

Léxico-estadística y selección del vocabulario para la enseñanza del español como lengua extranjera en Santiago de Cuba

Celia María Pérez Marqués
Leonor Grethel Sierra Salas
Haydeé Fernández Leyva

El vocabulario desempeña un rol primordial por su valor comunicativo en el aprendizaje de una lengua, además de constituir un facilitador del desarrollo cognitivo del individuo. Aunque el aprendiente domine las estructuras y relaciones gramaticales, un insuficiente repertorio lexical le impedirá una comunicación efectiva. Sin embargo, por diversas causas, entre las que se encuentran su casi ilimitado número de unidades y su carácter variable en los órdenes diacrónico, diatópico, diastrático y diafásico, el nivel lexical es considerado el de mayor complejidad. Ello hace suponer la necesidad de una explícita selección del léxico que se deberá enseñar al aprendiz de una lengua extranjera (Bartol, 2010).

En este sentido, la disponibilidad léxica resulta una herramienta léxico-estadística idónea ya que, al determinar el uso real que hace del léxico una comunidad de habla específica, proporciona al programador del vocabulario una base sólida, objetiva y científica para realizar esta tarea adecuadamente. De ahí la importancia y actualidad de la temática abordada.

Este trabajo tiene como propósito dar a conocer los principales aportes que en Cuba se han realizado desde la léxico-estadística a la enseñanza del español como lengua extranjera (ELE), particularmente en lo concerniente a la selección del vocabulario.

La selección del vocabulario en la enseñanza de ELE

Para planificadores curriculares y profesores la tarea de determinar qué vocabulario hay que enseñar en cada etapa y qué criterios deben guiar este proceso constituye un verdadero reto. En muchas ocasiones, solo la subjetividad y la intuición se han tomado en cuenta, trayendo como consecuencia selecciones arbitrarias que dificultan la progresión efectiva del aprendizaje.

El *Marco Común Europeo de Referencia para las lenguas* (MCER) plantea que aquellas personas encargadas de confeccionar materiales de exámenes y manuales tienen la obligación de hacer una selección léxica (Consejo de Europa, 2002). Entre las opciones que ofrece para cumplimentar esta tarea se encuentra el seguimiento de principios léxico-estadísticos que permitan elegir las palabras más frecuentes mediante recuentos generales y amplios, como los estudios de vocabulario básico, o por medio de una pesquisa de disponibilidad léxica.

Muchos autores coinciden en plantear que la disponibilidad léxica constituye un campo de investigación indispensable para la selección y planificación del vocabulario de la lengua extranjera que se pretende enseñar, pues ayuda a reducir la posibilidad de arbitrariedades en dicho proceso. Según Paredes (2015, p. 7), la disponibilidad léxica constituye la “base empírica sobre el contenido léxico temático que debe incorporarse al currículo”; y como afirma Carcedo (1998a), es “un valiosísimo instrumento de aplicación a la enseñanza de idiomas [...] que proporciona al programador del léxico una base sólida de la que servirse para cumplir eficazmente su cometido” (p. 2).

Aunque internacionalmente son más numerosos los trabajos dedicados a investigar cómo enseñar el vocabulario, algunos también se dedican a definir qué vocabulario enseñar, el cual debería constituir el primer paso. Entre ellos, se pueden destacar aquellos que consideran la disponibilidad léxica como parámetro rector para dicha selección. Al respecto es relevante el de Ángela Bombarelli (2005) o los muy actualizados de Carmela Tomé Cornejo (2015) y Ávila (2016).

Sin embargo, Rodríguez y Muñoz (2009, p. 2) afirman que “[...] a pesar de la evolución que han experimentado las investigaciones sobre disponibilidad léxica en el ámbito hispánico, pocas veces se han sabido aprovechar las nuevas metodologías de análisis léxico y su aplicabilidad a la enseñanza...”. Esta situación no ha variado mucho en nuestros días, y aun cuando en el contexto cubano en los últimos años se han realizado algunas investigaciones descriptivas sobre la disponibilidad léxica (Pérez, 2011; Almeida, 2012; Sierra, 2013; Suardíaz, 2014; Pérez y otros, 2015; Pacheco, 2016; Cárdenas, 2016), ninguna realiza propuestas concretas para la didáctica del español.

Disponibilidad léxica y su aplicación a la enseñanza del vocabulario de ELE en Santiago de Cuba

De los estudios antes citados, solo el trabajo de diploma de Leonor Sierra Salas (2013), cumple con los requisitos básicos para aplicar sus resultados a la enseñanza de ELE. Ello se debe a que se rige por los lineamientos del Proyecto Panhispánico de Disponibilidad Léxica (PPDL): primero, porque utiliza informantes de duodécimo grado; segundo, porque trabaja los 16 centros de interés tradicionales, y tercero, porque sigue fielmente la metodología para el tratamiento de la información primaria.

El objetivo de esa tesis fue caracterizar cuantitativa y cualitativamente el léxico disponible de los estudiantes preuniversitarios del duodécimo grado del municipio Santiago de Cuba. La selección de informantes de duodécimo grado garantizó resultados bastante próximos a la norma comunitaria adulta, ya que eran hablantes formados desde el punto de vista léxico, pero todavía no deformados por estar realizando estudios muy específicos.

