

# RIQUEZA LÉXICA DE TEXTOS REDACTADOS POR ALUMNOS DE BACHILLERATO DE LAS PALMAS DE GRAN CANARIA

## 1. INTRODUCCIÓN

El caudal léxico de cualquier individuo está constituido por todas las lexías que ha empleado y utiliza en sus expresiones orales y escritas, por todas las palabras de uso virtual (puesto que cualquier hablante está provisto de mecanismos lingüísticos que le permiten crear neologismos, creaciones momentáneas, a partir de modelos conocidos y memorizados) y, además, por todas las unidades léxicas que, como interlocutor, reconoce en la recepción de los mensajes en el curso de una conversación o lectura. Esto es, la competencia léxica individual es igual a la suma del vocabulario conocido más las lexías pasivas y las derivables. Si tenemos en cuenta la existencia, en la nomenclatura lingüística, de los conceptos de *vocabulario activo* y *vocabulario pasivo*, que aluden respectivamente a las voces que emplea un individuo como hablante y entiende como oyente (Menard 1983: 13), es necesario señalar que la labor del docente tiene por misión procurar enriquecer la competencia léxica de cualquier alumno y que el vocabulario pasivo de cada uno vaya convirtiéndose paulatinamente en activo, que “vaya pasando unidades de su léxico a su vocabulario”, como ha observado Lamíquiz (1974: 401). Pero para programar el vocabulario que se ha de enseñar a un grupo de estudiantes es necesario conocer de antemano de qué léxico dispone. Y para la comprobación y la medición tanto del léxico adquirido, como de la producción léxica de una muestra, se emplean distintas pruebas, descritas hace una veintena por López Morales (1984: 91). Las pruebas que miden la riqueza léxica de los textos de un corpus son de naturaleza cuantitativa y comparativa. Al respecto, Menard (1983:16) expresa que la *riqueza léxica* ha de entenderse como un lugar de comparación entre dos o varios textos en función de su extensión respectiva (número de palabras) y del

número de vocablos que figura en cada uno de ellos. Precisa que esta clase de descripciones comparativas requiere un análisis cuantitativo basado en modelos estadísticos y que los criterios de orden cualitativo son inoperantes para el fin que se pretende, aunque no sean prescindibles, que no lo son, porque ayudan a precisar el alcance y los límites del análisis cuantitativo. Por esto, afirma que la noción de riqueza léxica es de naturaleza cuantitativa, puesto que mide el vocabulario activo de un corpus textual determinado.

Los indicadores de riqueza léxica que vienen teniéndose en cuenta en esta modalidad descriptiva son:

1. Una de las pruebas permite la obtención del porcentaje de vocablos, esto es, la proporción de palabras diferentes del total de palabras de una composición. Para hallar este índice de riqueza léxica se cuenta el número total de palabras del discurso y el número de vocablos (palabras diferentes) con el fin de reconocer la proporción entre ambas cifras.
2. Otra, pretende establecer el intervalo de aparición de las palabras nocionales: N/PN. Este índice considera la cantidad total de palabras del escrito y la divide por el conjunto formado por nombres, adjetivos, verbos y adverbios, para obtener el intervalo de aparición en el texto de estas palabras. Se dirá que este índice aumenta conforme disminuya el intervalo de aparición de palabras nocionales; por el contrario, cuanto más dilatado sea el intervalo, menor será la riqueza léxica de la muestra en cuestión. Así, un intervalo de voces de contenido léxico de 2,1 detecta menor riqueza léxica que si se consigue un intervalo entre léxias de 1,2.
3. El índice de hápax detecta el empleo de palabras que sólo se utilizan una vez en el texto analizado. La variedad de hápax se obtiene con la aplicación de la fórmula  $V/V_1$ , donde  $V$  es el número total de vocablos y  $V_1$  la suma de las palabras de una ocurrencia. Para determinarlo se divide la cifra total de vocablos por aquellos que sólo se emplean una vez en un texto determinado, de modo que cuanto más pequeño sea el índice, más riqueza léxica caracteriza al texto que se confronta con otro. Conforme el índice va aumentando, la variedad léxica va disminuyendo; así, una medida de 1,34 revela mayor cantidad de unidades de una sola aparición en un texto que otra de 1,37.

Los procedimientos reseñados calculan respectivamente la producción de vocablos, de palabras nocionales y de hápax de un conjunto de textos con el mismo número de palabras. Según estos cálculos, los textos que contienen el mayor porcentaje de vocablos y el menor intervalo de palabras nocionales y de

hápax son los calificados como los más cualificados. Por eso, este tipo de evaluación permite conocer el dominio del léxico activo de un sujeto o de un grupo. Y, por esta razón, su empleo es recomendable en el ámbito de la enseñanza, ya que un profesor, después de tener constancia del vocabulario que un conjunto de alumnos maneja sobre un tema específico, puede programar de una manera más realista la enseñanza del léxico desconocido (Reyes Díaz 2007).

