

Índices de riqueza léxica en estudiantes de 11 y 16 años de La Habana

Study of lexical richness in students in Havana between the ages of 11 and 16 years

Laimí Álvarez Martín^{1*}

¹Facultad de Artes y Letras, Universidad de La Habana, Cuba.

*laimi.alvarez88@gmail.com

Resumen

El presente trabajo desarrolla una de las líneas de investigación impulsadas por el proyecto internacional «La enseñanza del español en el Mundo Hispánico». Se concentra en el área de los estudios de riqueza léxica y, particularmente, expone y analiza los resultados obtenidos a partir de la aplicación de las pruebas de riqueza a una muestra de textos de escolares de 11 y 16 años de La Habana. Si bien esta investigación responde a los presupuestos del proyecto, su interés ha ido más allá en la medida en que establece un análisis crítico de esos presupuestos. Por otra parte, se ocupa de demostrar la importancia del estudio y el reconocimiento de las locuciones para la determinación de los índices de riqueza léxica.

Palabras clave: riqueza léxica, índices de riqueza léxica, palabras, vocablos, locuciones.

Abstract

This is a work on one of the research lines developed by the international project «The teaching of Spanish in the Spanish-speaking world». Its main area of study is the study of lexical richness. In this paper the results of a lexical richness test applied to a corpus of texts written by a sample group of 11 and 16 year old school students in Havana is presented and analyzed. The work not only meets the requirements of the project but also includes a review of the given requirements. The research also aims to show the importance of the study of idioms and the ability to recognize them in order to determine the indexes of lexical richness.

Keywords: Lexical Richness, Indexes of Lexical Richness, Words, Lexemes, Idioms.

Recibido: 14/07/2018

Aceptado: 03/09/2018

«La enseñanza del español en el Mundo Hispánico»¹ es un proyecto internacional creado con el fin de evaluar el estado actual de la enseñanza de la lengua y proponer, a partir de los resultados de las investigaciones, soluciones posibles que, de alguna manera, ayuden a corregir las deficiencias encontradas. Nuestro país está integrado al proyecto desde los inicios de su creación. Una de sus áreas de estudio es el análisis de la riqueza léxica de los textos escolares. Las pruebas de riqueza constituyen una fuente de información valiosa para los especialistas de la lengua, porque a través de ellas se puede determinar el desarrollo cuantitativo del léxico de los estudiantes y la influencia en él de determinadas variables. El propósito esencial es incrementar el fondo léxico con que cuentan los educandos; y, una vez conocido el vocabulario que utilizan, podrían buscarse soluciones alternativas que contribuyan a enriquecer ese caudal léxico.

El haber trabajado particularmente en el área de los estudios de riqueza,² y haber reconocido las especificidades y complejidades que supone, así como su utilidad para el mejoramiento de la enseñanza de la lengua, son las motivaciones fundamentales de la realización de este estudio.

Este trabajo tiene como objetivos generales analizar la riqueza léxica de una muestra³ de textos escritos por escolares de 11 y 16 años de La Habana, describir su comportamiento cuantitativo y ofrecer un acercamiento parcial desde el punto de vista cualitativo.

Como objetivos específicos me propongo describir el procedimiento de análisis del indicador de medición que se utiliza –porcentaje de vocablos de contenido nocional–, según la primera y segunda variantes metodológicas propuestas por la directiva del proyecto; compararlas y demostrar que la primera es la más acertada para realizar el análisis de la riqueza de los textos; presentar los resultados del análisis de la muestra, de acuerdo con una y otra metodologías; comparar los índices de riqueza léxica atendiendo a las variables «edad» y «tipo de escuela»; establecer una comparación entre los resultados de los análisis realizados según la primera variante metodológica y los

¹ Este Proyecto comenzó en 2007. Su dirección está representada por el investigador Humberto López Morales, natural de Puerto Rico.

² A partir de aquí al referir el término «riqueza» se estará haciendo alusión a la «riqueza léxica».

³ Más adelante, en «Exposición y análisis de los resultados» (primera parte) aparecen referidas las características de la muestra.

resultados obtenidos de acuerdo con la segunda; y demostrar la importancia del estudio y el reconocimiento de las locuciones para determinar los índices de riqueza.

Esta investigación se plantea tres hipótesis fundamentales: los estudiantes de las escuelas «buenas» tienen mayor riqueza léxica que los alumnos de las escuelas «malas», la riqueza léxica crece según aumenta la edad y la riqueza léxica resultará mayor mientras menor sea el uso de locuciones.

Algunos lingüistas como Lidia Cuba (2006) le han otorgado importancia al estudio de las locuciones dentro de las investigaciones de riqueza, pero ninguno hasta hoy ha realizado un trabajo que considere este aspecto en todos sus matices, lo cual podría conferirle cierta originalidad a este estudio.

A los efectos de esta investigación, «riqueza léxica» es entendida como «el grado de madurez y de dominio léxico general de un estudiante, medidos en su producción escrita por el porcentaje de vocablos de contenido nocional».⁴

Estudios de esta naturaleza se han realizado pocos. En el ámbito hispanoamericano destacan algunos trabajos como el de Humberto López Morales (1984), en el que se analiza la riqueza léxica a partir de cálculos empíricos; y los estudios del mexicano Raúl Ávila (1988-1989), quien la evalúa mediante procedimientos comparativos. Conocemos, además, el trabajo de Gloria E. Báez Pinal (1989), que utiliza un corpus de textos producido por alumnos de cuarto y sexto grados del Distrito Federal de México para llevar adelante su investigación. Por su parte, la dominicana Ana Margarita Haché de Yunén (1991) trabaja la riqueza con textos de escolares de sexto y octavo grados de algunas escuelas primarias de su país.