Se emplearon las técnicas tradicionales en estos trabajos para la obtención de los datos primarios. En primer lugar, se aplicó un test de disponibilidad léxica sobre 18 centros de interés, a cada uno de los cuales se le dedicó dos minutos, tiempo en el cual cada estudiante debía escribir, sin restricción de cantidad,

todas las palabras que acudieran a su mente relacionadas con el área temática. Estos centros de interés fueron: 1) partes del cuerpo humano, 2) la ropa, 3) la casa (sin los muebles), 4) los muebles de la casa, 5) alimentos, 6) objetos sobre la mesa para comer, 7) la cocina: muebles y utensilios; 8) la escuela: muebles y material escolar, 9) iluminación y aire acondicionado, 10) la ciudad, 11) el campo, 12) medios de transporte, 13) trabajos del campo y del jardín, 14) animales, 15) juegos y diversiones, 16) profesiones y oficios, 17) las artes, 18) los deportes.

Los primeros 16 de ellos se corresponden con los que se han tenido en cuenta tradicionalmente en este tipo de estudios; los dos últimos se añadieron teniendo en cuenta la existencia de informantes provenientes de centros de enseñanza artística y deportiva. Luego se transcribieron, editaron y corrigieron los datos primarios obtenidos siguiendo al PPDL, con vistas a que pudieran ser procesados de manera informática por el programa Lexidisp y tener acceso a los datos estadísticos.

Los resultados obtenidos en Sierra (2013) son de relevancia al tratarse del primer estudio de disponibilidad léxica realizado en Santiago de Cuba, y sentaron bases para su posterior aplicación a otras investigaciones.

Así, en las dos últimas ediciones de la Maestría Enseñanza del Español como Lengua Extranjera que coordinó el departamento de Letras de la Universidad de Oriente, se defendieron dos tesis que se basaron en los listados léxicos de frecuencia obtenidos por Sierra (2013). A continuación se presentan ambas investigaciones y se comentan sus principales aportes a la enseñanza de ELE.

Tesis de maestría La selección y ocurrencias del vocabulario por centros de interés en el manual Español para Todos 1

Con esta tesis, defendida en diciembre de 2013, Haydeé Fernández Leyva, profesora de la Universidad de Oriente, se trazó como objetivo analizar el vocabulario presente en el manual didáctico *Español para Todos I* a partir de criterios léxico-estadísticos, con

el fin de precisar si este y su frecuencia de aparición en el citado manual eran los más indicados para propiciar una mejor adquisición por parte de los estudiantes a los que dirige.

El manual objeto de estudio por dicha autora es creación de un colectivo de profesores del departamento de Idiomas de la Universidad de Oriente. El texto está destinado a estudiantes de origen extranjero que inician sus estudios de idioma español como lengua extranjera en contexto de inmersión sociocultural. El léxico empleado en este manual didáctico (MD) se corresponde con la norma estándar de la variante cubana del español y da a conocer al alumno, por medio de las manifestaciones culturales, algunas palabras propias del hablante santiaguero.

Para el análisis del vocabulario en el referido MD, Fernández (2013) partió de su comparación con el índice léxico obtenido por Sierra (2013). Se tomó esta obra como referencia por ser la única investigación concluida hasta el momento en Cuba que había considerado el estudio de los 16 centros de interés tradicionales en una comunidad preuniversitaria.

El análisis efectuado posibilitó revisar el texto y valorar su nivel de suficiencia en función del resultado final que se pretendía lograr: la mejor apropiación por parte de los estudiantes de español de las palabras de mayor utilidad, en correspondencia con cada tema.

Por estar destinado el manual *Español para Todos I* a los niveles A1 y A2, la autora consideró necesario prestar atención a si ofrece al estudiante las palabras más útiles en dependencia del tema, lo cual guarda estrecha relación con el vocabulario más disponible.

Una palabra disponible es la que, sin ser particularmente frecuente, está siempre lista para ser empleada y viene de manera inmediata y natural a la mente en el momento en que se tiene necesidad de ella. Es una palabra que, al formar parte de las asociaciones de ideas usuales, existe en potencia en el sujeto hablante en cuanto esas asociaciones entran en juego (Carcedo, 1998b, p. 14). Por ello,

para la selección léxica, tienen mayor valor para el alumno las palabras más productivas, lo cual se determina por el alto índice de disponibilidad.

Fernández (2013) destaca que si bien escoger las unidades léxicas apropiadas es importante, también lo es la ocurrencia alcanzada en el MD. Para ello se basa en lo planteado por Benítez y Zebrowski (1991): “El hecho de que una forma se aprenda pero luego no se actualice en los ejercicios y lecturas es un hecho que no favorece el aprendizaje. Esa lexía podrá pertenecer al conocimiento pasivo del estudiante, pero tardará tiempo en pasar a formar parte de su conocimiento activo” (p. 226).

Benítez (1995) reconoce que el MD no es el único medio por el que los estudiantes adquieren las formas léxicas que pasarán a su competencia lingüística, mucho menos si se encuentran en contexto de inmersión sociocultural, pero enfatiza en que a veces son demasiadas para que todas se actualicen las veces necesarias. “Desde un punto de vista subjetivo no parece que las bajas frecuencias de aparición en los libros ayuden al aprendizaje; parece lógico que el aprendizaje de la lengua necesita un mayor número de exposiciones de un vocablo para que este se incorpore al lexicón mental del aprendiz” (Benítez, 1995, p. 78).

