

El análisis de los datos textuales en la investigación educativa
Una experiencia sobre la opinión que tienen los alumnos de educación primaria referida al concepto de lectura

Núria Rajadell I Puiggros*

Justificación de la importancia del análisis de datos textuales en el ámbito educativo

El ámbito educativo así como las materias que con él se relacionan, ha pretendido desde sus orígenes, un enfoque lo más humano posible, debido a que es el propio ser humano su principal protagonista. A pesar de ello, la investigación educativa ha estado trabajando de manera prácticamente exclusiva y a lo largo de muchos años, con datos cuantitativos y resultados numéricos, con la idea de mostrar una mayor científicidad y, consecuentemente, teniendo abandonado el contenido específico que englobaba la expresión y opinión de los individuos.

Pero desde hace algunos años parece ser que estamos viviendo un período de transformación a favor del estudio directo, trabajando con metodologías observacionales que profundizan en la personalidad e individualidad, a las cuales será necesario incidir de manera directa si con ello se pretenden ciertos cambios. De aquí que el programa S.P.A.D. (**Systeme Portable pour l'Analyse des Données**) en su versión textual, se pueda considerar como una interesante posibilidad metodológica para el análisis del contenido educativo, sin tener que recurrir a extensos vaciados de información a través de métodos manuales y bastante imperfectos. A pesar de que este programa permite el análisis objetivo de contenidos textuales, la intervención directa del investigador sigue jugando un papel básico e indispensable.

Experiencia realizada en educación primaria

La experiencia que presentamos forma parte de una amplia investigación realizada en Barcelona, destinada al conocimiento del ambiente lector en el nivel de la educación primaria, concretamente en el ciclo medio, en sus diferentes facetas: personales, familiares, escolares y ambientales.

La población seleccionada se caracteriza por los siguientes aspectos:

- Edad: niños y niñas de 5º de E.G.B. –Educación General Básica– (la mayoría tienen entre 10 y 11 años).
- Ubicación: escuelas públicas (56%) y privadas (34%) del cinturón industrial barcelonés.
- Muestra de los individuos: 895 escolares (51.2% niños y 49.6% niñas).

* Profesora de la Facultad de Educación de la Universidad de Barcelona. Investigadora en el ámbito de la lengua: literatura infantil, intervención logopédica y, especialmente, lectura.

Para medir las actitudes lectoras de estos escolares se elaboran dos inventarios específicos, uno de los cuales contiene dos preguntas abiertas con el siguiente enunciado:

1. **Para mi leer es...**
2. **Creo que leer es importante porque...**

La primera cuestión está pensada para conocer el concepto de lectura concreto que poseen estos niños, el cual, en cierta manera, nos proporcionará una interesante información previa referida a las actitudes y los intereses lectores.

La segunda cuestión intenta conocer pequeños detalles sobre este concepto lector, como su relación con la escuela o con otros ámbitos (como imaginación, relax...).

Para realizar el análisis textual de las respuestas emitidas por los escolares a estas dos cuestiones abiertas, efectuamos en primer lugar una búsqueda de materiales existentes por todo el mundo, de acuerdo con los objetivos pretendidos; y así llegamos a conocer la existencia del programa S.P.A.D. del cual seleccionamos aquellas etapas que consideramos más apropiadas e interesantes para nuestro estudio.

El procedimiento efectuado para cada una de las dos cuestiones ha sido muy paralelo. En primer lugar, aseguramos el registro de datos textuales (fijando los espacios correspondientes, número de respuestas, tipología del texto así como su longitud) y efectuamos una selección del texto con el que pretendemos trabajar a partir de ahora, en nuestro caso el primero.

Seguidamente, a partir de la etapa COTEX realizamos la numeración del texto, reconociendo las diferentes palabras utilizadas, calculando sus frecuencias, seleccionando las más frecuentes y construyendo una tabla de contingencias (respuestas en línea y palabras en columna). Permite además obtener el listado de formas lexicales a partir del orden alfabético o de la frecuencia de aparición.

En nuestra experiencia observamos que el número total de palabras es de 5.723, siendo 626 diferentes (correspondientes al 11%). A partir del análisis alfabético, y teniendo además en cuenta su frecuencia, decidimos efectuar una agrupación de palabras con significado idéntico (p.e.: **aburrido**, **aburrida**, **aburridos**,... engloban un total de diez palabras diferentes, que concentramos en una sola).

Mostramos una primera visión general de los resultados de esta etapa en el **Cuadro 1**, en el que agrupamos los nombres a partir de características muy generales, las cuales solamente nos proporcionan una primera visión muy aproximativa.