En Cuba no existían investigaciones sobre riqueza léxica hasta que, en 2001, la doctora Lidia Cuba Vega presentó un estudio del tema en su tesis de maestría. Ya en 2006, Cuba Vega adquirió el grado de Doctora en Ciencias Filológicas por el trabajo «Índices de riqueza léxica en una muestra de escolares de primaria de Ciudad de La Habana», que constituye la continuación de su tesis anterior. En 2004, la investigadora Celia María Pérez Marqués presentó «Nuevo enfoque para un diagnóstico del desarrollo léxico en escolares de primaria». En este estudio la riqueza léxica es entendida como uno de los indicadores de la «actuación léxica»⁵ de los estudiantes. Posteriormente, Pérez Marqués (2005) investiga el «Impacto de los Programas de la Revolución

⁴ Para conceptualizar la riqueza léxica hemos partido de la definición que propone la Dra. Lidia Cuba en su tesis de doctorado: «el grado de madurez y dominio léxico general –competencia léxica– de un hablante, medidos en la producción de sus textos, orales o escritos, por índices estadísticos como el porcentaje de vocablos, el intervalo de aparición de palabras de contenido nocional y el porcentaje de vocablos de contenido nocional» (Cuba Vega, 2006, p. 31).

⁵ La «actuación léxica» se define como la manifestación, en la producción de textos orales o escritos, del grado de madurez y dominio léxico general de un hablante. Constituye una manifestación de la competencia léxica de dicho hablante y puede ser medida, en su mayor parte, mediante procedimientos estadísticos (Pérez Marqués, 2011).

Educacional en el desarrollo léxico de los escolares del municipio Guamá», que resume el informe del proyecto territorial I+D «Medición del impacto de los Programas de la Revolución en el desarrollo léxico de los escolares guamenses», desarrollado entre 2005 y 2008 en Santiago de Cuba.

Estos son, a grandes rasgos, los referentes teóricos más importantes de esta investigación.

El presente trabajo contribuiría modestamente a ampliar la bibliografía sobre riqueza léxica existente en nuestro país; si se considera que, hasta el presente, los estudios realizados en Cuba solo han trabajado con un corpus de textos de escolares de primaria. La muestra que se analiza en esta investigación también incluye textos de estudiantes de preuniversitario.

Aspectos metodológicos

Selección de la muestra y definición de las variables independientes

Características de la muestra

En este estudio se trabaja con una muestra intencionada, no con una aleatoria ni empírica por cuotas proporcionales. El propósito no es el de exponer unos datos rigurosamente representativos –en el plano estadístico–, sino el de mostrar unas tendencias que indiquen con suficiente claridad la realidad en que nos encontramos (López Morales, 2008).

Para ello se seleccionaron dos escuelas primarias de ciudad de La Habana: Vo Thi Tang, situada en el municipio Playa, y Tomás Alba Edison, de 10 de Octubre, a las que se ha denominado primaria A «buena» y primaria B «mala», respectivamente; y dos preuniversitarios: el Instituto Preuniversitario Vocacional de Ciencias Exactas (IPVCE) Vladimir I. Lenin y el Instituto Preuniversitario en el campo (IPUEC) Cuba socialista, nombrados el primero Preuniversitario A «bueno» y el segundo B «malo». Las categorías «bueno» / «malo» fueron otorgadas por el Ministerio de Educación (MINED) y responden a la valoración general de la comunidad a la que pertenecen los centros escolares elegidos.

De cada una de las escuelas se escogieron dos grupos de estudiantes para que realizaran composiciones de «tema libre». Del total de redacciones fueron seleccionados al azar 20 textos por escuela. Para obtener así una muestra de 80 composiciones: 40 escritas por alumnos de 11 años de la enseñanza primaria y 40 de estudiantes preuniversitarios de 16 años de edad.

Variables independientes

Tipo de escuela

Se tuvieron en cuenta dos tipos fundamentales de escuela: «buena» (A) y «mala» (B). Al primer tipo corresponden la escuela primaria Vo Thi Tang, del municipio Playa, y el IPVCE Vladimir I. Lenin; al segundo tipo, la escuela Tomás Alba Edison, de 10 de Octubre, y el IPUEC Cuba socialista.

Edad

Trabajamos con dos grupos etarios: 11 y 16 años. El primero atañe a los escolares de sexto grado de la enseñanza primaria y el segundo a los de oncenno grado de preuniversitario.

Porcentaje de vocablos de contenido nocional

Llevar adelante un análisis de esta naturaleza implica seguir un riguroso método de trabajo, que suele variar en dependencia del indicador que se haya escogido para medir la riqueza. Dentro del marco de estos estudios se han utilizado los cuatro indicadores siguientes:

1. porcentaje de vocablos del total de palabras,
2. frecuencia acumulada de vocablos,⁶
3. intervalo de aparición de palabras de contenido nocional y
4. porcentaje de vocablos de contenido nocional.

En el proyecto «La enseñanza del español...» se ha trabajado este último índice y de dos maneras distintas: estableciendo la proporción entre vocablos nocionales y total de palabras,⁷ y entre vocablos de contenido nocional y total de vocablos del texto.⁸ Estas variantes desarrollan procedimientos de análisis semejantes pero, al mismo tiempo, divergentes. La tabla 1 representa un cuadro comparativo que ayudará a reconocer los aspectos comunes y diferenciadores.

Tabla 1. Variantes metodológicas del procedimiento de análisis

⁶ Raúl Ávila (1989) denomina a este indicador «frecuencia acumulada de tipos».

⁷ Esta variante es la empleada por Gloria E. Báez y Lidia Cuba para medir la riqueza léxica a través del indicador del porcentaje de vocablos de contenido nocional.

⁸ Se define «palabra» como la unidad lexical y semántica que puede ser simple o compleja en dependencia de su construcción gráfica. «Palabras de contenido nocional» (o nocionales) son aquellas que portan un contenido léxico: sustantivos, adjetivos, verbos, adverbios y las unidades léxicas complejas que tienen esas mismas funciones. (Adaptado de Cuba Vega, 2006, p. 48). «Los vocablos se definen como esas formas que se lematizan en un repertorio lexicográfico (*niña* no, sino *niño*; *comió* tampoco, sino *comer*)» (Dominguez Hernández, 2011).