## 2. ESTUDIOS HISPÁNICOS DE RIQUEZA LÉXICA

A partir de la publicación del libro *Enseñanza de la lengua materna* (López Morales 1984), donde su autor expone las pruebas de riqueza léxica que sirven para la medición de la producción léxica de textos confeccionados por escolares, fueron surgiendo en el mundo hispánico trabajos destinados a la descripción del vocabulario activo de muestras específicas de estudiantes. Estos análisis en cierta manera han estado vinculados entre sí, pues cada nueva investigación se diseña teniendo en cuenta las conclusiones de los experimentos precedentes. Desde Hispanoamérica han ido llegando a esta orilla del Atlántico con generosidad los resultados de las investigaciones de Haché de Yunén, Echeverría, Portela y Cintrón Serrano. A ellos se han sumado las aportaciones de otros estudiosos españoles, como los de Torres González y Andrés Pérez.

A la pregunta ¿qué finalidades persigue este tipo de análisis cuantitativo de producción léxica? responde Haché de Yunén (1991) con un artículo publicado hace varios lustros. La autora presenta una enumeración abierta de las posibles utilidades de los estudios de riqueza léxica:

1. Las pruebas son de utilidad para diagnosticar qué vocabulario conocen los alumnos, de manera que se pueda actuar en consecuencia.
2. El vocabulario activado por los niños en las pruebas indicaría qué voces habrían de incluirse en los cuentos de los más pequeños para facilitarles la comprensión del contenido (que tal vez podría bloquearse por el sobreesfuerzo que exigiría el empleo de un lenguaje plagado de voces desconocidas) y el gusto por la lectura.
3. Las listas de los lexemas activados podrían contrastarse con el vocabulario de los textos escolares y de las lecturas infantiles con el fin de comprobar si existe adecuación.
4. Además, aportarían información para la elaboración de diccionarios infantiles adecuados al entorno lingüístico y cultural circundante.
5. Asimismo, constituye un dispositivo recomendable para la actividad docente a la hora de establecer una valoración certera del desarrollo del

vocabulario activo de cada uno de los alumnos y de la totalidad de estos en un grupo determinado.

Haché de Yunén tomó como muestra informativa a escolares dominicanos de sexto y octavo grado de la escuela primaria, y les aplicó las variables sexo y nivel sociocultural. La investigadora considera la hipótesis de que existe covariación entre las variables sociales y los índices de riqueza léxica que estudia: porcentaje de vocablos e intervalo de aparición de las palabras de contenido nocional. El objetivo fundamental que persigue el diseño de la investigación consiste en la puesta en práctica de las pruebas presentadas por López Morales (1984) para evaluar la riqueza léxica de estudiantes. El corpus del plan de trabajo experimental quedó constituido por 80 narraciones sobre un mismo tema, que consistió en contar por escrito “Un paseo” hipotético realizado con un grupo de amigos a la playa o a cualquier otro lugar. La decisión de que todos escribieran sobre un tema determinado, utilizando el modo narrativo, se hizo para que el contraste de todas las composiciones fuera más fiable, dado que el modo discursivo repercute en la riqueza léxica. El análisis considera la palabra como la unidad gráfica, pero separando los elementos aglutinados del tipo de “al, del, dámelo” como procedimiento previo a los cálculos léxicos. Los resultados de la investigación dejaron la siguiente información: los niños alcanzan índices de riqueza léxica superiores a los alcanzados por sus compañeras en las redacciones; el nivel sociocultural de los alumnos de la muestra no influye en el caudal de vocablos que emplean en las composiciones, aunque sí parecen repercutir en el intervalo; por el contrario, sí repercute la variable edad, que favorece a los mayores. La observación de contrastes de textos sin límites dejó ver que el índice que mide el porcentaje de vocablos en relación con el total de palabras era ventajoso para los textos más exigüos en palabras y, por el contrario, era desfavorable para los textos más largos. Este resultado llevó a la investigadora a ratificar análisis anteriores (Ham 1979: 79, Ávila 1986: 512) que mostraban que la proporción de vocablos decrece con el alargamiento del texto. Por eso propone que en investigaciones futuras se parta de textos con un mismo tamaño, a fin de facilitar el cotejo de distintas investigaciones.

