restantes 14 centros de interés. Pero sí mina las posibilidades de dichos estudios el que la muestra estuviera constituida por estudiantes universitarios (en vez de los de 9-12 años con los que trabajaron Gougenheim y sus seguidores), ya que el vocabulario de ambas poblaciones es cualitativa y cuantitativamente muy distinto. Por otro lado, en la argumentación del autor se privilegia, como antes se hizo en el caso del *Léxico del habla culta*, el objetivo secundario de los estudios comparativos, sobre el propósito fundamental, postulado explícitamente en la Introducción y en el título del libro, de describir el léxico disponible de una comunidad. Ya vimos los desastrosos resultados de la utilización del cuestionario de Alvar y Flórez en el estudio anterior: claramente, los instrumentos de la investigación tienen que seleccionarse en función de su objetivo primordial, y no de acuerdo con propósitos secundarios.

En este contexto, cabe preguntarse en qué medida unos centros diseñados para describir los hechos lingüísticos de determinado grupo o sociedad pueden servir para el análisis de los de otros.

Ya ha sido abundantemente demostrado que existen vínculos estrechos entre el mundo físico, social y cultural de una comunidad de hablantes y su lengua, y que éstos van desde las más evidentes determinaciones de la realidad sobre el lenguaje, hasta las maneras en que, como han establecido Whorf (1971) y otros, cada lengua (o dialecto) organiza la experiencia en formas que le son peculiares. Los centros de interés, en cuanto regiones nocionales y semánticas, no son ajenos a estas relaciones: tanto el número de campos, como los límites y el valor relativo de cada uno, dependen de las realidades de la comunidad, y de la forma en que ésta ha codificado el mundo y la experiencia.

Muy poco “interesantes” serán para los integrantes de alguna remota tribu amazónica ciertos campos de Gougenheim, como *La ciudad*, *La escuela: muebles y material escolar o Iluminación y calefacción*; probablemente tampoco recibirían muchas respuestas de ese grupo centros en apariencia más universales como *La ropa*, *Los muebles de la casa*, *Objetos situados en la mesa para las comidas*, *Trabajos del jardín* e, incluso, *El campo* (que posiblemente no constituya una realidad diferenciada). Además, aun en el caso de centros universales como *Los alimentos* y *Las partes del cuerpo*, el número de respuestas variará de acuerdo con la importancia relativa que éstos tengan en la cultura y con la manera en que se hayan categorizado lingüísticamente esas realidades. Por otro lado, de seguro la lista del investigador francés no incluye campos que constituyen áreas del saber y la experiencia importantes para ese hipotético grupo del Amazonas.

El mismo López Morales (1999) nos recuerda que, frente a la idea de Gougenheim y sus colaboradores de que los centros propuestos por ellos representaban los sectores semánticos más universales, estudiosos como Mackey han evidenciado que la importancia de estos campos está muy vinculada con factores culturales y sociales.

Si bien Puerto Rico difiere en importantes aspectos de esa imaginada comunidad del Amazonas —como prueba el hecho de que todos los centros del investigador francés utilizados en este estudio recibieron respuestas—, no es menos cierto que nuestro país

tampoco es Francia —lo que de inmediato se hace patente con la sustitución, a la que se vio obligado el autor, de la calefacción por el aire acondicionado. Además, aunque los campos sean comunes a ambas comunidades, queda sin resolver el asunto del valor relativo de estos centros en cada cultura. En este contexto, no nos parece fortuito el que tres de los dieciséis campos incluidos en el estudio francés versen sobre la cocina, práctica que goza de gran prestigio en ese país. Tampoco debe extrañarnos el que el campo que menos respuestas recibió en el estudio de nuestro léxico haya sido el de los *Medios de transporte*, en vista de la escasa variedad de éstos con que contamos.

Es más, el instrumento no sólo traza un mapa cultural, sino también ideológico: la Francia dibujada por los centros de Gougenheim, con esos bucólicos Trabajos del campo y del jardín (pero sin los de las fábricas y las oficinas), de seguro dista mucho de las experiencias y los intereses de la mayoría de los trabajadores de ese país. Y se trata además de un instrumento marcado por la población de niños/as para la que se diseñó: de ahí que se incluya un centro como *Juegos y diversiones* (para no hablar del más evidente de *La escuela: muebles y material escolar*) y, a la vez, se excluyan otros relacionados con realidades más complejas o abstractas, como la tecnología, la industria o las finanzas.