Primera variante metodológica	Segunda variante metodológica
<p>1. Realizar un conteo inicial de palabras.⁹ (Los verbos reflexivos de forma se consideran una sola pieza léxica [<i>alegrarse</i>]).</p>	<p>1. Realizar un conteo inicial de palabras. (Los verbos reflexivos de forma se contabilizan como dos palabras [<i>alegrar/se</i>]).</p>
<p>2. Descontar las siglas, números o sustantivos propios. (Esta operación deja establecido el total de palabras sobre el cual se determina el porcentaje de riqueza).</p>	<p>2. Descontar las siglas, los números o los sustantivos propios.</p>
<p>3. Delimitar las palabras de contenido nocional. (No se reconocen como nocionales los adverbios relativos o interrogativos ni los pronombres).</p>	<p>3. Delimitar las palabras de contenido nocional. (Todos los adverbios –incluidos los relativos e interrogativos– son reconocidos como palabras nocionales. Los pronombres ordinales en función adjetiva también se consideran palabras de contenido nocional).</p>
<p>4. Delimitar los vocablos nocionales del texto.</p>	<p>4. Delimitar los vocablos nocionales y no nocionales, para luego sumarlos y así obtener el total de vocablos del texto sobre el cual se habrá de determinar el índice de riqueza léxica.</p>
<p>5. Establecer la proporción entre vocablos de contenido nocional y el total de palabras del texto.</p>	<p>5. Establecer la proporción entre los vocablos de contenido nocional y el total de vocablos del texto.</p>

La diferencia fundamental entre las dos variantes radica en la quinta operación. La primera establece la proporción entre vocablos nocionales y total de palabras, la segunda entre vocablos nocionales y total de vocablos del texto. Esta divergencia influye de manera determinante en el resultado final del análisis. El porcentaje de riqueza obtenido de acuerdo con la segunda variante metodológica será mayor, sin dudas, que el índice obtenido según la primera, ya que en un texto siempre aparecerá menor cantidad de vocablos que de palabras.

El procedimiento de análisis de cada una de las variantes se explica mejor en el siguiente texto de un alumno de la escuela primaria «buena».

1101A⁹

«Un viaje a Marte»¹⁰

El domingo **estaba** en mi casa **leyendo**¹¹ el periódico y desayunando, y **al poco rato** recibo una llamada de que si quería / ir¹² al planeta Marte.¹³ Acepté y me vestí y me fui.

Al llegar al lugar me sorprendí por la gran belleza del lugar. Me recibieron muy bien y me dijeron que me sentara en el salón de espera, a esperar al jefe. Cuando¹⁴ llegó me dijo que entrara a su oficina y me dijo que la misión era para ver si los extraterrestres **iban a atacar** a nuestro planeta. Entonces me dijo que me vistiera con un extraño traje y después entré en la nave y me asombré pues qué bonita era la nave /¹⁵ y la prendí y partí. [...]

Análisis según la primera variante metodológica de «Un viaje a Marte»

- Primer conteo de palabras: 106
- Segundo conteo de palabras (sin siglas, números o sustantivos propios): 105
- Palabras de contenido nocional: 50
- Vocablos de contenido nocional: 38

⁹ La numeración de los textos responde a la siguiente ordenación: edad (11 años / 16 años), numeración del estudiante (01-20), tipo de escuela («buena»: A; «mala»: B).

¹⁰ La directiva del proyecto determinó que los títulos de los textos no se consideraran en los análisis.

¹¹ El subrayado indica las locuciones que aparecen en el texto.

¹² Recordemos que los elementos que integran las perífrasis verbales que tienen la estructura *querer + infinitivo* se contabilizan como palabras separadas de acuerdo con los presupuestos del proyecto.

¹³ Los sustantivos propios no se reconocen como palabras de contenido nocional, son descontados desde la segunda operación del procedimiento.

¹⁴ La segunda variante metodológica sí considera a los adverbios relativos como palabras nocionales, por ello en el segundo análisis reconocemos el adverbio temporal *cuando* como «una palabra más» de contenido nocional.

¹⁵ Aquí interrumpimos el conteo de palabras. Como se ha dicho, el corte cuantitativo no debe sobrepasar las primeras 115 palabras del texto.

- Índice de riqueza léxica: 36.19 %

Análisis según la segunda variante metodológica de «Un viaje a Marte»

- Primer conteo de palabras: 112¹⁶
- Segundo conteo de palabras (sin siglas, números o sustantivos propios): 111
- Palabras de contenido nocional: 51
- Vocablos de contenido nocional: 39
- Vocablos no nocionales: 23
- Total de vocablos: 62
- Índice de riqueza léxica: 62.90 %

La directiva internacional del proyecto dejó establecida de manera oficial la segunda variante metodológica. Los principales investigadores sugirieron una nueva manera de calcular el porcentaje de riqueza y modificaron, además, algunos de los presupuestos de la primera metodología. Algunas de estas modificaciones resultaron ventajosas para la investigación. Por ejemplo, se empezaron a considerar los adverbios relativos e interrogativos como palabras de contenido nocional, lo cual responde al criterio de algunos lingüistas como López Morales (1984) y Lidia Cuba (2006), quienes consideran todos los adverbios como palabras nocionales. Pero, al mismo tiempo, se aprecian ciertas desventajas que, en mi opinión, deben ser atendidas en tanto influyen desfavorablemente en la eficacia del análisis. Una de ellas es que se dejaron de considerar los verbos reflexivos de forma como una sola palabra. Las formas que acompañan al verbo reflexivo no constituyen complementos, «su función es plenamente lexical y deben considerarse como parte de él» (De la Cueva Iglesias (II), 1989, p. 79), ej.: *Me voy para la escuela*. Otra de las desventajas es que comenzaron a reconocerse como nocionales los pronombres ordinales en función adjetiva. En los trabajos de riqueza revisados, todos los pronombres son excluidos del grupo de las palabras nocionales, lo cual es consecuente con el criterio de que «los pronombres se definen por no tener una significación concreta y fija, [...] y se refieren a los seres u objetos sin caracterizarlos» (De la Cueva Iglesias (I), 1989, p. 61).

Pero quizás la mayor desventaja de la segunda variante es el hecho de que se determine el índice de riqueza de los textos a partir de la proporción entre los vocablos nocionales

¹⁶ La segunda variante metodológica contabiliza los verbos reflexivos de forma como dos palabras, por tanto, el resultado del primer conteo de palabras aquí obtenido difiere del resultado del análisis anterior.

y el total de vocablos; esta operación define, esencialmente, la variante, de modo que pudiera cuestionársele.