Un año después se publica otro trabajo sobre el mismo tema en Chile firmado por Echeverría *et al.* (1992), quienes seleccionan una muestra de 308 estudiantes chilenos de último curso de enseñanza media. En las operaciones de recuento, este grupo decide no identificar el concepto de palabra con la unidad gráfica, como se había hecho en trabajos anteriores, sino con la unidad semántica. Las variables analizadas son sexo, nivel sociocultural, tipo de escuela (pública y privada), procedencia geográfica de los sujetos de la muestra (zonas norte, sur y metropolitana) y tipo discursivo (narrativo, descriptivo, expositivo y argumentativo). El recuento da a conocer que la riqueza léxica de un sujeto no la determina la variable

sexo, aunque sí parece tener cierta relevancia la variable nivel sociocultural, pues los textos de los individuos que proceden del nivel sociocultural alto presentan una cantidad superior de vocablos y palabras léxicas que los de los niveles medio y bajo. Los datos del escrutinio también indican que el tipo de escuela incide en la riqueza léxica de los sujetos, como apuntan los índices de riqueza léxica superiores alcanzados en porcentajes de vocablos y palabras nocionales por los niños de la enseñanza privada. Asimismo, muestran la existencia de relación asociativa entre la variable zona geográfica y las variables lingüísticas, pues los resultados de las mediciones favorecen a los estudiantes de la zona metropolitana. Para el modo discursivo, las conclusiones que se desprenden son que el discurso narrativo propicia la mayor afluencia de vocablos y el discurso argumentativo, la menor.

Por las mismas fechas, en la República Dominicana, Portela (1992) presenta los resultados de otro experimento léxico, para cuyo diseño tuvo en cuenta los trabajos hispanoamericanos que sobre el tema habían visto la luz en ese momento. En la elaboración de su tesis decide trabajar con dos versiones y con los tipos de medición léxica propuestos por López Morales (1984) y Ávila (1986). En la versión A, al realizar el recuento del vocabulario, delimita la palabra como unidad gráfica; en la versión B, en cambio, sigue el criterio que establece que una palabra es una unidad significativa aunque sus elementos componentes, dos o más, no se hayan fusionado gráficamente. Portela trabajó con la producción de dos textos narrativos distintos y con temas determinados, uno sobre “Un paseo” para la preprueba y otro sobre “El argumento de una novela” para la postprueba, por parte de una población de 84 individuos dominicanos de nuevo ingreso de la Pontificia Universidad Católica Madre y Maestra. Este grupo quedó estratificado a partir de las variables sexo, nivel socioeconómico, procedencia geográfica (urbana o rural) de los universitarios, tipo de centro, estrato sociocultural (nivel alto y nivel bajo) y momento de la prueba (preprueba y postprueba). Las conclusiones que se desprenden del análisis de los datos muestran que la variable sexo no está asociada a los índices de riqueza léxica; los resultados de las pruebas se inclinan a favor del nivel alto frente al estrato bajo y, de igual modo, la mayor riqueza léxica se concentra en los textos de los estudiantes de las zonas urbanas, de la enseñanza privada y, en términos generales, de los alumnos del nivel 101, que repiten. Además, no se percibe un crecimiento de los índices de riqueza léxica en la postprueba con respecto a la preprueba.

En 1992, también Cintrón Serrano da a conocer de modo oficial los resultados de una tesis similar realizada en Puerto Rico con una muestra integrada por 120 individuos de cuatro grupos de 30 escolares cada uno: tercero, sexto, noveno y duodécimo. Como en la investigación del círculo chileno, para la medición de la riqueza léxica la autora delimita la palabra, no a partir del aspecto gráfico, sino de la unidad semántica. Por otro lado, le pide al alumnado evaluado que escriba sobre el tema “Mis vacaciones” y que use como expresión discursiva el modo narrativo. Para

averiguar la riqueza léxica de los estudiantes emplea las dos medidas propuestas por López Morales (1984). Se plantea como objetivo principal averiguar la corroboración de la existencia de relación asociativa entre índices de riqueza léxica y grado escolar. Examina, además, el comportamiento de las variables sexo y nivel sociocultural con los índices de riqueza léxica. Al finalizar todos los cómputos, descubre que las variables sociales están correlacionadas con las variables lingüísticas, pues los indicadores de riqueza léxica favorecen a las mujeres, a los niveles socioculturales superiores y a los grados escolares más avanzados. En cuanto al empleo de las categorías, comprueba que el orden no es sorprendente: sustantivo, verbo, adjetivo y adverbio. Por otra parte, observa el aumento de vocabulario de la categoría verbal en los dos cursos superiores y el ascenso progresivo del empleo de términos sinónimos con el avance de grado escolar de los informantes.

Otra investigación hispánica, citada por Benítez Pérez (2003), es la de Andrés Pérez (1997), que se realiza en la Universidad de Alcalá de Henares. El autor se propuso la descripción del grado de riqueza léxica de una muestra de 60 alumnos de cuarto, sexto y octavo de EGB. Además del grado escolar (edad), consideró las variables sexo y nivel socioeconómico. El tema de las narraciones versó sobre “Un fin de semana”. Tras la observación de los datos suministrados por el análisis, concluye que los textos elaborados por las chicas contienen un mayor grado de riqueza léxica que los textos compuestos por los chicos, que la edad o nivel escolar es un factor que determina la riqueza léxica de los alumnos; y que el factor social influye, sobre todo, en las primeras etapas en la producción de vocablos.