Cuadro 1	
Distribución general de las palabras del texto global	
Partículas sueltas (preposiciones, adverbios, artículos...)	2698
Nombres comunes	1061
Adjetivos	601
Verbos	1261

Podemos desglosar la especificación concreta de las palabras más frecuentes en tres bloques diferenciados (**Cuadro 2**):

- Palabras con sentido positivo: encontramos palabras que pretenden definir la lectura como **diversión** (347), **aventura** (64); otras que van más allá de este concepto y que acostumbran a ser adjetivos: **importante** (145), **bonito** (77), **entretenido** (70), **interesante** (41) y, por último, un reflejo de la importancia que posee la lectura en el ámbito escolar: **aprender** (201), **saber** (40).

- Palabras con sentido negativo: **aburrir** (64).

- Palabras neutras: la mayoría de ellas no poseen significación por sí mismas, como **ser** (293), **cosa** (174), **mucho** (89), **estar** (70), **pasar** (69); a pesar de que encontramos **leer** (88), **libro** (73).

Cuadro 2			
Palabras más frecuentes ordenadas en función del sentido positivo, negativo o neutro			
Palabras con sentido positivo		Palabras con sentido neutro	
—diversión	347	—ser	293
—aventura	64	—cosa	174
—importante	145	—mucho	89
—bonito	77	—estar	70
—entretenido	70	—pasar	69
—interesante	41	—leer	88
—aprender	201	—libro	73
—saber	40		
Palabras con sentido negativo			
—aburrir	64		

La etapa SETEX se relaciona directamente con la etapa anterior, ya que realiza la numeración del texto pero añadiendo los caracteres específicos de cada palabra. En nuestro caso trabajamos con todas las palabras recopiladas, obteniendo (según se muestra en el **Cuadro 3**) como más frecuentes las que poseen 5, 6 y 7 caracteres o letras.

Cuadro 3
Distribución de las palabras en función de sus caracteres o letras

Caracteres que incluye la palabra															
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4	23	33	59	115	91	92	56	55	46	29	15	5	1	2	0
Efectos observados															

En el **Cuadro 4** podemos observar estas cifras concretas, traducidas en palabras específicas.

Cuadro 4
Palabras más frecuentes en función de sus caracteres

Palabras con cinco caracteres					
—cosas	119	—libros	39	—mejor	18
—gusta	56	—veces	35		
—mucho	48	—pasar	20		
Palabras con seis caracteres					
—porque	119	—nuevas	37	—enseña	17
—bonito	66	—libros	32		
—cuando	45	—puedes	17		
Palabras con siete caracteres					
—leyendo	22	—algunas	15	—algunos	15
—aprendo	19				

La etapa SEGME reconoce los diferentes segmentos repetidos o cadenas de palabras vecinas no delimitadas por un separador fuerte, calcula además su frecuencia, seleccionando los más significativos y construyendo una tabla de contingencia (respuestas en línea y segmentos en columna) en la que aparecen ordenados alfabéticamente. En general los segmentos se relacionan con aspectos positivos de la lectura considerándola interesante, divertida o agradable, según se muestra en el **Cuadro 5**.

El hecho de considerar la lectura como un tema importante no incluye forzosamente interés y agrado por ella, a pesar de dejar entrever un objetivo a realizar. La palabra **muy** aparece a menudo; en algunos casos no podemos llegar a descubrir su último destino, cuando se nos presentan segmentos como **es una cosa** o **para mi leer es**. Poseemos un único segmento negativo, **no me gusta**, con una frecuencia de 12.

Cuadro 5
Segmentos más frecuentes

Segmentos de palabras más frecuentes	
—muy importante	94
—muy divertido	91
—divertido y	60
—me gusta	55
—una diversión	52

La etapa ASPAR efectúa los cálculos y las ediciones de un análisis de correspondencias a partir de una tabla de contingencia en la que cruzamos las

respuestas individuales con las palabras, y además, nos permite contemplar las palabras más representativas desde dos ángulos diferentes pero complementarios: listado de palabras y representación gráfica.

Para la interpretación de los ejes hay que tener en cuenta dos coeficientes: contribuciones absolutas (reflejan la importancia que posee una palabra para la construcción de un eje) y las contribuciones relativas (reflejan la importancia que posee un eje para una palabra). Mientras las absolutas permiten conocer las variables responsables de la construcción de un factor, las relativas explican en qué medida el eje refleja la posición que tiene la palabra.

En nuestra investigación trabajamos a partir de las 92 palabras más significativas, ya que son las que más han contribuido para la elaboración de un eje (**Cuadro 6**).