El indicador de medición trabajado –porcentaje de vocablos de contenido nocional– es uno de los menos utilizados en las investigaciones de riqueza léxica. Hasta el momento solo lo incluyen en sus estudios Gloria E. Báez Pinal (1989) y Lidia Cuba Vega (2006). Ambas especialistas calculan el porcentaje de vocablos de contenido nocional al establecer la proporción entre vocablos nocionales y total de palabras de un texto; lo que coincide con la primera variante metodológica del Proyecto, por la cual tomo partido en esta investigación.

Esta variante resulta más acertada en tanto el procedimiento de análisis ofrece resultados acordes y consistentes con una de las principales reglas del indicador trabajado: «En este índice, el porcentaje obtenido es directamente proporcional al grado de riqueza léxica: a mayor porcentaje, mayor riqueza léxica» (Cuba Vega, 2006, p. 54). La segunda variante metodológica suele otorgar resultados que no se ajustan a esa regla. Dicho de otro modo: si la riqueza léxica aumenta a medida que lo hace el porcentaje de vocablos nocionales –relación directamente proporcional–, y este porcentaje será mayor mientras mayor sea el número de vocablos de contenido nocional, entonces la riqueza léxica ha de crecer si aumenta el número de vocablos nocionales.

Esto, efectivamente, se aprecia en los resultados del análisis de los textos, cuyos porcentajes de riqueza fueron calculados según la primera variante metodológica. Pero si se revisan los textos analizados de acuerdo con la segunda variante, se encuentran resultados incoherentes.

Lo anterior se explica mediante ejemplos: las redacciones 1108A y 1115A de la muestra –ambas escritas por estudiantes de primaria– comparten el mismo total de palabras (107), pero difieren en el número de vocablos nocionales (38 y 42, respectivamente). De acuerdo con la primera variante, el porcentaje de riqueza del primer texto (36.53 %) es menor que el índice del segundo (39.25 %), lo cual corrobora el hecho de que mientras mayor es el número de vocablos nocionales, mayor resulta la riqueza léxica.

La segunda variante metodológica ofrece resultados opuestos: el índice de riqueza del primer texto (62.5 %) es mayor que el porcentaje del segundo (58.90 %). Evidentemente, estos resultados no responden a la regla de la que se había hablado. Aquí el texto de más alto porcentaje no es el que posee mayor cantidad de vocablos nocionales y, por consiguiente, no representa el de mayor grado de riqueza. El aumento o la disminución del número de vocablos no nocionales determinan, en este análisis, el decrecimiento o crecimiento del porcentaje de riqueza léxica. El primer texto contiene

un menor número de vocablos no nocionales que el segundo (24 y 30, respectivamente); por ello su índice de riqueza es mayor.¹⁷ En este caso, la diferencia entre el número de vocablos nocionales no es la más influyente; más significativa es la diferencia entre vocablos no nocionales. Por tanto, esta última determina la desigualdad entre los índices porcentuales. Si el objetivo es obtener el porcentaje de vocablos de contenido nocional, el número de vocablos nocionales ha de determinar dicho porcentaje, no el número de vocablos no nocionales.

Queda demostrado que la primera variante metodológica resulta, sin dudas, la más acertada para la determinación de los índices de riqueza, lo cual constituye uno de los principales propósitos de esta investigación. Sin embargo, hay algunos presupuestos de dicha variante que no se comparten; por ello se harán tres propuestas que cancelan esos presupuestos y que se estiman que contribuyen a la eficacia del procedimiento de análisis, las cuales se presentan a continuación:

1. Reconocer todos los adverbios como palabras de contenido nocional.¹⁸
2. Reconocer todas las locuciones como una unidades de sentido equivalentes a la palabra.¹⁹
3. Descontar los sustantivos propios en la tercera operación del procedimiento – donde se delimitan las palabras nocionales–, y no en la segunda como lo ha establecido el proyecto.

Nuestro trabajo responde a las orientaciones de la directiva del proyecto, de modo que no se consideran estas propuestas en el análisis de los textos. No obstante, sería de gran utilidad tenerlas en cuenta en otros estudios sobre riqueza, cuyos resultados seguramente podrían aportar otros datos y, quizás, ser más precisos.

Exposición y análisis de los resultados

El análisis de los resultados de la muestra, examinada de acuerdo con la primera variante metodológica, revela que los estudiantes de la escuela primaria «buena» tienen mayores índices de riqueza léxica que los de la primaria «mala» –la media calculada de

¹⁷ Según la segunda variante metodológica, el porcentaje de vocablos nocionales se calcula estableciendo la proporción entre vocablos nocionales y total de vocablos (nocionales + no nocionales). El porcentaje de vocablos nocionales aumenta si crece el número de vocablos nocionales (relación directamente proporcional), pero también tiende a ser mayor en la medida en que disminuye el número de vocablos no nocionales (relación inversamente proporcional).

¹⁸ Recordemos que la primera variante metodológica no reconoce los adverbios relativos ni interrogativos como palabras nocionales.

¹⁹ Las perífrasis verbales con la estructura *querer + infinitivo*, *esperar + infinitivo* habían sido excluidas de este grupo.

la escuela «buena» es de 40.39 %, la de la «mala» es de 37.36 %–; lo mismo sucede con los escolares del preuniversitario «bueno» respecto de los alumnos del preuniversitario «malo» –la media del preuniversitario «bueno» es de 43.20 %; la de la escuela «mala» es de 41.57 %–. De tal manera, se corrobora la primera hipótesis de este trabajo: los estudiantes de las escuelas «buenas» tienen mayor riqueza léxica que los alumnos de las escuelas «malas». Se hace evidente entonces la pertinencia de la variable «tipo de escuela» que aplica en este estudio. La variable «edad» es una de las más influyentes. Los alumnos de 16 años de preuniversitario suelen tener mayores índices de riqueza que los de primaria de 11 años. Por tanto, se ratifica la segunda hipótesis de esta investigación: la riqueza léxica crece según aumenta la edad. Las diferencias porcentuales entre los estudiantes de uno y otro grupos etarios pudieran considerarse más o menos significativas en dependencia del tipo de escuela a la que pertenezcan.