También en la Universidad de Alcalá de Henares, en la década de los noventa, se llevó a cabo un estudio, “Riqueza léxica en textos de estudiantes de español como lengua extranjera”, firmado por García Rosas (1996). La muestra incluye a 41 estudiantes, que confeccionan textos en tres modos discursivos: descriptivo, narrativo y argumentativo. A esta variable se añade el nivel de enseñanza (medio, avanzado, superior), el sexo y la lengua materna de los estudiantes. Los resultados de la investigación manifiestan que el tipo de discurso que ofrece mayor riqueza léxica es el narrativo y el que menos, el descriptivo; que la edad y la lengua materna del individuo no se correlacionan con los índices de riqueza léxica; y que el léxico de los informantes crece paralelamente con el nivel de enseñanza.

Torres González (1999, 2003a, 2003b) analiza la producción léxica de alumnos tinerfeños de cuarto de Primaria, segundo de Enseñanza Secundaria Obligatoria y segundo de Bachillerato. Junto a la variable curso considera las variables tipo de centro, sexo y nivel sociocultural. Las variables lingüísticas que considera son las pruebas que investigan la proporción de vocablos, el intervalo de las palabras nocionales, el índice de hápax y los promedios de las categorías nocionales. El material de estudio comprende 140 narraciones de tema libre. Todas las composiciones escritas tienen la misma longitud: cien palabras. Finalizada la descripción

cuantitativa, la investigadora concluye que, por lo general, los mayores valores de riqueza léxica corresponden a los centros privados y a la variable femenina, asimismo, comprueba que el ascenso en el nivel sociocultural de los estudiantes va parejo con el aumento de los indicadores de riqueza léxica y que la progresión escolar conlleva el incremento de lexías adjetivas.

La lectura de los resultados obtenidos por las investigaciones hispánicas citadas hace pensar que los índices de riqueza léxica están asociados a la variable enseñanza privada, al factor zona urbana, a los niveles socioculturales superiores, a las mujeres, a los cursos superiores y a los textos narrativos.

### **3. DESCRIPCIÓN DE LA RIQUEZA LÉXICA DE UNA MUESTRA DE ALUMNOS GRANCANARIOS**

El trabajo que aquí exponemos describe los resultados de la aplicación de las pruebas de riqueza léxica a dos grados educativos del antiguo Bachillerato: segundo de BUP y COU, cuyos alumnos tenían una edad media de 16 y 18 años respectivamente. Los grados equivalentes en el actual Plan de Estudio son cuarto de la ESO y segundo de Bachillerato. En cada curso escolar analizamos la variable tipo de centro (privado y público), la variable sexo y nivel sociocultural de los informantes, con el fin de averiguar si los resultados de las tres pruebas de riqueza léxica de los materiales de estas variables confirman las conclusiones que derivan de la lectura de los hallazgos de los estudios reseñados en el apartado anterior.

Los procedimientos encaminados a la investigación de la riqueza léxica de las narraciones escolares son los que ya hemos expuesto en la introducción.

#### **3.1. Instrumento de evaluación**

El presente análisis lo hemos realizado sobre un corpus de redacciones tomado en 1997 de todos los individuos de los cursos encuestados y elaborado en el aula en presencia del profesor del centro respectivo y del responsable de este estudio. Pedimos a los alumnos que escribieran un texto en el que contaran qué hacían en su “Tiempo libre”. Dado que el modo discursivo que habíamos elegido para la recogida de los datos era el modo narrativo, insistimos en que narraran lo que suelen hacer habitualmente en su tiempo de ocio. Así quedó fijado el tema y el modo discursivo. De todo el material originado por esta tarea, desechamos aquellas narraciones que no cumplían con los requisitos preestablecidos. Por ejemplo, eliminamos las redacciones que no alcanzaron el número total de 100 palabras, ya que previamente habíamos resuelto que todos los textos debían tener esa longitud. Además, suprimimos los textos de los informantes que no aportaban todos los datos personales que les pedíamos en los cuestionarios para la elaboración del perfil

sociocultural de la muestra. Una vez concluida la fase de revisión, seleccionamos de manera aleatoria las 32 narraciones que debían constituir el material básico del estudio de cada curso escolar. Luego unificamos el tamaño de todos los textos del corpus. Esta labor supuso la elección previa de ciertos criterios de lematización para la delimitación de las palabras, siguiendo las pautas marcadas por los trabajos de Portela (1992), Cintrón Serrano (1992) y Torres González (1999), quienes segmentan como una sola unidad léxica las construcciones perifrásticas (*ir+infinitivo*, *tener+que+infinitivo*, *haber+de+infinitivo*), los vocablos compuestos Sontagmáticos (*silla de ruedas*, *pez espada*, *un pobre diablo*) y los verbos pronominales (*olvidarse*, *alegrarse*, *marcharse*). Como recomendó Muller (1973: 253), contabilizamos como voces diferentes los casos de polisemia y homonimia homógrafa. Asimismo, agrupamos en un mismo lema la variación de género y de número tanto de adjetivos como de sustantivos, los elementos derivados apreciativos (expresivos) y la flexión verbal.