Cuadro 6			
Palabras más significativas para la construcción de los ejes			
Eje N° 1		Eje N° 2	
un	11.9	aburrimiento	19.2
muy	11.5	divertido	9.0
aburrimiento	11.2	un	4.7
divertido	9.2		
importante	6.0		
interesante	6.0		
Eje N° 3		Eje N° 4	
divertirme	20.0	pasar	6.3
una	17.6	el	4.6
aprender	10.3	interesante	4.4
diversión	9.1	bueno	4.3
cosa	7.2	cosas	4.3
Eje N° 5			
divertirme	34.1		
diversión	7.5		
aburrimiento	5.4		

La etapa GRAFH ofrece la posibilidad de realizar representaciones gráficas sin embargo cuantos más datos tengamos más difícil será su interpretación. Hay que diferenciar en ella tres partes: representación del eje de coordenadas (las palabras se van distribuyendo a su alrededor, siendo más comunes a medida que se acercan más al eje central), relación de palabras ubicadas en los extremos (aparecen representadas de manera parcial o no aparecen) y relación de palabras céntricas (aparecen representadas de manera parcial o se encuentran escondidas debajo de otras).

Las representaciones gráficas más significativas se muestran en la **Figura 1**, **Figura 2**, **Figura 3** y **Figura 4**.