Basado en estudios precedentes en lengua materna, este autor aplicó esta metodología a la enseñanza de ELE considerando aproximadamente de seis el número de exposiciones que debe recibir el aprendiz para llegar a aprender el vocabulario nuevo. Esto evidencia la ardua tarea a la que se enfrentan los autores de MD, pues tan solo en cuanto al contenido léxico deben preocuparse por utilizar métodos objetivos y de gran fiabilidad para su selección, así como por la frecuencia en que lo actualizan en el manual.

Siguiendo la metodología empleada por Benítez (1990, 1995) y por Benítez y Zebrowski (1991) para el análisis del léxico en los manuales, lo primero que hizo Fernández (2013) fue realizar el vaciado de los textos para agrupar todos los vocablos relativos a

los 16 centros de interés considerados tradicionalmente en investigaciones de este tipo. Para ello tuvo en cuenta las explicaciones gramaticales, lecturas, ejercicios, índices y diseños gráficos; en fin, todo el cuerpo del manual. Este procedimiento se hizo lematizando las formas (palabras).¹ Sin embargo, como entre los objetivos de Fernández (2013) estaba evaluar la ocurrencia de los vocablos, no olvidó la cuantificación de sus ocurrencias o realizaciones en cada centro de interés. Así, por ejemplo, inventarió *perro* como vocablo y cuantificó la aparición de sus formas flexivas correspondientes: *perro*, *perra*, *perrito*. De este modo, el vocablo *perro* alcanzó tres ocurrencias en el MD, porque cada una de estas formas tuvo una sola frecuencia de aparición.

Luego comparó el registro obtenido con las listas del léxico disponible de los estudiantes preuniversitarios santiagueros. Esto lo hizo para evaluar cuáles de los vocablos por centros de interés que propone el MD debían permanecer o excluirse, así como el número de apariciones que se debía recomendar. Ello también permitió conocer si existía léxico superfluo. Por este término se entiende, siguiendo a Benítez y Zebrowski (1991, p. 227), el conjunto de vocablos utilizados por los autores de un MD que no están presentes en la lista de disponibilidad.

Para saber cuántos vocablos debería tener un centro de interés en el manual, Fernández (2013) tomó en cuenta el promedio de respuestas por sujeto (PR) del estudio de Sierra (2013). El PR se calcula dividiendo el número total de unidades léxicas de un centro de interés (repetidas o no) entre el número total de informantes, e indica el promedio de palabras que poseen los sujetos en el ámbito de conocimiento que explora el centro de interés en cuestión. Así, por ejemplo, tomando en consideración que en *Animales* se produjeron 178 vocablos y alcanzó un PR

¹ Müller (1973) define la lematización como el proceso que consiste en “reunir bajo una misma rúbrica las palabras que representan un mismo vocablo” (p. 254).

de 24,3; la autora propone incluir los 24 vocablos más disponibles del centro.

Fernández (2013) detectó que la cantidad de vocablos diferentes por áreas temáticas en el manual *Español para Todos I* es superior al PR de cada una de estas en Sierra (2013), lo cual representa mediante el gráfico 1.

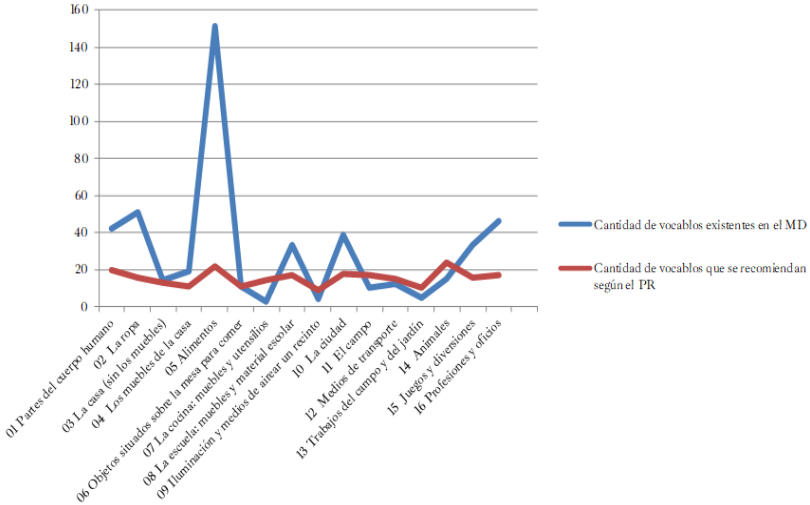


Gráfico 1. Productividad de vocablos por centros de interés en el MD respecto a lo recomendado

Fuente: Fernández (2013)

En cuanto a la cantidad de vocablos diferentes existentes en el manual, con independencia del centro de interés, plantea que su número es excesivo (489). Considera que el criterio de fijar su cantidad a partir del promedio de respuesta por sujeto, si bien no es todavía el recurso ideal, ha demostrado ser eficaz para su reducción por centros de interés. De ahí que proponga para el citado MD un total de 250 vocablos diferentes, lo que está en correspondencia con el PR total obtenido por Sierra (2013) en el estudio léxico de estudiantes preuniversitarios santiagueros. No obstante, puntualiza que esto es solo una sugerencia, ya que los

autores del manual serán los responsables de seleccionar, entre los de mayor índice de disponibilidad, los más adecuados a la situación que abordan.