La fijación del tema y del modo discursivo, así como la delimitación de la unidad léxica y del texto facilitan la tarea descriptiva y el contraste de datos.

### 3.2. Información sobre la muestra

Antes de elegir por sorteo los centros de enseñanza que debían suministraros la muestra de segundo de Bachillerato y COU, tomamos la decisión de que uno de cada tipo debía estar ubicado en zonas distintas de Las Palmas de Gran Canaria, con el fin de asegurar la representatividad de la muestra. Los centros que participaron en la investigación fueron los colegios “San Ignacio de Loyola” y “San Antonio María Claret” (centros de enseñanza privada), y los institutos “Alonso Quesada” y “Tomás Morales” (centros de enseñanza pública). Cada institución se encargó de la asignación de los grupos, que en todos los casos sobrepasaban el número de alumnos necesarios, lo cual nos concedió cierta libertad para la elección aleatoria del número total de informantes (32) que se requería según la estratificación que se había establecido de antemano. Los sujetos de la muestra se distribuyeron equitativamente en las submuestras de las variables tipo de centro de enseñanza y sexo, como habíamos determinado.

Después de recoger las redacciones de los estudiantes, les pasamos un cuestionario con varios ítems para que los cumplimentaran con la información acerca de la ocupación profesional, la instrucción y el salario mensual de los padres a fin de establecer su nivel sociocultural. La información suministrada por cada informante con respecto a la ocupación laboral de sus padres nos permitió su inclusión en una de las seis categorías profesionales que consideramos siguiendo el modelo utilizado por Samper Padilla (1940: 40),



pero suprimimos la categoría de altos directivos y grandes empresarios, porque su número es insignificante en la población de Las Palmas de Gran Canaria:

Obreros sin cualificar.	1
Obreros cualificados.	2
Empleados medios.	3
Pequeños empresarios y autónomos.	4
Medianos empresarios.	5

En el parámetro referente al grado de instrucción de los padres distinguimos seis categorías. A cada una le otorgamos el valor que señalamos a continuación:

Sin estudios	1
Estudios primarios	2
Bachillerato elemental o asimilado	3
Bachillerato Superior o asimilado	4
Estudios universitarios medios	5
Estudios universitarios superiores	6

El parámetro de ingresos que establecimos para la recogida de la información sobre el salario en pesetas que necesitábamos para la estratificación social de la muestra lo estructuramos y cuantificamos de esta manera:

Hasta 65.000	1
De 65.000 a 85.000	2
De 85.000 a 110.000	3
De 110.000 a 150.000	4
De 150.000 a 200.000	5
De 200.000 a 275.000	6
De 275.000 a 350.0000	7
Más de 350.000	8

Después de efectuar la suma de los valores conferidos a las categorías de los tres parámetros, calculamos la media de los datos relativos a los cónyuges. Para ponderar los parámetros asignamos un valor cuantitativo a cada uno: 5 a la ocupación profesional, 3 al grado de instrucción, y 6 a los ingresos. Con la información suministrada por este baremo delineamos el nivel sociocultural a que pertenecen los estudiantes. Observamos que los informantes de segundo se distribuyeron en tres niveles. De los 32 alumnos, 8 conforman el nivel medio y

los 24 restantes se reparten a partes iguales en el nivel medio-bajo y bajo. En COU, la variable nivel sociocultural adopta la siguiente estructura: 3 estudiantes conforman el estrato sociocultural medio-alto; 5, el medio; 12, el medio-bajo; y 12, el nivel bajo de la escala.

NSC	Segundo	COU
Medio-alto		3
Medio	8	5
Medio-bajo	12	12
Bajo	12	12
Total	32	32

Cuadro 1: Distribución sociocultural

Los niveles socioculturales inferiores de esta muestra están más densamente poblados que el resto, lo que pone de relieve con total nitidez la extracción social de los sujetos.

### 3. 3. Presentación de los resultados del estudio de las variables de segundo curso de BUP

A continuación exponemos los resultados obtenidos de la aplicación de las pruebas de riqueza léxica a las variables tipo de centro, sexo y nivel sociocultural de segundo de Bachillerato.

Curso	60,65%	
Tipo de centro	Privado	61,25%
	Público	60,06%
Sexo	Masculino	61,37%
	Femenino	59,93%
Nivel sociocultural	Medio	63,25%
	Medio-bajo	60,16%
	Bajo	59,41%

Cuadro 2: Proporción de los vocablos de segundo de BUP

### a) Proporción de vocablos

El análisis de vocablos de las composiciones de los alumnos de segundo curso de Bachillerato descubre que la proporción de estas unidades es de 60,65%. El vocabulario de las narraciones correspondientes a los centros privados (61,25%) es ligeramente más extenso que el de los centros públicos (60,06%). Los estudiantes, que alcanzan un 61,37%, superan en esta prueba a sus colegas femeninas, que logran un porcentaje de 59,93%. Asimismo la descripción cuantitativa revela la existencia de correlación entre la variable lingüística y el factor sociocultural: al nivel medio le corresponde un porcentaje de 63,25%, al medio-bajo, 60,16%; y al bajo, 59,41%.