El análisis de los resultados de la muestra, examinada de acuerdo con la segunda variante metodológica, indica que los estudiantes de la escuela primaria «buena» tienen mayores índices de riqueza léxica que los de la primaria «mala» –la media calculada de la escuela «buena» es de 63.12 %; la de la «mala» es de 62.58 %–. Los escolares del preuniversitario «malo» tienen mayores índices de riqueza que los alumnos del preuniversitario «bueno» –la media del preuniversitario «malo» es de 66.21 %; la de la escuela «buena» es de 65.63 %–. En el primer caso se confirma la primera hipótesis; en el segundo no se ratifica. Los escolares de 16 años de preuniversitario tienen mayores índices de riqueza léxica que los de primaria de 11 años de edad. Nuevamente se verifica la segunda hipótesis de esta investigación.

Es así que los resultados de la muestra analizada de acuerdo con la primera variante metodológica corroboran las dos primeras hipótesis presentadas en la introducción del presente trabajo. La variante que dejó establecida la directiva del proyecto –dígase segunda variante metodológica– ofrece resultados que confirman la segunda hipótesis, pero la primera no siempre se cumple.

Al evaluar los resultados generales se percibe que las mayores diferencias se hayan entre la primaria «mala» y el preuniversitario «bueno», y las menores entre la primaria «buena» y el preuniversitario «malo».

Según la primera variante metodológica, las diferencias porcentuales entre la primaria «mala» y el preuniversitario «bueno» son notables, en tanto en la escuela «buena» 18 porcentajes sobrepasan el 40 % y en «la mala» solo 5 índices superan este valor (figura 1).

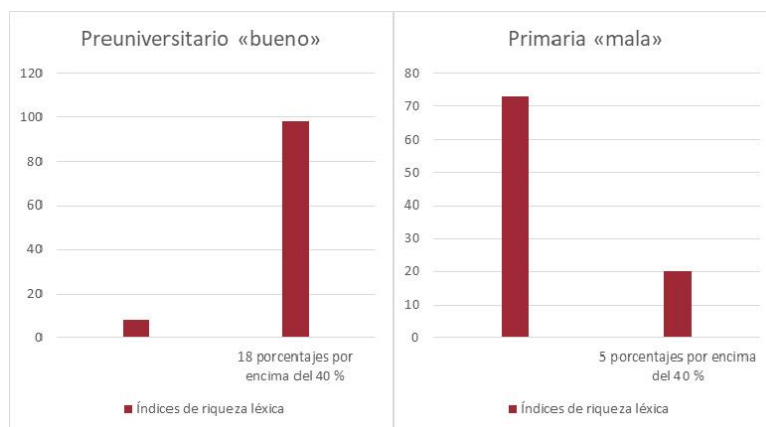


Figura 1. Diferencias porcentuales entre la primaria «mala» y el preuniversitario «bueno», según la primera variante metodológica

En el preuniversitario «bueno» hay 9 porcentajes por encima del 44 % y en la primaria «mala» solo un índice sobrepasa este porcentaje (figura 2).

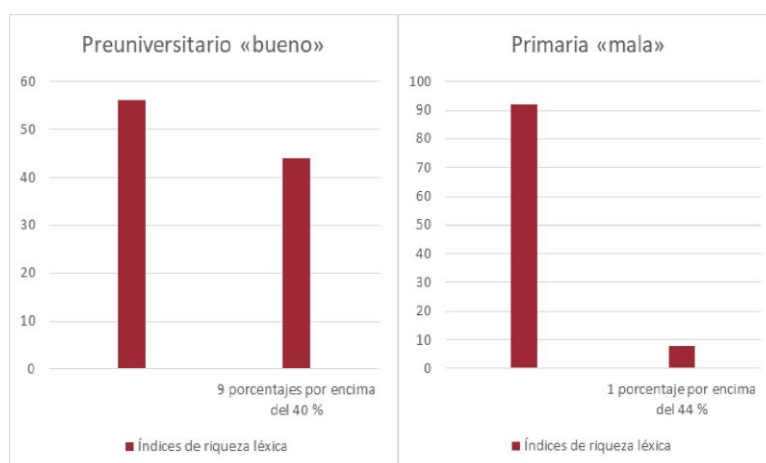


Figura 2. Diferencias porcentuales entre el preuniversitario «bueno» y la primaria «mala», según la primera variante metodológica

El análisis de acuerdo con la segunda metodología igualmente indica que las diferencias entre estas dos escuelas son significativas. Por encima del 64 % hay catorce índices en el preuniversitario «bueno», mientras que en la primaria «mala» hay solo siete porcentajes que superan ese valor. En el preuniversitario «bueno» ocho índices sobrepasan el 68 %, mientras que en la primaria «mala» dos valores porcentuales se encuentran por encima de ese porcentaje.

Los resultados estadísticos favorecen ampliamente a los alumnos del preuniversitario «bueno» por las dos razones básicas siguientes:

1. El hecho de que estos estudiantes tengan más edad que los alumnos de primaria, lo que supone un mayor grado de instrucción.

- 2- La condición de que pertenezcan a una escuela «buena», con gran prestigio social y educacional, a diferencia de la escuela «mala» que no suele tener este reconocimiento. Obviamente, en este caso las variables «edad» y «tipo de escuela» resultan muy influyentes.

Según la primera variante metodológica, las diferencias porcentuales entre la primaria «buena» y el preuniversitario «malo» son mínimas, en tanto en la escuela buena hay 13 porcentajes por encima del 38 % y 15 en la escuela mala (figura 3).

Figura 3. Diferencias porcentuales entre la primaria «buena» y el preuniversitario «malo», según la primera variante metodológica

En la primaria «buena» 9 índices sobrepasan el 40 %, mientras que en el preuniversitario «malo» 12 porcentajes superan ese valor (figura 4).

Figura 4. Diferencias porcentuales entre la primaria «buena» y el preuniversitario «malo», según la primera variante metodológica

El análisis según la segunda metodología igualmente indica que las diferencias entre estas dos escuelas son pequeñas: la escuela primaria «buena» contiene 13 porcentajes que sobrepasan el 62 %, mientras que en el preuniversitario hay 15 índices que superan dicho valor.

En este caso, las diferencias porcentuales entre las dos escuelas son mínimas a causa de la influencia de la variante «tipo de escuela». Los estudiantes de primaria de 11 años tienen un nivel escolar más bajo, pero pertenecen a una escuela «buena» de reconocido prestigio educacional. Los estudiantes de preuniversitario de 16 años pertenecen a un grado superior, pero son de una escuela «mala», lo que podría suponer ciertas deficiencias en el funcionamiento de la enseñanza de ese centro de estudio.