Como resultado de la adopción de estos centros, el estudio de López Morales de seguro recoge voces que, a pesar de que forman parte del léxico disponible, tienen un índice de disponibilidad mucho menor que el de otros vocablos que, por no corresponder a los campos establecidos, quedan fuera del recuento. Más preocupante es la exclusión de centros relacionados con tantas otras áreas del saber y la experiencia (el gobierno y la política; la industria, el comercio y las finanzas; las letras y las artes; los medios de comunicación; la ciencia y la tecnología; la salud y el ambiente; las relaciones humanas; la geografía física; la vida mental y espiritual, etc.). Aun reconociendo la imposibilidad de un inventario exhaustivo, entendemos que la omisión de todos estos centros no puede sino producir un perfil marcadamente incompleto y desfigurado del léxico que se pretende describir.

Otro problema que plantea la adopción de estos centros es que, por la manera en que están formulados, tienden a propiciar la obtención de sustantivos sobre otros tipos de palabras. Diez de los

dieciséis centros están representados por conceptos genéricos cuya extensión se limita a cosas y seres (*Partes del cuerpo, La ropa, Los muebles de la casa, Alimentos, Objetos situados en la mesa para las comidas, La cocina: muebles y utensilios, La universidad (o la escuela): muebles y materiales, Medios de transporte, Animales y Profesiones y oficios*), y otros tres (*La casa, La ciudad, El campo*), por “todos” espaciales que también requieren la enumeración de cosas. Sólo en los nombres con que se designan dos centros, *Trabajos del campo y del jardín* y *Juegos y diversiones*, figuran semas que estimulen la producción de verbos. Por otro lado, la preponderancia de los centros que exigen sustantivos, y el que 12 de ellos precedan al de los *Trabajos*, de seguro influye en las respuestas a éste, en el que cerca del 26% corresponde a nombres. El predominio de esta clase de palabras tampoco es de extrañar en vista de que todos los centros están representados mediante sintagmas nominales; como ha demostrado De ese (1965, citado en Aitchison, 1987), en las pruebas asociativas los hablantes tienden a responder con palabras de la misma categoría que las del estímulo, especialmente en el caso de los nombres.

De hecho, otros recuentos de léxico disponible realizados con anterioridad han estado constituidos fundamentalmente por sustantivos, ya que sus centros se han codificado en términos análogos a los de Gougenheim. Sólo unos pocos investigadores han incluido estímulos diseñados expresamente para recoger otros tipos de palabras, y aun así éstos se suelen reducir a uno o dos por estudio: entre los 15 campos de la investigación de Azurmendi (1983, cit. en López Morales, 1999), figuran *Acciones que normalmente se realizan todos los días* y otro que incluye lo que se *hace en la ciudad o el pueblo*; Canízal Arévalo (1987, cit. en López Morales, 1999) incluye entre sus campos el de *Defectos físicos y morales*, con el que busca conseguir adjetivos; y Valencia (1994, cit. en López Morales, 1999) incluye *Procesos mentales*. El propio López Morales recurre al apartado de Gougenheim sobre Los trabajos del campo y el jardín “con el objeto de conseguir verbos”; casi el 74% de los vocablos de este centro, y 20% de los de *Juegos y diversiones*, correspondieron a esa categoría gramatical.

No ignoramos que los verbos (y, en menor medida, los adjetivos) suelen ser vocablos de mayor generalidad que los sustantivos, y que por ello tiendan a predominar, junto a las palabras

gramaticales, en los diccionarios del léxico básico (Morales, 1986; López Morales, 1986). También somos conscientes de que es precisamente a partir de este hecho que los centros de interés de los estudios de disponibilidad tradicionalmente se han diseñado con el propósito de obtener nombres. Sin embargo, no es menos cierto que hay numerosos adjetivos y verbos, muy conocidos, que sólo se utilizan en ciertos contextos comunicativos (vg., verbos relacionados con acciones del cuerpo como pestañear, pellizcar, rascarse, estornudar, evacuar y orinar), por lo que deberían recogerse en los estudios del léxico disponible.

Pero aun si las pruebas incluyeran los centros necesarios para obtener estos vocablos, es muy posible que la cantidad de ellos que se recoja no refleje adecuadamente su disponibilidad. En el estudio de Deese antes citado, el 80% de los informantes respondió con sustantivos ante estímulos nominales, pero sólo poco más de la mitad produjo verbos y adjetivos cuando los estímulos eran de esas clases. Otros estudios citados por Aitchison sugieren que, en las pruebas asociativas, los nombres son más fáciles de recordar, no sólo porque hay más, sino también porque carecen de las restricciones sintácticas que suelen acompañar a los verbos. Esto plantea la posibilidad de que este tipo de pruebas, en las que se trabaja sin contextos lingüísticos, no sea el más apropiado para recoger estos verbos y adjetivos temáticos.