### b) Intervalo de las palabras nocionales

La operación que indaga el intervalo de uso de las palabras de contenido léxico muestra que 2,05 es la medida del intervalo medio de segundo en su conjunto. También determina que la cantidad de expresiones léxicas de los centros públicos (2,03) sobrepasan a los privados (2,07), como atestiguan los intervalos obtenidos por ambas variables. Además descubre que el intervalo de la variable masculina (1,98) es menor que el de la femenina (2,13). En el factor sociocultural se observa que los intervalos más cortos se asientan en los niveles extremos de la escala sociocultural (2,03 en el medio y 2,04 en el bajo) y el más amplio se ubica en el estrato medio-bajo, con una medida de 2,08.

Curso	2,05	
Tipo de centro	Privado	2,07
	Público	2,03
Sexo	Masculino	1,98
	Femenino	2,13
Nivel sociocultural	Medio	2,03
	Medio-bajo	2,08
	Bajo	2,04

Cuadro 3: Intervalo de las palabras nocionales de segundo de BUP

### c) Índice de hápax

Los resultados de la operación que identifica y cuantifica las palabras de una sola ocurrencia en las narraciones indican que su valor es de 1,40 en el conjunto examinado; de 1,42 en los informantes de los centros privados y de 1,37 en los centros públicos, lo que constata la superioridad léxica del material

de estos últimos. Asimismo, es superior el número de hápax en la variable masculina (1,36) que en la femenina (1,43). Por otro lado, la investigación prueba que las composiciones más copiosas en hápax son los escritos por los sujetos del nivel sociocultural bajo.

Curso	1,40	
Tipo de centro	Privado	1,42
	Público	1,37
Sexo	Masculino	1,36
	Femenino	1,43
Nivel sociocultural	Medio	1,38
	Medio-bajo	1,43
	Bajo	1,37

Cuadro 4: Resultados de la descripción de hápax de segundo de BUP

#### d) Recapitulación

La investigación arroja la siguiente información: el caudal de vocablos de los centros privados supera a la de los centros públicos. Las tres pruebas avalan la superioridad léxica de la variable masculina. Los valores de riqueza léxica no están claramente correlacionados con la posición social del informante, aunque el estrato social medio sobresale sobre el resto en la producción de vocablos y en la actualización de palabras de contenido léxico.

Curso	Segundo	Vocablos	Palabras nacionales	Hápax
Tipo de centro	Privado	61,25%	2,07	1,42
	Público	60,06%	2,03	1,37
Sexo	Masculino	61,37%	1,98	1,36
	Femenino	59,93%	2,13	1,43
Nivel sociocultural	Medio	63,25%	2,03	1,38
	Medio-bajo	60,16%	2,08	1,43
	Bajo	59,41%	2,04	1,37

Cuadro 5: Cotejo de los resultados de las tres pruebas de riqueza léxica de segundo de BUP

Los resultados de la variable ‘sexo’ corroboran los hallazgos de la investigación de Haché de Yunén (1991), pues en ambos escrutinios los índices favorecen a la variable masculina.

### 3. 4. Presentación de los resultados del estudio de las variables de COU

Como hicimos en el apartado dedicado a segundo de Bachillerato, expondremos a continuación los resultados de las tres pruebas de riqueza léxica aplicadas a las variables ‘tipo de centro’, ‘sexo’ y ‘nivel sociocultural’ de COU.

#### a) Proporción de vocablos

El porcentaje de los vocablos reunidos de todas las narraciones de la muestra seleccionada de COU asciende al 63,96% de las 3200 palabras que constituyen el corpus de este curso. Por otro lado, el análisis de vocablos determina que los materiales de los centros privados (65%) son más ricos que los de los centros públicos (62,93%). Por el contrario, no hay distinción notoria en el caudal que actualizan la variable masculina (64%) y la femenina (63,93%). En cuanto al factor sociocultural, las operaciones indican que las redacciones elaboradas por los informantes de los niveles sociales intermedios contienen más vocabulario y, por consiguiente, son más exiguos los extremos.

Curso	63,96%	
Tipo de centro	Privado	65%
	Público	62,93%
Sexo	Masculino	64%
	Femenino	63,93%
Nivel sociocultural	Medio-alto	63,33%
	Medio	64,6%
	Medio-bajo	65,16%
	Bajo	62,66%

Cuadro 6: Proporción de vocablos de COU

#### b) Intervalo de aparición de las palabras nocionales

Los cálculos evidencian que el intervalo de aparición de palabras nocionales en COU es igual a 2,08; que el corpus de los alumnos que estudian en los centros privados posee un intervalo de palabras nocionales (2,05) menor que el de los centros públicos (2,12); que los resultados de esta prueba obtenidos por los chicos y las chicas son exactamente iguales (2,08); y que el factor social no incide en el uso de este tipo de unidades lingüísticas, pues el intervalo más estrecho corresponde al

estrato medio-bajo (2,04) y el más dilatado, al medio (2,18). La diferencia de amplitud del intervalo en el estrato alto y en el bajo es la meno diferenciada.

