

## 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,