A continuación, a manera de ejemplo, se muestra la lógica del análisis realizado por Fernández (2013) en dos centros de interés:

01. Partes del cuerpo humano

Este centro de interés cuenta en el manual con 42 vocablos diferentes; de ellos 24 con una ocurrencia, 6 con dos ocurrencias, 3 con tres ocurrencias, 3 con cuatro ocurrencias, 2 con cinco ocurrencias; para un total de 38 vocablos con una cantidad insuficiente, lo que representa el 90,47 % del total. Con ocurrencias por encima de 6 se encontraron los vocablos pelo, estómago, muela y cabeza, lo que representa el 9,53 % del total. Ninguno de ellos constituye léxico superfluo.

Según la autora, el promedio de respuestas por sujeto en este centro de interés es de 19,9, por lo que sugiere eliminar 22 vocablos y enfatizar en los 20 más disponibles. Por esta razón, recomienda que aparezcan los siguientes: cabeza (0,55), pierna (0,54), brazo (0,53), ojo (0,53), pie (0,46), mano (0,45), nariz (0,44), dedo (0,44), boca (0,43), corazón (0,33), oreja (0,30), riñón (0,22), pulmón (0,21), pelo (0,20), uña (0,18), cuello (0,15), oído (0,14), sangre (0,14), estómago (0,13), hígado (0,13). La cifra que aparece entre paréntesis es el índice de disponibilidad léxica alcanzado por cada vocablo en Sierra (2013). De los vocablos anteriores, no se registraban en el manual corazón, riñón, pulmón, cuello, sangre ni hígado.

Una vez seleccionados los vocablos de esta área temática, Fernández (2013) recomienda la ocurrencia de ellos aproximadamente seis veces. Respecto de *cabeza*, que tiene once repeticiones, debe disminuirse su aparición para así dejar cobertura a la ocurrencia de otros.

02 .La ropa

Samper (2002, p. 318) apunta que: “El centro acoge no solo lo relacionado estrictamente con ‘todo género de tela que, con variedad de cortes y hechuras, sirve para el uso o adorno de las cosas’ (DRAE), también se incluyen términos para designar el calzado y otros accesorios (collar, broche, reloj, bolso, gafas)”. Por esta razón, se incluyeron estas formas en la lista de asociaciones.

En el MD este centro de interés contiene 51 vocablos diferentes; de ellos 33 con una ocurrencia, 9 con dos ocurrencias, 2 con tres ocurrencias, 3 con cuatro ocurrencias, y 1 con cinco ocurrencias, para un total de 48 vocablos con un número insuficiente, lo que representa el 94,11 % del total. Con ocurrencias por encima de 6 se encontraron los vocablos zapato, camisa y blusa, lo que representa el 5,89 % del total. Ninguno de ellos constituye léxico superfluo.

En el estudio de Sierra (2013) el promedio de respuestas por sujeto alcanzado en este centro de interés fue de 16,1 por lo que Fernández (2013) sugiere eliminar muchos de los vocablos existentes en el MD. Recomienda la presencia de: pantalón (0,72), short (0,49), blusa (0,48), camisa (0,43), pulóver (0,36), blúmer (0,34), camiseta (0,32), saya (0,32), calzoncillo (0,30), ajustador (0,30), media (0,26), pescadora (0,24), zapato (0,20), chancleta (0,18), vestido (0,17), abrigo (0,13).

Excepto abrigo, todos estos vocablos aparecen en el manual, aunque enfatiza en que debe corregirse su frecuencia de aparición, pues solo tres alcanzan una ocurrencia de 6. Además propone eliminar el resto para no sobrecargar al estudiante de información nueva con menos posibilidades de adquisición.

Por otro lado, aunque reconoce en el MD la presencia de vocablos no pertenecientes a nuestra variante lingüística, como pendientes, calcetines, braga, sujetador y brasier, recomienda mantenerlos por su importancia para la comunicación a nivel internacional, pese a que no aparecen entre los resultados de Sierra (2013).

El análisis realizado por Fernández (2013) ha permitido comprobar la utilidad de los criterios léxico-estadísticos para evaluar la pertinencia de la selección léxica de un MD, lo que posibilita un mejor acercamiento de los estudiantes de ELE al vocabulario de los hablantes nativos. La autora concluye que los vocablos mejor seleccionados en el MD se relacionan con los centros de interés 2) la ropa, 4) los muebles de la casa, 6) objetos situados sobre la mesa para comer, 7) la cocina: muebles y utensilios, 10) la ciudad, 15) Juegos y distracciones. Por otro lado, reconoce mayor cantidad de vocablos innecesarios según el promedio de respuesta por sujeto en los centros de interés 1) partes del cuerpo humano, 5) alimentos y 16) profesiones y oficios.

En cuanto a las ocurrencias, se observa la tendencia a emplear muchos vocablos pocas veces. Así, existe un alto porcentaje de vocablos con ocurrencia inferior a 6 en los centros de interés 15) juegos y distracciones (34,3 %), 5) alimentos (29,14 %), 3) partes de la casa (28,58 %) y 8) la escuela: muebles y material escolar (27,28 %). Esto determina la necesidad de que las ocurrencias de los vocablos en el MD sean revisadas por sus autores.