Curso	2,08	
Tipo de centro	Privado	2,05
	Público	2,12
Sexo	Masculino	2,08
	Femenino	2,08
Nivel sociocultural	Medio-alto	2,08
	Medio	2,18
	Medio-bajo	2,04
	Bajo	2,09

Cuadro 7: Intervalo de las palabras nocionales de COU

### c) Índice de hápax

La operación que analiza los hápax del corpus de esta muestra detecta que su intervalo en COU disminuye (1,35) con respecto a segundo de Bachillerato (1,40), lo que indica que sus informantes emplean con más frecuencia en los textos producidos voces de una sola ocurrencia. Esto supone que la edad y la formación repercuten en la producción de hápax. Además, los resultados ponen de relieve que los centros privados (1,34) aventajan a los centros públicos (1,36), y los hombres (1,33), a las mujeres (1,37). Asimismo, manifiestan que los datos de los niveles extremos del perfil sociocultural, que coinciden (1,36), revelan que el corpus de estos niveles es más exiguo en voces de un solo uso que el resto.

Curso	1,35	
Tipo de centro	Privado	1,34
	Público	1,36
Sexo	Masculino	1,33
	Femenino	1,37
Nivel sociocultural	Medio-alto	1,36
	Medio	1,32
	Medio-bajo	1,34
	Bajo	1,36

Cuadro 8: Resultados de la prueba de hápax de COU

#### d) Recapitulación

La comparación de los hallazgos derivados de la aplicación de los tres indicadores de riqueza léxica marca la superioridad del dominio del vocabulario de los alumnos de los centros de enseñanza privada frente al de la muestra procedente de los centros públicos. Asimismo, permite apreciar que la diferencia entre variable masculina y femenina en COU se especifica especialmente en el examen de hápax, que pone de manifiesto que los varones aventajan a las mujeres. Y, además, que las redacciones de los sujetos pertenecientes al nivel medio-bajo son las más copiosas en vocabulario y en palabras de contenido léxico.

Curso	COU	Vocablos	Palabras nocionales	Hápax
Tipo de centro	Privado	65%	2,05	1,34
	Público	62,93%	2,12	1,36
Sexo	Masculino	64%	2,08	1,33
	Femenino	63,93%	2,08	1,37
Nivel sociocultural	Medio-alto	63,33%	2,08	1,36
	Medio	64,6%	2,18	1,32
	Medio-bajo	65,16%	2,04	1,34
	Bajo	62,66%	2,09	1,36

Cuadro 9: Cotejo de los resultados de las tres pruebas de riqueza léxica de COU

Los resultados de la variable ‘tipo de centro’ de COU corroboran los hallazgos de las investigaciones de Portela y de Torres González.

### 3. 5. Análisis comparativo de los resultados de segundo y COU

Se nota un crecimiento de vocablos y de hápax de la muestra de COU en contraste con la de segundo de Bachillerato. Sin embargo, en los resultados de las pruebas que miden el intervalo de palabras nocionales no se advierte progreso.

Curso	Vocablos	Palabras nocionales	Hápax
Segundo de BUP	60,65%	2,05	1,40
COU	63,96%	2,8	1,35

Cuadro 10: Comparación de los resultados globales de segundo de BUP y COU

Los datos expuestos en los cuadros 5 y 9 denotan que el comportamiento de la variable tipo de centro es diferente en los dos grados educativos: en segundo curso sólo en el volumen de vocablos supera la muestra de los centros privados a los públicos, mientras que en COU lo hace en todas las pruebas de riqueza léxica. También es diferente la forma en que procede la variable sexo, ya que en segundo de Bachillerato la masculina supera a la femenina en todos los índices de riqueza léxica, sin embargo en COU se aprecia que los jóvenes aventajan con nitidez a las mujeres en el índice de hápax. Asimismo, en lo que a la variable nivel sociocultural se refiere, es dispar la distribución de los mayores valores en uno y otro curso de Bachillerato, pues en segundo, los resultados de las pruebas de proporción de vocablos y de establecimiento del intervalo de palabras nocionales favorecen al estrato medio; en COU, al inmediato inferior; esto es: al nivel medio-bajo. En el estudio de hápax, en segundo de Bachillerato, sobresale el nivel bajo; y, en COU, el nivel medio. En resumen, los comportamientos de los factores extralingüísticos no son similares en las dos muestras estudiadas.

Por otro lado, los resultados de segundo de Bachillerato no permiten corroborar la hipótesis de que la muestra de los centros privados aventaja a la de los centros públicos. Tampoco el cotejo de los datos ratifica las suposiciones previas relativas a las variables sexo y a la sociocultural.