Figura 1. Representación gráfica del eje de coordenadas 1,2

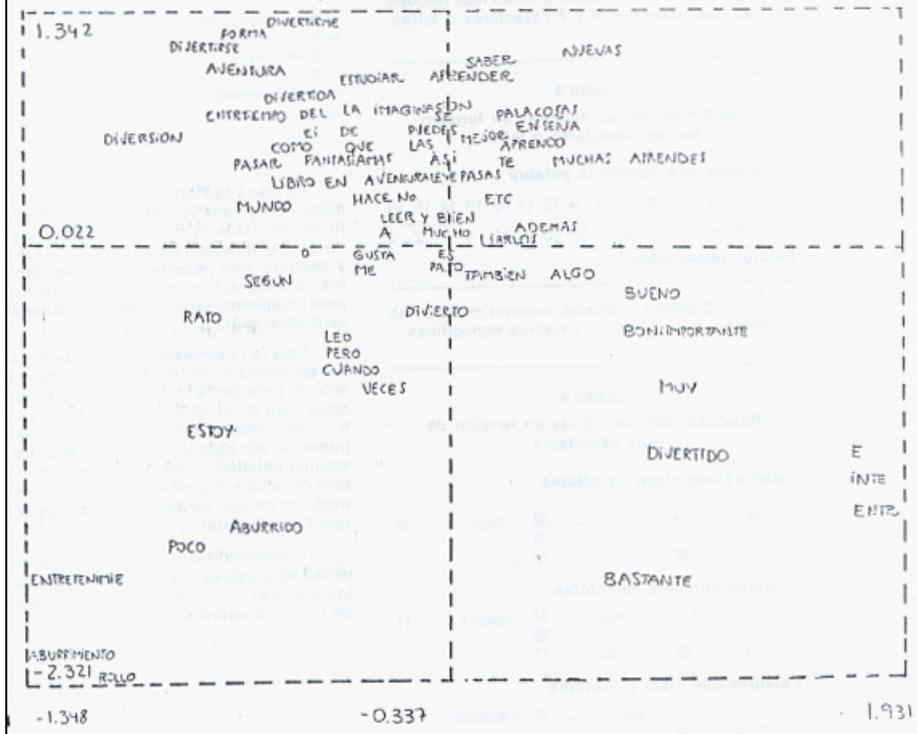
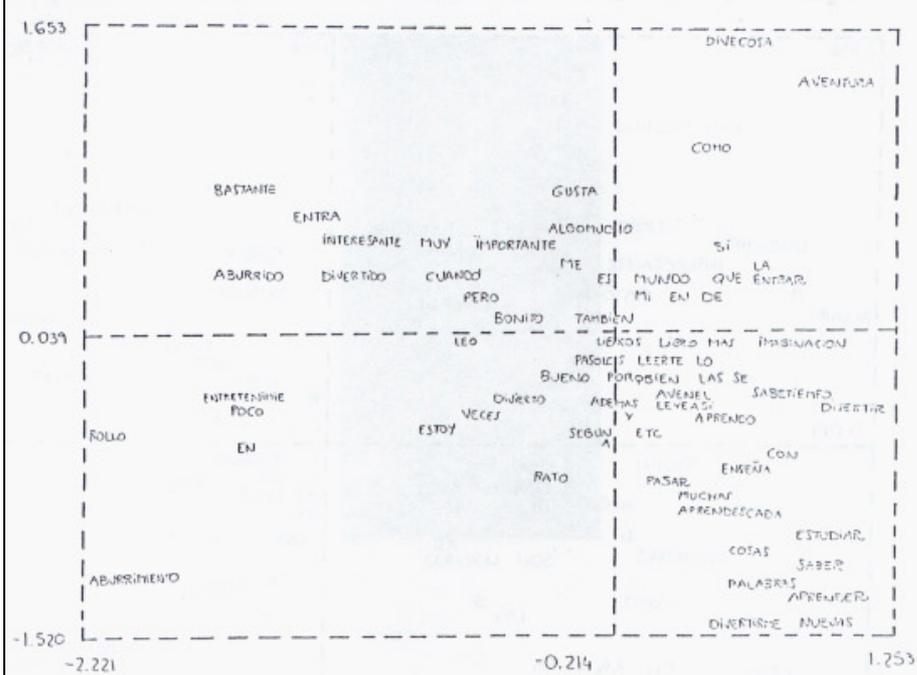
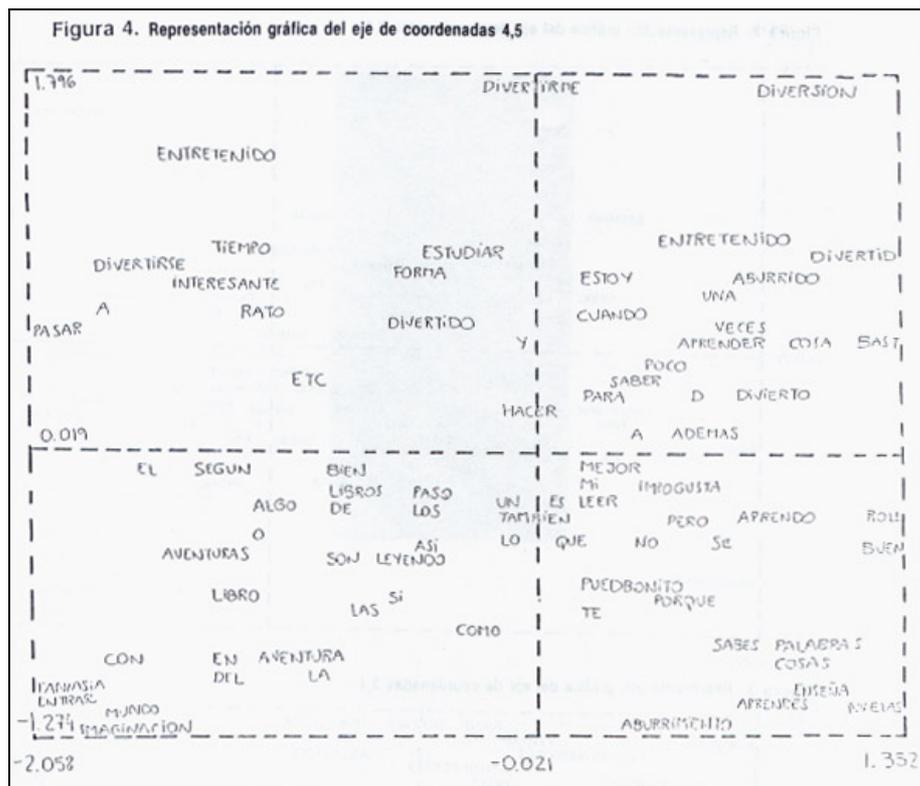
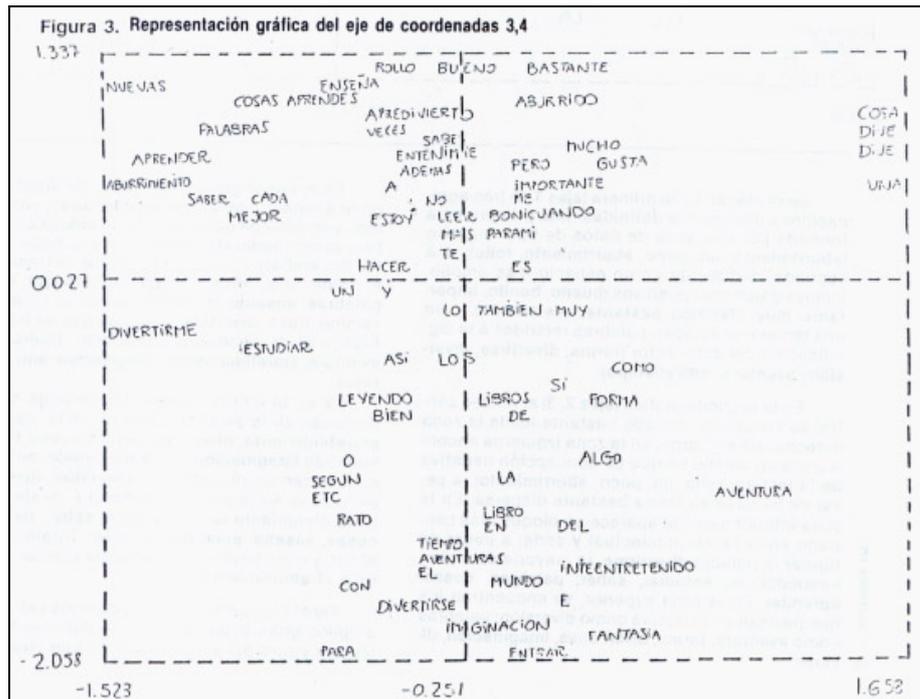