Tesis de maestría *Selección del vocabulario para la enseñanza de ELE basada en la disponibilidad léxica (nivel B1)*

Con esta tesis, defendida en junio de 2017, Leonor Grethel Sierra Salas se trazó como objetivo elaborar una metodología que permitiera la selección científica del vocabulario que debe ser enseñado en las clases de ELE (nivel B1).

Escogió este nivel, en primer lugar, por ser muy poco tratado en Cuba. La investigación de Fernández (2013) se centra en el nivel A1 al analizar las ocurrencias léxicas en el manual *Español para todos*. Además, en el ámbito internacional tampoco existen muchos trabajos sobre el nivel B1. El alumno de este nivel debe ser capaz de interactuar y de hacerse entender en diversas situaciones habituales y de enfrentarse de forma flexible a los problemas cotidianos.

A partir de encuestas realizadas a profesores de la Universidad de Oriente y la Facultad de Ciencias Médicas II de Santiago de Cuba, el análisis del manual en uso y del programa del nivel B1, Sierra (2017) pudo constatar que en la actualidad la selección del vocabulario que se enseña en las clases de ELE se realiza de forma subjetiva, haciéndola depender exclusivamente de las decisiones de profesores y alumnos, lo que no garantiza que las unidades léxicas sean representativas y útiles al aprendizaje de ELE, ni que reflejen la norma léxica de la comunidad en la que este se inserta. Ello también se verifica en la inadecuación de parte del vocabulario presente en el libro de texto con respecto a los resultados de los escasos estudios sobre disponibilidad léxica efectuados en Cuba.

La metodología empleada por Sierra (2017) es una propuesta flexible, que no responde a una delimitación objetiva del vocabulario, basada solo en “frías” cotas cuantitativas, sino que incorpora otros elementos de juicio en aras de una “corrección”. Según Paredes (2015), el predominio de la selección objetiva, pero con una complementación de criterios subjetivos, es el proceder más adecuado para una efectiva selección léxica.

De forma individual, la selección objetiva podría no ajustarse a las necesidades de los alumnos y al contexto de enseñanza-aprendizaje, y en el caso de la selección subjetiva, sería demasiado riesgoso dejar dicha elección a la intuición del profesor o programador, lo cual propiciaría diferencias significativas entre los materiales y textos de enseñanza. Por tanto, Sierra (2017) plantea que tomar los aspectos positivos de ambas ofrecerá los resultados más favorables.

Sierra y Pérez (2018a, p. 16), llegan a la conclusión de que la disponibilidad léxica suple las deficiencias de la frecuencia y revela, a partir de la garantía de un corpus obtenido con la aplicación de una metodología rigurosa, el léxico de uso real de la comunidad de habla donde se inserta el aprendiz de ELE, así como su grado de empleo.

A continuación explican los pasos de la metodología propuesta en Santiago de Cuba, que permite una adecuada selección del léxico para la enseñanza del español en el nivel B1:

1) Efectuar un estudio de disponibilidad léxica en hablantes nativos de la comunidad donde se inserta el aprendiz de ELE

Sierra (2017) partió del estudio de disponibilidad léxica realizado a estudiantes preuniversitarios santiagueros (Sierra, 2013), ya explicado con anterioridad, considerando su pertinencia para revelar con carácter objetivo el estado léxico de la comunidad de habla donde se inserta el aprendiz de ELE (Benítez, 2009 y Tomé, 2015), en este caso, Santiago de Cuba.

2) Efectuar un estudio de disponibilidad léxica en los aprendices de ELE

Se realiza un estudio del léxico de español disponible de los estudiantes de ELE en Santiago de Cuba, lo que posibilitaría corregir y complementar la selección primaria que realizaría a partir de los resultados obtenidos con los nativos santiagueros.

Para desarrollar el estudio de disponibilidad léxica de los aprendices, se siguieron los mismos presupuestos metodológicos contemplados para la investigación con hablantes nativos, solo que con algunas especificidades en ELE. La muestra estuvo constituida por un total de 50 estudiantes de ELE que asistieron a los cursos impartidos por la Universidad de Oriente y la Facultad de Ciencias Médicas II, en Santiago de Cuba, entre octubre y noviembre del 2015, lo que representa un 98 % de la matrícula.

La depuración de la información obtenida tras la aplicación de la prueba asociativa es un paso esencial para que los listados finales que se alcancen sean el reflejo fiel de la realidad de los hablantes y se encuentren aptos para futuras comparaciones. Esta etapa es compleja ya que, “al tratarse de informantes no nativos, nos encontramos con fenómenos tales como la interferencia de la lengua materna de los hablantes o la semejanza de una unidad determinada a una palabra española sin llegar a pertenecer sin embargo a nuestro idioma” (Samper, 2002, p. 2).

Al igual que en el estudio realizado con nativos, el programa informático Lexidisp posibilitó el cálculo del índice de disponibilidad (ID), el porcentaje de aparición y la frecuencia relativa.

Como se ha podido apreciar, el primer y segundo pasos son de gran importancia, pues constituyen la base de esta metodología.