Por otro lado, entendemos que, dentro de las limitaciones de tiempo de estas encuestas, resulta imposible incluir todos los centros de interés imaginables. Sin embargo, su número podría ser mayor si se aumentara el tiempo que se dedica a las pruebas, o se trabajara con submuestras paralelas de tamaño adecuado. En todo caso, nos parece inaceptable el que, con el pretexto de que resultan útiles para los estudios comparativos, se siga recurriendo a modelos que no corresponden con la realidad lingüística que se pretende describir. Y, aunque el *desideratum* de exhaustividad resulte siempre inalcanzable, el valor de estos estudios ha de determinarse precisamente en la medida en que se acerquen a ese ideal.

Por último, llaman la atención ciertas inconsistencias en la codificación y análisis de los datos, las cuales señalan hacia otros problemas metodológicos aún no resueltos. Para el proceso de segmentación de las unidades léxicas se adoptó el criterio de incluir todos los vocablos diferentes, por lo que se recogen no sólo formas

simples (*arroz*), sino también aquellas secuencias, ya lexicalizadas, que funcionan como un todo (*arroz con pollo*). También responde a este criterio el uso de entradas separadas para las voces que representan un mismo referente (*laundry*, lavandería, cuarto de lavar, cuarto de *laundry*), así como la decisión de lematizar los vocablos, llevando “a las categorías nominales que lo permitieran al masculino singular y a las verbales, al infinitivo”. Sin embargo, esta pauta no se sigue de forma consistente, pues reiteradamente se incluyen voces femeninas junto a las masculinas (*perro* y *perra*, *abogado* y *abogada*, etc.). Tampoco hay uniformidad en la representación de los plurales: junto a las formas, muchas de ellas extrañas a los usos establecidos, que resultan de este criterio de lematización (*arroz con habichuela*, *arroz con gandul*, *cocktail de fruta*, *guineo en escabeche*, etc.), figuran algunos vocablos en plural (*platanutres*, *zucaritas*, *panties*, *medias cortas*, entre otros).

López Morales así mismo decidió someter las palabras a un proceso de nivelación ortográfica, el cual, además de ser “imprescindible para el trabajo de computación electrónica”, resulta a nuestro entender cónsono con el criterio de segmentación adoptado. A partir de este criterio, por ejemplo, respuestas como “biftek”, “biftéc” y “biftec” figurarían todas bajo una entrada con grafía correcta o estándar (en este caso, *biftec*). Sin embargo, en el libro aparecen palabras con errores ortográficos, como *anis* (anís), *manhattann* (manhattan), *zoombie* (zombie), *ice tea* (iced tea), *mayorca* (mallorca) y *te* (té), para limitarnos al campo de Los alimentos. También podemos considerar en este grupo los casos de *maizena* y *congrís*, que constituyen, frente a *maicena* y *congrí*, variantes de poca aceptación.

La falta de consistencia ortográfica también se observa en el uso de las grafías originales, escritas en bastardillas, de ciertas palabras extranjeras para las cuales existen en nuestro país formas castellanizadas de uso general: *spaghetti* (espagueti), *lasagna* (lasaña), *cocktail* (coctel), *champagne* (champán), *cognac* (coñac).

Estas inconsistencias se tornan problemáticas en el caso de entradas como *kaka*, *petipúa* y *petipo*. En vista de la escasa competencia ortográfica de los sujetos que constituyeron la muestra, y de que se les instruyó para que “no repararan en la ortografía”, resulta evidente que *kaka*, la cual el investigador considera como voz extranjera de uso idiolectal, no es sino una variante de “caca”.

En el caso de las segundas, se trata a todas luces de dos variantes ortográficas de “petipúa” y “petipó” (las cuales son a su vez variantes fonéticas comunes de pitipúa, que se incluye por separado).