Las locuciones y su importancia para la determinación de los índices de riqueza léxica

Luego de haber analizado la muestra, pude darme cuenta de que los resultados de las pruebas de riqueza dependen, en cierta medida, del mayor o menor uso de locuciones en los textos. Frecuentemente, las composiciones con mayor número de locuciones tienen un índice de riqueza léxica menor que las que escasamente las contienen. Por tanto, se

pudiera deducir que el porcentaje de riqueza es inversamente proporcional al total de locuciones de los textos: mientras menor resulte el número de locuciones, mayor será la riqueza léxica. Ofrezco como ejemplos dos redacciones de estudiantes preuniversitarios de la escuela «buena» que difieren, significativamente, en cuanto al número de locuciones utilizadas.²⁰ La primera contiene 12 locuciones en total (2 adverbiales y 10 verbales, todas de contenido nocional), y la segunda posee solo 6 (3 verbales de contenido nocional, 2 conjuntivas y 1 preposicional no nocionales). Lo cual incide en el hecho de que la primera tenga una riqueza léxica menor que la segunda.

1620A

Francamente, mientras más trato de entender / la y predecir lo que viene, más me cuesta continuar sin tropezar. La vida es impredecible y quien trate de coger / le la seña terminará desilusionado. Varias veces me atreví a pensar que era fácil y posible; y ven, como lo cuento ¡volví a tropezar y volví a caer al suelo! Desconcertado *me + levantaba* y tomaba luego decisiones incorrectas, solo para demostrar / me que podía vencer; que podía adivinar lo que venía, o por lo menos imaginár/me/lo. Di vuelta y vuelta, y al fin: ¡lo descubrí! Lo importante no era adivinar nada sino lograr/lo todo; lo correcto no era *prepararse* mucho para lo que debía venir, sino *prepararse* un poquito en todo para que nada te sorprenda demasiado. La vida te sorprende /, te engaña, te miente, te hace reír, llorar, cantar; y por ultimo te hace amar, que es lo mejor que te puede hacer. [...]

- Total de locuciones: 12 (2 adverbiales y 10 verbales).

1601A

No podemos hablar de poemas, novelas, versos y cartas de gran significación si no mencionamos a nuestro José + Martí. Martí es uno de los lujos que la lengua española puede ofrecer a un público universal aunque apenas tuvo tiempo de ensayar/se a las letras. Fue cronista, ensayista y orador y con gran frecuencia se alcanza en sus versos altura poética. Con él concluye el esfuerzo romántico hacia una prosa estéticamente elaborada. Solía dar a esta forma de discurso, sermón, oración ya quea través de la reiteración de sustantivos y adjetivos desproporciona, recalca, amplifica y subordina excesivamente sus emociones e ideas con metáforas de gran sinceridad. Fue un hombre original ya que llegó a los más profundo de sí / con una sencillez inalcanzable, [...].

- Total de locuciones: 6 (3 verbales, 2 conjuntivas y 1 preposicional).

²⁰ Se analizará la riqueza léxica de los textos de acuerdo con cada una de las variantes metodológicas utilizadas por el proyecto para medirla.

La tabla 2 muestra el análisis de ambos textos a partir de las dos variantes metodológicas.

Tabla 2. Análisis de los textos 1620A y 1601A según las dos variantes metodológicas propuestas.

Textos	1620	1601
Análisis según la primera variante metodológica	Primer conteo de palabras: 111 Segundo conteo de palabras (sin siglas, cifras o sustantivos propios): 111 Vocablos de contenido nocional: 47 Índice de riqueza léxica: 42.34 %	Primer conteo de palabras: 113 Segundo conteo de palabras (sin siglas, cifras o sustantivos propios): 111 Vocablos de contenido nocional: 55 Índice de riqueza léxica: 49.54 %
Análisis de acuerdo con la segunda variante metodológica	Primer conteo de palabras: 114 Segundo conteo de palabras (sin siglas, cifras o sustantivos propios): 114 Total de vocablos: 70 Vocablos de contenido nocional: 48 Vocablos no nocionales: 22 Índice de riqueza léxica: 68.57 %	Primer conteo de palabras: 113 Segundo conteo de palabras (sin siglas, cifras o sustantivos propios): 111 Total de vocablos: 76 Vocablos de contenido nocional: 55 Vocablos no nocionales: 21 Índice de riqueza léxica: 72,36 %

De tal manera, se verifica que la frecuencia de aparición de estas construcciones incide en el crecimiento o la disminución de los índices de riqueza léxica, en dependencia de si aumentan o disminuyen sus ocurrencias. Por tanto, queda así demostrada la importancia del estudio y el reconocimiento de las locuciones para la aplicación de las pruebas de riqueza, lo cual constituye uno de los objetivos del presente trabajo.

La variable edad influye de manera determinante en este análisis: los escolares de la enseñanza primaria utilizaron 171 locuciones en total, mientras que los de preuniversitario emplearon 152. Estos resultados se corresponden con los del análisis de la muestra: los escolares de 11 años suelen tener menores índices de riqueza léxica que los de 16 años de edad.²¹ Se puede corroborar así otra de las hipótesis planteadas al inicio: la riqueza léxica será mayor mientras menor resulte el uso de locuciones.

Conclusiones

Los análisis realizados permiten arribar a las siguientes conclusiones:

- -La primera variante metodológica ofrece resultados que se corresponden con una de las principales reglas del indicador trabajado –porcentaje de vocablos de contenido nocional–: el porcentaje obtenido es directamente proporcional al grado de riqueza léxica: si aumenta el porcentaje, crece también la riqueza. La

²¹ Recordemos que la relación que existe entre el total de locuciones y el índice de riqueza de un texto es entendida aquí como una relación inversamente proporcional.

segunda variante metodológica suele otorgar resultados que contradicen esta regla. Por tanto, la variante más acertada es la primera, y por ello se ha tomado partido en esta investigación.