3) Elegir las palabras en ambos listados, que tienen un $ID \geq 0,05$ y han sido mencionadas por al menos el 10% de los informantes, para ser enseñadas en el nivel B1

El ID, valor que revela los vocablos de mayor empleo en la comunidad de habla estudiada, fue el parámetro que rigió la elección, auxiliado por el distribución social. Al consultar la literatura científica no se pudieron encontrar parámetros o criterios de corte que definieran qué léxico elegir para el nivel B1, por lo tanto, la autora decidió establecer estas cotas.

Tras un detallado análisis cualitativo y valorando los criterios que se han tenido en cuenta para la selección del vocabulario en niveles inferiores, pudo determinar que los vocablos que se encuentran por encima de los valores propuestos son los representantes prototípicos de cada centro de interés. Por tanto, no deberían, debido a su utilidad, ser obviados de la planificación léxica destinada a estudiantes del nivel B1, los cuales deben llegar a ser capaces de comunicarse de forma independiente en un espectro de situaciones variadas.

Por debajo de estos valores, los vocablos empiezan a perder representatividad al alejarse del núcleo semántico del centro de interés en cuestión, y pueden ser incluidos en niveles superiores (B2 y C1) porque son menos disponibles, esto es, producidos por menos informantes en posiciones más bajas en los listados, producto de asociaciones más individuales. Pero esto no es una propuesta inflexible, no significa que no pueda hacerse uso de otros vocablos que aparecen en posiciones inferiores o no se registran en los listados, en dependencia de los intereses o situación de aprendizaje (Sierra y Pérez, 2018a).

4) Comparar los resultados de ambos estudios de disponibilidad aplicando este criterio de corte

Este paso es significativo, pues en él se valora también el léxico más disponible entre los extranjeros. Una vez definidos los vocablos con un $ID \geq 0,05$ y mencionados por al menos el 10 % de los informantes, tanto en el listado resultante de los hablantes santiagueros como en el de los extranjeros, se procedió a comparar cuáles fueron mencionados por los nativos y no por los extranjeros, que serían los que se deben seleccionar para este nivel en primer lugar.

Luego, en el listado de los extranjeros, se analiza si en este índice de corte hubo vocablos que estos incluyeron, pero que los hablantes nativos no. Y se valora si esta(s) unidad(es) léxica(s) altamente disponible(s), puede(n) ser incluida(s) también en la selección.

5) Contrastar los resultados con la propuesta del Plan Curricular del Instituto Cervantes (PCIC) para concluir la selección

Dado que el *PCIC* constituye una obra de obligada consulta en el campo de la enseñanza de lenguas y es el punto de partida en la actualidad de toda iniciativa relacionada con el diseño y desarrollo de cursos y materiales de ELE, se consideró valorar la propuesta de selección del *PCIC* en el nivel B1. No obstante, se han observado para el nivel de referencia en cuestión unidades léxicas que no se corresponden con nuestra realidad lingüística, como ocurre con *braga* y *vaqueros* en el centro de interés la ropa. En ese caso, se incorporó el equivalente en nuestra área, *blúmer* y *pitusa*, respectivamente.

Por tanto, la propuesta de selección léxica realizada termina por definirse con aquellos vocablos mencionados en el *PCIC* que no aparecen en las listas de disponibilidad léxica considerando la cota establecida (vocablos con $ID \geq 0,05$ y mencionados por al menos el 10 % de los informantes); esto, por supuesto, si ameritaran su elección.

Sierra y Pérez (2018a, p. 15) plantean que la metodología propuesta responde a una selección objetiva corregida, ya que combina las formas de selección subjetiva y objetiva. Se apoya en métodos léxico-estadísticos, pero matiza dichos resultados con los juicios y criterios personales del programador o profesor, es decir, realiza un ajuste posterior de naturaleza subjetiva. El programador, según sus intereses particulares, podrá considerar la eficacia de la palabra (si es empleada en muchos contextos), su productividad (si da lugar a otras siguiendo los dictados de la formación de palabras en una lengua), y su universalidad (en el caso del español, las voces más generalmente utilizadas en todos los países hispanos).

Sierra (2017) realiza por vez primera una propuesta científica concreta, basada en la disponibilidad léxica de hablantes nativos y extranjeros, sobre el vocabulario que se debe enseñar en el nivel B1 en Cuba, de ahí la importancia de la investigación y de los aportes metodológicos y prácticos realizados, como el que se explicará a continuación.

Diccionario temático comparativo de léxico disponible para la enseñanza de ELE

Como resultado de las dos investigaciones de disponibilidad léxica ya referidas, Sierra y Pérez (2018b) elaboraron un diccionario temático contrastivo entre el léxico disponible de los estudiantes nativos santiagueros y aquellos que estudian el ELE en la ciudad. Este recoge todos los vocablos obtenidos por centro de interés o área temática, no solo los incluidos en el corte para el nivel B1.

Esta obra lexicográfica se realizó con el fin de proporcionar a profesores y estudiantes de ELE en Cuba un material de consulta y apoyo para el perfeccionamiento de la enseñanza-aprendizaje del nivel lexical a partir del uso real efectuado por la comunidad de habla cubana y los estudiantes extranjeros. Además, con el propósito de ofrecer una base científica fiable para la selección léxica en otros niveles de español, ya que la metodología de tra-

bajo en la disponibilidad léxica garantiza el reflejo del léxico más representativo y empleado por área temática.