MARÍA JOSEFA REYES DÍAZ

*Universidad de Las Palmas de Gran Canaria*

#### REFERENCIAS BIBLIOGRÁFICAS

- ANDRÉS PÉREZ, B. 1997. *Riqueza léxica en textos escritos de tres niveles de EGB*. Memoria de licenciatura. Alcalá: Universidad de Alcalá.
- ÁVILA, R. 1986. Léxico infantil de México: palabras, tipos y vocablos. *Actas del II Congreso Internacional sobre el "Español de América"*. México: Universidad Nacional Autónoma de México. 510-517.
- BENÍTEZ PÉREZ, P. 2003. Consideraciones en torno a la enseñanza del vocabulario. En Francisco Moreno Fernández, Francisco Gimeno Menéndez, José Antonio Samper, M. Luz Gutiérrez Araus, María Vaquero y César Hernández (eds.). *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*. Madrid: Arco/Libros. 145-156.
- CINTRÓN SERRANO, F. 1992. *Índices de riqueza léxica en escolares de Barranquitas*. Tesis de maestría. Puerto Rico: Universidad de Puerto Rico.
- DUGAST, D. 1978. Sur quoi se fonde la notion d'étendue théorique du vocabulaire. *Le français moderne*. vol. 46. 1: 25-33.
- EACHEVERRIA, M.; A. VALENCIA *et al.* 1992. Evaluación de la riqueza léxica de estudiantes de último año de enseñanza media. *Estudios Filológicos*. 27: 59-71.



- GARCÍA ROSAS, M. P. 1996. *Riqueza léxica en textos de estudiantes de español como lengua extranjera*. Memoria de máster. Alcalá: Universidad de Alcalá.
- GUIRAUD, P. 1960. *Problèmes et méthodes de la statistique linguistique*. Paris: U.U.F.
- HACHÉ DE YUNÉN, A. M. 1991. Aportes de las pruebas de riqueza léxica a la enseñanza de la lengua materna. En *La enseñanza del español como lengua materna*. En Humberto López Morales (ed.). Puerto Rico: Universidad de Puerto Rico. 49-90.
- HAM CHANDE, R. 1979. Del 1 al 100 en lexicografía. En *Investigaciones lingüísticas en lexicografía*. México: El Colegio de México. 41-83.
- LAMÍQUIZ, V. 1974. *Lingüística Española*. Sevilla: Universidad de Sevilla.
- LÓPEZ MORALES, H. 1984. *Enseñanza de la lengua materna*. Madrid: Playor.
- MENARD, N. 1983. *Mesure de la richesse lexicale. Théorie et vérifications expérimentales. Etudes stylométriques et sociolinguistiques*. Genève-Paris: Slatkine-Champion.
- MULLER, CH. 1962. Polysémie et homonymie dans lexicque contemporain. *Etudes de linguistique appliquée*. 1: 49-54.
- \_\_\_\_\_. 1969. Comment mesurer l' étendue d'un texte. *Revue de linguistique romane*. T. 33. 131-132: 73-84.
- \_\_\_\_\_. 1971. Sur la mesure de la richesse lexicale. Théorie et expériences. *Etudes de linguistique appliquée*. 1: 20-46.
- \_\_\_\_\_. 1973. *Estadística lingüística*. Madrid: Gredos.
- PORTELA, C. 1992. *Índices de riqueza léxica en estudiantes del primer año universitario*. Tesis de maestría en Lingüística. Santiago de los Caballeros: Pontificia Universidad Católica Madre y Maestra.
- REYES DÍAZ, M. J. 2007. Apuntes para la enseñanza del vocabulario. *Revista de Filología*. 25: 529-538.
- SAMPER PADILLA, J. A. 1990. *Estudios sociolingüísticos del español de Las Palmas de Gran Canaria*. Las Palmas de G.C.: La Caja de Canarias.
- TORRES GONZÁLEZ, A. N. 1999. Incidencia de las variables sociales en los índices de producción léxica de estudiantes del último curso de la enseñanza no universitaria. *Contribuciones al estudio de la Lingüística Aplicada*. Edita Julián de las Cuevas y Dalila Fasla. Castellón: Asociación Española de Lingüística Aplicada. 393-401.
- \_\_\_\_\_. 2003a. Riqueza léxica en textos narrativos escritos por estudiantes de Tenerife. En Francisco Moreno Fernández, Francisco Gimeno Menéndez, José Antonio Samper, M<sup>o</sup> Luz Gutiérrez Araus, María Vaquero y César Hernández. (eds.). *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*. Madrid: Arco / Libros. 435-449.
- \_\_\_\_\_. 2003b. La producción léxica de estudiantes no universitarios de Tenerife. En *Estudios sobre el español de Canarias. Actas del I Congreso Internacional sobre Español de Canarias*. La Laguna y Las Palmas de Gran Canaria: Academia Canaria de la Lengua. 989-1009.