- -Los resultados de la muestra analizada de acuerdo con la primera variante metodológica corroboran las dos primeras hipótesis presentadas en la introducción del presente trabajo: los estudiantes de las escuelas «buenas» tienen mayor riqueza léxica que los alumnos de las escuelas «malas»; la riqueza léxica crece según aumenta la edad. La variante que dejó establecida la directiva del Proyecto –dígase segunda variante metodológica– ofrece resultados que confirman la segunda hipótesis, pero la primera no siempre se cumple.
- -Los estudiantes de primaria suelen utilizar un mayor número de locuciones que los de preuniversitario y estos últimos suelen tener mayores índices de riqueza que los primeros. Se corrobora, entonces, la última de las hipótesis planteadas al inicio del estudio: la riqueza léxica será mayor mientras menor resulte el uso de las locuciones.
 - El análisis desde el punto de vista cualitativo mostró que los factores principales que influyen en la disminución de los índices de riqueza léxica son la repetición de palabras y el uso indiscriminado de locuciones, lo cual está unido a otros aspectos determinantes como el nivel educacional y la influencia de factores sociales, económicos y culturales.

La planificación de textos y diccionarios escolares ha de realizarse sobre la base de un conocimiento científico que indique el grado de dominio léxico que poseen nuestros estudiantes. Conocer el vocabulario que estos utilizan también nos ayuda a descifrar las principales carencias de su léxico, las cuales pudieran subsanarse si se elaboran los textos escolares en función de las necesidades cognitivas de los educandos. En el ámbito de la enseñanza de la lengua, los estudios de riqueza léxica constituyen una valiosa herramienta para dar cumplimiento a esos fines. Como parte de este tipo de estudios, el presente trabajo se suma a los esfuerzos por perfeccionar el proceso de enseñanza-aprendizaje de la lengua materna y tiene el propósito de impulsar otras investigaciones encaminadas a resolver las principales problemáticas de dicho proceso.

REFERENCIAS

- ALVARADO ORTEGA, MARÍA BELÉN (2008): *Las fórmulas rutinarias en el español actual*, Universidad de Alicante. Tesis de doctorado.
- ÁVILA CELL, RAÚL (1988): «Índices de Riqueza Léxica», en «Lengua hablada y estrato social: un acercamiento léxico-estadístico», *NRFH* XXXVI, n.º 1, Costa Rica, pp. 131-148.
- ÁVILA CELL, RAÚL (1989): «Las palabras de los niños: tercer grado de primaria», *Estudios sobre Español de América y Lingüística Afroamericana*, Publicaciones del Instituto Caro y Cuervo LXXIII, Bogotá, pp. 19-39.
- BÁEZ PINAL, GLORIA E. (1989): *Índices de riqueza léxica en escolares de primaria de México D. F.* (Material impreso).
- BARKELIUS, JOSEFIN (2007): *Riqueza léxica y fluidez. Un estudio comparativo entre usuarios avanzados de castellano y hablantes nativos*, Universidad de Estocolmo. Tesis de investigación.
- CASTILLO CARBALLO, MARÍA AUXILIADORA (1997-1998): «El concepto de unidad fraseológica», *Revista de Lexicografía*, vol. IV, Universidad de Sevilla, pp. 67-79.
- CORPAS PASTOR, GLORIA (1996): *Manual de fraseología española*, Gredos, Madrid.
- CUBA VEGA, LIDIA (2006): «Índices de riqueza léxica en una muestra de escolares de primaria de Ciudad de La Habana», Universidad de La Habana. Tesis de doctorado.
- DE LA CUEVA IGLESIAS, OTILIA *et al.* (1989): *Manual de Gramática Española*, Editorial Pueblo y Educación, tomos I y II, La Habana.
- DOMÍNGUEZ HERNÁNDEZ, MARLEN A. (2011): Seminario «La enseñanza de la lengua materna», Conferencia «Riqueza Léxica» (transcrita por la autora de este trabajo), febrero, La Habana.
- EICHEVERRÍA, MAX SERGIO *et al.* (1992): «Evaluación de la riqueza léxica en estudiantes de último año de la Enseñanza Media», *Estudios Filológicos*, n.º 27, Universidad Austral de Chile, pp. 59-72.
- GONZÁLEZ MAFUD, ANA MARÍA (1999): «La lengua escrita de los escolares de primaria de Ciudad de la Habana: descripción y análisis de los indicadores principales de madurez sintáctica», Universidad de La Habana. Tesis de doctorado.
- HACHÉ DE YUNÉN, ANA MARGARITA (1991): «Aportes de las pruebas de riqueza léxica a la enseñanza de la lengua materna», *La enseñanza del español como lengua materna, Actas del II Seminario Internacional sobre Aportes de la lingüística a la enseñanza del español como lengua materna*, Editorial de la Universidad de Puerto Rico, pp. 49-60.

- HUHTINIEMI, SAIJA (2007): «El desarrollo de la morfología verbal en el aprendizaje “espontáneo” de E/LE», Universidad de Estocolmo. Tesis de investigación.
- ILINÁ, NATALIA (2000): «La fraseología española contemporánea: estado de la cuestión», Facultad de Lenguas Extranjeras adjunta a la Universidad Estatal Lomonosov de Moscú.
- LINARES TERRY, KELLY (2011): «Del léxico a la producción textual. Análisis lingüístico de una muestra de textos escritos por escolares de primaria de ciudad de La Habana», Universidad de La Habana. Tesis de grado.
- LÓPEZ CHÁVEZ, JUAN (1995): «Léxico y enseñanza de la lengua materna», *Actas del I Encuentro sobre Problemas de la Enseñanza del Español en México*, Zacatecas, mayo, pp. 93-105.
- LÓPEZ MORALES, HUMBERTO (1984): *Enseñanza de la lengua materna*, Puerto Rico, pp. 86-92.
- LÓPEZ MORALES, HUMBERTO (2008): *Documentos de trabajo del proyecto La enseñanza del español en el Mundo Hispánico*, s. e., La Habana.
- MARTÍNEZ MONTORO, JORGE (2002): «La fraseología en J. Casares», *Estudios de Lingüística*, Departamento de Filología Española, Lingüística General y Teoría de la Literatura, n.º 16, Universidad de Alicante, pp. 30-56.
- MORALES, AMPARO (1986): *Léxico básico del español de Puerto Rico*, Academia Puertorriqueña de la Lengua, San Juan.
- MÜLLER, CHARLES (1973): *Estadística lingüística*, Editorial Gredos, Madrid.
- MURILLO ROJAS, MARIELOS (2009): «Diversidad de vocabulario en los preescolares. Aportes para valorar su competencia léxica», *Filología y Lingüística*, vol. XXXV, n.º 1, pp. 123-138.
- NUEVA GRAMÁTICA DE LA LENGUA ESPAÑOLA (NGRLE): Edición Espasa Libros, volúmenes I y II, Madrid, 2009.
- OBEGI GALLARDO, NADIA (2005): «La Riqueza Léxica como parámetro de distinción en los textos del Inglés Antiguo», *Interlingüística*, vol. 16, n.º 2, Universidad de Málaga, pp. 837-846.
- PÉREZ MARQUÉS, CELIA MARÍA (2004): «Nuevo enfoque para un diagnóstico de desarrollo léxico en escolares de primaria», Universidad de Oriente. Tesis doctoral (inédito).
- PÉREZ MARQUÉS, CELIA MARÍA (2005): *Desarrollo de la competencia léxica en escolares de primaria de Guamá, Cuba*, Centro de Lingüística Aplicada, Santiago de Cuba, <<http://www.eumed.net>> [14/06/2011].