Este aporte práctico de 80 páginas se estructura en los 16 centros de interés empleados tradicionalmente en las investigaciones de disponibilidad léxica. Cada una de estas áreas temáticas consta de un conjunto de unidades léxicas, de las que se ofrecen los siguientes valores:

1. **Disponibilidad:** en la primera columna, los vocablos se encuentran ordenados de forma descendente por su índice de disponibilidad léxica. Como se explicó antes los vocablos que ocupan las posiciones superiores resultan los nucleares de la categoría y, por ello, los más indicados para considerados en los manuales y programas.
2. **Frecuencia:** resulta de la suma de la frecuencia relativa de cada palabra. Muestra las unidades más frecuentes en relación con el tema correspondiente. Estos valores se hallan en la segunda columna.
3. **Porcentaje de aparición:** ofrece el porcentaje de informantes que han anotado la palabra en su cuestionario. Se encuentra en la tercera columna.

Para tener una representación de la apariencia del *Diccionario*, la tabla 1 muestra a un pequeño fragmento.

Tabla 1. Fragmento del Diccionario temático comparativo de léxico disponible para la enseñanza de ELE

CI 01 EL CUERPO HUMANO								
ESTUDIANTES NATIVOS SANTIAGUEROS					ESTUDIANTES DE ELE			
No	Palabra	DL	Frec.	% Aparición	Palabra	DL	Frec.	% Aparición
1.	cabeza	0,553911	3,296704	65,625	cabeza	0,829165	6,349206	88
2.	pierna	0,549561	3,610675	71,875	ojo	0,6484971	5,772006	80
3.	brazo	0,538978	3,610675	71,875	mano	0,5544521	5,339106	74
4.	ojo	0,537961	3,453689	68,75	boca	0,4934864	5,339106	74
5.	pie	0,467335	3,610675	71,875	brazo	0,4675446	5,339106	74

Las potencialidades de uso de este diccionario, según destaca Sierra (2017, p. 66), son bastante amplias, ya que sirve:

- a) Para la selección léxica en otros niveles, valorando y justificando otros criterios de corte y gradaciones.
- b) Para una exhaustiva comparación cuantitativa y cualitativa del vocabulario de hablantes nativos santiagueros y extranjeros, evaluando su competencia léxica.
- c) Para el trabajo con los términos más empleados y prototípicos de cada categoría, así como la elaboración de ejercicios.
- d) Para formular hipótesis acerca de cómo se organizan y almacenan las palabras en el lexicón mental de los hablantes bilingües, cómo se recuperan y actualizan. Al comparar los listados de los hablantes nativos con los no nativos se puede interpretar si los mecanismos de asociación coinciden o divergen.
- e) Para detectar las deficiencias y errores léxicos, interferencias de la lengua materna, así como los rasgos culturales presentes en el léxico producido.
- f) Para tomarse como referente en la evaluación de los materiales didácticos existentes en nuestro contexto de enseñanza.

En el caso de los ejercicios, sugiere los siguientes: la lectura de los listados por parte del alumno, la explicación del significado de las palabras desconocidas, la clasificación gramatical de las unidades, la formación de frases o textos con las unidades léxicas, la agrupación de las unidades léxicas contenidas en los listados de acuerdo con criterios semánticos, la elaboración de mapas conceptuales con las unidades léxicas actualizadas.

También se puede realizar individualmente la prueba y contrastar los resultados particulares con los generales, seleccionar unidades de diferentes centros de interés y pedir a los alumnos que los encuadren en el tema correspondiente, dada una unidad léxica, identificar a qué centro o centros de interés puede corresponder, localizar unidades que se repiten en diferentes centros de interés y observar su diferente grado de disponibilidad (Paredes, 2010).

El estudio de disponibilidad léxica realizado por Sierra (2013) sirve de base a las dos tesis de maestría que tratan de la selección del léxico que debe ser enseñado. Por su parte, la tesis de Fernández (2013) aportó un valioso análisis sobre la pertinencia del vocabulario presente en el manual didáctico *Español para Todos I*, y una propuesta concreta del léxico que debe contener y cómo debe ser tratada su frecuencia de aparición en el manual.

La tesis de Sierra (2017) aportó no solo una metodología para la selección del léxico que se debe enseñar en el nivel B1, sino también un diccionario de frecuencias contrastivo, útil en cualquier nivel para la selección del léxico de ELE que se debe enseñar. Esto permite concluir que, en Santiago de Cuba, los estudios léxico-estadísticos han realizado importantes aportes a la selección léxica para la enseñanza del español como lengua extranjera en Cuba.

Referencias bibliográficas

- ALMEIDA, A. (2012). *Disponibilidad léxica en escolares de La Habana*. (tesis inédita de diploma). Facultad de Artes y Letras, Universidad de La Habana, La Habana.
- ÁVILA, A. M. (2016). El léxico disponible y la enseñanza del español. Propuesta de selección léxica basada en la teoría de los conjuntos difusos. *Journal of Spanish Language Teaching*. Recuperado de <http://dx.doi.org/10.1080/23247797.2016.1163038>.
- BARTOL, J. A. (2010). Disponibilidad léxica y selección del vocabulario. En Castañer Martín, R. M. y Laguéns Gracia, V. (eds.), *De moneda nunca usada. Estudios dedicados a José M. Enguita Utrilla* (pp. 85-107). Zaragoza: Institución Fernando el Católico.
- BENÍTEZ, P. (1990): Léxico real/irreal en los manuales de español como lengua extranjera (pp. 225-233). *Actas del Segundo Congreso Nacional de ASELE*. Madrid.