Pero aun si se nivelara la ortografía de estos vocablos, quedaría sin resolver el problema de las variantes fonéticas. Aunque el autor no alude a ellas (acaso porque se trataba de una prueba escrita), aparentemente ha optado por considerarlas como unidades distintas; de ahí que figuren por separado *petipo* y *petipúa* (así como *batipoit*), a pesar de los indicios que sugieren que se tratan de meras variantes. Sin embargo, el criterio no se ha usado de forma uniforme. La entrada para *biftec* probablemente recoge no sólo variantes ortográficas, sino también fonéticas, como *bisté* y *bistec*. Así mismo, las grafías de los extranjerismos antes mencionados implican un proceso de nivelación fonética (/espageti/ se transforma en /spageti; /champán/, /champana/ y /champéin/ se reducen a /shampéin/ —o a /shampan/, si en francés—, etc.). En este contexto, *petipúa*, *petipó* y *batipoit* debieron considerarse como variantes de *pitipúa*, y computarse juntas bajo dicha entrada (a diferencia, por otro lado, de *chícharos* y *guisantes*, que también designan el mismo referente, pero constituyen vocablos distintos).

Limitándonos nuevamente al campo de los alimentos, el investigador también considera que son formas idiolectales respuestas como *moronga*, *morongo* y *ave con habichuelas*. La inclusión de las dos primeras entradas aparentemente obedece a una lectura equivocada de la caligrafía del encuestado: de nuestro conocimiento del campo de los alimentos puertorriqueños, y de los alógrafos, podemos sin mayor esfuerzo deducir que el investigador confundió la *a* cursiva con la *o* en la primera voz, y la *f* con la *r* en ambas respuestas, las cuales se refieren a un plato conocido, el *mofongo*. Lo mismo parece haber ocurrido con *ave con habichuela*, en que la primera palabra debió interpretarse como *arroz*, particularmente en vista de que muchas personas escriben la *rr* cursiva en forma parecida a la *v*, y de que, por la aspiración u omisión de la *s* en fin de sílaba, son frecuentes, en el habla vulgar, grafías como *arro*. Incluso es posible que *ave* sea el resultado de la lematización, por parte del investigador, de la forma plural. Además, a diferencia del referente, desconocido, de la entrada incluida, el *arroz con habichuelas* es un plato muy común, y los dos sustantivos

del sintagma que lo representa son las palabras con mayor índice de disponibilidad en el campo de los alimentos. Acaso también responden a errores de interpretación de la caligrafía entradas como *panzza* (probablemente *pizza*) y *pocho* (pollo).

La palabra *henequén*, también de este campo, plantea otros problemas. Ya que se refiere a una planta textil, no debió aparecer en la lista, para ser consistentes con otro criterio de edición establecido por López Morales: el de excluir las respuestas que no correspondieran al centro en cuestión. Pero cabría otra interpretación, más conforme con la realidad puertorriqueña: el encuestado probablemente se refería a Heineken, una conocida marca de cerveza. En este caso, debió aparecer en bastardillas, como las otras voces extranjeras, y particularmente en vista de que la *h* inicial es aspirada. Sin embargo, se estaría con ello violando otra pauta del investigador, la de excluir los nombres de marcas comerciales que no estuvieran lexicalizadas.

En todos estos casos nos encontramos ante un mismo problema hermenéutico: en vista de la ausencia de un contexto (fuera del que provee el centro de interés), el investigador se ve continuamente obligado a tomar decisiones que implican adjudicarle una intención comunicativa al informante. Para ello, necesariamente se vale de su conocimiento del mundo (y en particular de los referentes que constituyen cada centro de interés) y de la lengua y el habla de la comunidad (incluyendo el vocabulario, y los aspectos fonéticos, ortográficos y aun caligráficos). Esto es lo que le permite, por ejemplo, incluir bajo una sola entrada variantes como *bistec* y *bistéc*, *bisté* y *biste*, *biftec* y *biftéc*, *bistek*, *viftec*, etc. Sin embargo, el autor de este estudio, acaso por su afán de objetividad científica, muestra gran vacilación en el proceso, lo que lo lleva a considerar como formas idiolectales muchas respuestas que claramente constituyen meras variantes gráficas o fonéticas.

El asunto se complica cuando consideramos respuestas más difíciles de nivelar, como *palta* (¿variante caligráfica o fonética de *malta*?), *emperador* (¿variante fonética de *emparedado*?), así como otras imposibles de recuperar, como *chalpa*, *pitana*, *longo en salsa*, y *like*. El investigador nuevamente optó por tratar estas respuestas como formas idiolectales, e incluirlas en el recuento. Sin embargo, esta decisión no es coherente con el criterio de exclusión de las respuestas que no correspondieran al centro de interés. Desde este

punto de vista, se incluiría en el apartado de los alimentos cualquier grafía (*bati, batu, boti, botto...*), siempre y cuando no coincidiera con la de otra palabra existente, pero ajena al campo (*bate, bata, bote, boto, boat...*). Sin embargo, ¿qué le dice al investigador que *bate* y *bata* no son los platos del día en la fonda que frecuenta un informante? ¿O que *bote* no designa el amarillo relleno en un barrio o un hogar particular? Si se aceptan “palabras” sin referente conocido, ¿por qué no hacerlo con aquellas que muy bien pudieran ser polisémicas?