- PÉREZ MARQUÉS, CELIA MARÍA (2008): «Impacto de los programas de la revolución educacional en el desarrollo léxico de los escolares del municipio Guamá», <<http://www.anmal.uma.es/numero29/Guama.htm>> [14/06/2011].
- PEREZ MARQUES, CELIA MARIA *et al.* (2011): *Desarrollo léxico en escolares de primaria. Ejercicios para su perfeccionamiento*, Ediciones Centro de Lingüística Aplicada, Santiago de Cuba.
- RUIZ GURILLO, LEONOR (1997): «Aspectos de fraseología teórica española», *Revista Cuadernos de Filología*, n.º 24, Valencia.
- SINTRÓN, F. (1992): «Índices de riqueza léxica en escolares de Barranquitas», Universidad de Puerto Rico. Tesis de maestría.

Notas aclaratorias

1. Este Proyecto comenzó en 2007. Su dirección está representada por el investigador Humberto López Morales, natural de Puerto Rico.
2. A partir de aquí al referir el término «riqueza» se estará haciendo alusión a la «riqueza léxica».
3. Más adelante, en «Exposición y análisis de los resultados» (primera parte) aparecen referidas las características de la muestra.
4. Para conceptualizar la riqueza léxica hemos partido de la definición que propone la Dra. Lidia Cuba en su tesis de doctorado: «el grado de madurez y dominio léxico general –competencia léxica– de un hablante, medidos en la producción de sus textos, orales o escritos, por índices estadísticos como el porcentaje de vocablos, el intervalo de aparición de palabras de contenido nocional y el porcentaje de vocablos de contenido nocional» (Cuba Vega, 2006, p. 31).
5. La «actuación léxica» se define como la manifestación, en la producción de textos orales o escritos, del grado de madurez y dominio léxico general de un hablante. Constituye una manifestación de la competencia léxica de dicho hablante y puede ser medida, en su mayor parte, mediante procedimientos estadísticos (Pérez Marqués, 2011).
6. Raúl Ávila (1989) denomina a este indicador «frecuencia acumulada de tipos».
7. Esta variante es la empleada por Gloria E. Báez y Lidia Cuba para medir la riqueza léxica a través del indicador del porcentaje de vocablos de contenido nocional.
8. Se define «palabra» como la unidad lexical y semántica que puede ser simple o compleja en dependencia de su construcción gráfica. «Palabras de contenido nocional» (o nocionales) son aquellas que portan un contenido léxico: sustantivos, adjetivos, verbos, adverbios y las unidades léxicas complejas que tienen esas mismas funciones. (Adaptado de Cuba Vega, 2006, p. 48). «Los vocablos se definen como esas formas que se lematizan en un repertorio lexicográfico (*niña* no, sino *niño*; *comió* tampoco, sino *comer*)» (Domínguez Hernández, 2011).
9. La numeración de los textos responde a la siguiente ordenación: edad (11 años / 16 años), numeración del estudiante (01-20), tipo de escuela («buena»: A; «mala»: B).
10. La directiva del proyecto determinó que los títulos de los textos no se consideraran en los análisis.
11. El subrayado indica las locuciones que aparecen en el texto.
12. Recordemos que los elementos que integran las perífrasis verbales que tienen la estructura *querer + infinitivo* se contabilizan como palabras separadas de acuerdo con los presupuestos del proyecto.
13. Los sustantivos propios no se reconocen como palabras de contenido nocional, son descontados desde la segunda operación del procedimiento.
14. La segunda variante metodológica sí considera a los adverbios relativos como palabras nocionales, por ello en el segundo análisis reconocemos el adverbio temporal *cuando* como «una palabra más» de contenido nocional.
15. Aquí interrumpimos el conteo de palabras. Como se ha dicho, el corte cuantitativo no debe sobrepasar las primeras 115 palabras del texto.
16. La segunda variante metodológica contabiliza los verbos reflexivos de forma como dos palabras, por tanto, el resultado del primer conteo de palabras aquí obtenido difiere del resultado del análisis anterior.
17. Según la segunda variante metodológica, el porcentaje de vocablos nocionales se calcula estableciendo la proporción entre vocablos nocionales y total de vocablos (nocionales + no nocionales). El porcentaje de vocablos nocionales aumenta si crece el número de vocablos nocionales (relación directamente proporcional), pero también tiende a ser mayor en la medida en que disminuye el número de vocablos no nocionales (relación inversamente proporcional).

18. Recordemos que la primera variante metodológica no reconoce los adverbios relativos ni interrogativos como palabras nocionales.
19. Las perífrasis verbales con la estructura querer + infinitivo, esperar + infinitivo habían sido excluidas de este grupo.
20. Se analizará la riqueza léxica de los textos de acuerdo con cada una de las variantes metodológicas utilizadas por el proyecto para medirla.
21. Recordemos que la relación que existe entre el total de locuciones y el índice de riqueza de un texto es entendida aquí como una relación inversamente proporcional.