- BENÍTEZ PÉREZ, P.(1990). Léxico real / irreal en los manuales de español como lengua extranjera. En *Actas del Segundo Congreso Nacional de ASELE* (pp.225-233). Madrid, España.
- BENÍTEZ, P. (2009). ¿Qué vocabulario hay que enseñar en las clases de español como lengua extranjera? *MarcoELE*, 8.
- BENÍTEZ, P.Y ZEBROWSKI, J. (1991). El léxico español en los manuales polacos. *AELE Actas III. Centro Virtual Cervantes* (pp. 223-230). Recuperado de http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/03/03_0221.pdf
- BOMBARELLI, Á. (2005). *La disponibilidad léxica como herramienta didáctica: una propuesta de selección del vocabulario para un nivel umbral de ELE*. (memoria Máster). Universidad de Salamanca, España.
- CARCEDO, A. (1998a). Sobre las pruebas de disponibilidad léxica para estudiantes de español/LE. *Rilce*, 14(2), 205-224.
- CARCEDO, A. (1998b). Tradición y novedad en las aportaciones hispánicas a los estudios de disponibilidad léxica. *Lingüística*, año 10, 5-68.
- CÁRDENAS, S. (2016). *La disponibilidad léxica en los estudiantes de oncenno grado del IPVCE Ernesto Guevara en la provincia de Villa Clara*. (tesis inédita de diploma). Universidad Central Martha Abreu de Las Villas, Villa Clara, Cuba.
- COLECTIVO DE AUTORES (2012). *Español para Todos I*. Universidad de Oriente, Santiago de Cuba. (Inédito).
- CONSEJO DE EUROPA (2002). *Marco Común Europeo de Referencia para las lenguas: aprendizaje, enseñanza, evaluación*. España: Ministerio de Educación, Cultura y Deporte, Madrid. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf
- FERNÁNDEZ, H. (2013). *La selección y ocurrencias del vocabulario por centros de interés en el manual Español para Todos I*.

- (tesis inédita de maestría). Universidad de Oriente, Santiago de Cuba, Cuba.
- INSTITUTO CERVANTES (2018). *Plan curricular*. Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/default.htm
- MÜLLER, C. (1973). *Estadística lingüística*. (trad. Arturo Amítes). Madrid: Editorial Gredos.
- PAREDES, F. (2015). Disponibilidad léxica y enseñanza de ELE: el léxico disponible como fuente curricular y como recurso en el aula. *LinRed, Lingüística en la red*, número monográfico. Recuperado de http://www.linred.es/monograficos_pdf/LR_monografico13-articulo2.pdf
- PACHECO, C. R. (2016). *Disponibilidad léxica en jóvenes pinaresños*. (resultados de proyecto de investigación inéditos). Universidad de Pinar del Río, Pinar del Río, Cuba.
- PÉREZ, C. M. (2011). Disponibilidad léxica de los escolares de Guamá en tres centros de interés. *Comunicación Social en el siglo XXI* (pp. 80-84). Santiago de Cuba: Centro de Lingüística Aplicada.
- PÉREZ, C. M. ET ALL. (2015). Proyecto Vocabulario disponible de los estudiantes de noveno grado en Santiago de Cuba: un monitoreo de la efectividad del sistema educativo cubano desde la léxico-estadística. *Comunicación Social. Retos y Perspectivas* (pp. 63-67). Santiago de Cuba: Centro de Lingüística Aplicada.
- RODRÍGUEZ, F. J. Y MUÑOZ, I. (2009). De la disponibilidad a la didáctica léxica. *Tejuelo*, 4. Recuperado de <http://mascvuex.unex.es/revistas/index.php/tejuelo/article/view/2336/1499>
- SAMPER, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera*. (Colección Monografías no. 4). Málaga: ASELE.

- SIERRA SALAS, L. G. (2013). *Disponibilidad léxica de estudiantes preuniversitarios santiagueros*. (tesis inédita de diploma). Universidad de Oriente, Santiago de Cuba, Cuba.
- SIERRA SALAS, L. G. (2017). *Selección del vocabulario para la enseñanza de ELE en Santiago de Cuba basada en la disponibilidad léxica (nivel b1)*. (tesis inédita de maestría). Universidad de Oriente, Santiago de Cuba, Cuba.
- SIERRA SALAS, L. Y PÉREZ, C. M. (2016). Disponibilidad léxica de estudiantes de español como lengua extranjera en Santiago de Cuba. *Santiago*, 141, 770-782.
- SIERRA SALAS, L. Y PÉREZ, C. M. (2018a). La selección léxica en el español como lengua extranjera. *Maestro y Sociedad*, 15(1), 3-17.
- SIERRA SALAS, L. G. Y PÉREZ, C. M. (2018b). *Diccionario temático comparativo de léxico disponible para la enseñanza de ELE*. Universidad de Oriente, Santiago de Cuba (inédito).
- SUARDÍAZ, L. (2014). *Léxico disponible de los escolares habaneros y matanceros en dos centros de interés: descripción, análisis y comparación*. (tesis inédita de diploma). Universidad de La Habana, Cuba.
- TOMÉ, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE*. (tesis doctoral). Universidad de Salamanca, España.