Aunque las respuestas consideradas como formas idiolectales resultan estadísticamente insignificantes, apuntan hacia algunos de los problemas que las investigaciones sobre disponibilidad léxica no han logrado resolver. Los de interpretación caligráfica y ortográfica podrían fácilmente solucionarse con el uso de pruebas orales; las variantes fonéticas, por su parte, podrían someterse a un proceso consistente de nivelación. Pero en los otros casos, la coherencia metodológica exigiría o la inclusión de todas las respuestas (lo que daría lugar a redundancias en los casos de homónimos y a respuestas extrañas a los campos y a la realidad lingüística que se busca describir) o la eliminación de todas aquellas que no correspondieran al léxico en cuestión (en cuyo caso se correría el riesgo de excluir voces existentes, pero desconocidas por el investigador); la primera alternativa tiene la ventaja de que los usos idiolectales son de poca importancia estadística. En todo caso, lo que nos parece inaceptable es que se utilicen, como en este estudio, criterios contradictorios.

Si a estos problemas les añadimos el más serio de la falta de adecuación de los centros de interés al objeto de estudio, resulta evidente que las investigaciones sobre disponibilidad léxica aún no han alcanzado el carácter científico que para ellas se reclama, y con el que se pretende legitimarlas. Su utilidad dependerá, pues, de la cautela con la que manejemos sus hallazgos.

BIBLIOGRAFÍA

- Aitchison, Jean. 1987. *Words in the Mind. An Introduction to the Mental Lexicon*. Oxford-Nueva York: Basil Blackwell.
- Alba, Orlando. 1998. "Variable léxica y dialectología hispánica", *La Torre*, III, 7-8: 299-316.
- Alvar, Manuel y L. Flórez. 1971. *Cuestionario para el estudio de la norma lingüística culta de Iberoamérica y de la Península Ibérica*, tomo III, Léxico. Madrid: Centro de Investigaciones Científicas.
- Azurmendi, María José. 1983. *Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial a la Comarca de San Sebastián*. Guipúzcoa: Caja de Ahorros de Guipúzcoa.
- Canízal Arévalo, Alma Valentina. 1987. Disponibilidad léxica en escolares de Primaria terminada. Análisis de seis centros de interés (tesis inédita). México: Universidad Nacional Autónoma.
- Deese, J. 1965. *The Structure of Associations in Language and Thought*. Baltimore: Johns Hopkins University Press.
- Gougenheim, Georges y otros. 1964. *L'élaboration du français fondamental (1er. degré). Étude sur l'élaboration d'un vocabulaire et d'un grammaire de base*. París: Didier.
- Juilland, Alphonse y E. Chang Rodríguez. 1964. *Frequency Dictionary of Spanish Words*. La Haya: Mouton.
- López Chávez, Juan. 1995. "Léxico fundamental panhispánico: realidad o utopía", *Actas del IV Congreso Internacional sobre el Español de América*, tomo II: 1006-1014. Santiago: Universidad Católica de Chile.

Comentarios en torno a tres recuentos del léxico puertorriqueño

López Morales, Humberto. 1986. *La enseñanza de la lengua materna: lingüística para maestros de español*. 2ª ed. Madrid: Playor.

López Morales, Humberto. 1999. *Léxico disponible de Puerto Rico*. Madrid: Arco/Libros.

López Morales, Humberto, coord. 1987. *Léxico del habla culta de San Juan de Puerto Rico*. San Juan: Academia Puertorriqueña de la Lengua Española.

Valencia, Alba. 1994. "Disponibilidad léxica en educación media: III y IV", en *Encuentro en torno a la Admisión Universitaria*. Santiago de Chile: Talleres Gráficos DIVEST.

Mackey, William C. 1971. *Le vocabulaire disponible du français*, 2 vols. París-Bruselas-Montreal: Didier.

Morales, Amparo. 1986. *Léxico básico de Puerto Rico*. San Juan: Academia Puertorriqueña de la Lengua Española.

Whorf, Benjamin Lee. 1971. *Lenguaje, pensamiento y realidad*. Barcelona: Barral.