Figura 2. Representación gráfica del eje de coordenadas 2,3





Se destacan en la primera (ejes 1, 2) tres agrupaciones claramente definidas: una primera está formada por una serie de datos de tipo negativo (**aburrimiento, un, poco, aburrimiento, rollo**); una segunda, distribuida en un espacio más amplio, incluye datos más positivos (**bueno, bonito, importante, muy, divertido, bastante, interesante**), y en una tercera se agrupan palabras referidas a la significación del acto lector (**forma, divertirse, diversión, aventura, entretiem**).

En la segunda gráfica (**ejes 2, 3**) el núcleo central se encuentra ubicado bastante hacia la zona derecha; sin embargo, en la zona izquierda encontramos un amplio bloque de concepción negativa de la lectura (**rollo, un, poco, aburrimiento**), a pesar de hallarse en forma bastante dispersa. En la zona inferior derecha aparece un bloque más centrado en la faceta intelectual y seria; a pesar de figurar la palabra **divertirme**, la mayoría se sitúan alrededor de: **estudiar, saber, palabras, cosas, aprender**. En la zona superior, se encuentran los que piensan en la lectura como diversión: palabras como **aventura, forma, diver, cosa, imaginación, divertir**.

En el tercer gráfico (**ejes 3, 4**), de diseño alargado alrededor del eje de coordenadas, encontramos una serie de palabras relacionadas con la lectura como medio de distracción y entretenimiento. Sin embargo, nos interesan los extremos: en el noroeste aparece el grupo de **aprender, cosas, palabras, enseñar**; mientras que en el noreste **diversión, cosa, divertida**, palabras que se irán perfilando al descender en el espacio: **forma, algo, aventura, entretenimiento, imaginación, entrar, fantasía**.

Y en la última representación (**ejes 4, 5**) se destacan en la parte superior hacia la izquierda: **entretenimiento, diversión, pasar un rato, tiempo**; mientras **imaginación, fantasía, mundo, entrar**, se encuentran en el sudoeste. **Diversión, divertirse**, en la zona norte y noroeste. En el sudeste se entrevé claramente la faceta seria: **saber, palabras, cosas, enseña, aprendes, nuevas**. Totalmente en el sur, y contrariamente a la faceta diversión, aparece el aburrimiento.

De estos cuatro gráficos podemos extraer dos amplios grupos: uno en función del concepto de lectura y otro del estado anímico que viven en el momento de la práctica lectora. En el primer grupo, para unos lectura equivale a una manera de aprender (**aprender, estudiar, saber palabras nuevas,...**); para otros no es más que una forma de diversión (**diversión, entretenimiento, pasar el rato...**). En el segundo grupo, teniendo presente el concepto lector inmerso, encontramos aquellos que lo pasan bien (**es divertido, es hermoso...**) y los que parece que no tanto (**es aburrido, es un rollo,...**).

Las etapas MOTEX, MOCAR y RECAR permiten observar la distribución específica de los textos presentados, precisada en nuestro caso a través de seis clases (**Tabla 1**):

Nº Texto	Identificador	Nº Individuos	Nº Respuestas
1	aa1a clase 1/6	104	104
2	aa2a clase 2/6	146	146
3	aa3a clase 3/6	315	269
4	aa4a clase 4/6	23	23
5	aa5a clase 5/6	105	105
6	aa6a clase 6/6	202	202
Totales		895	849

Los textos con mayor densidad son el tercero y el sexto, mientras el cuarto es extremadamente reducido; reflejado de nuevo en la **Tabla 2**.

Identificador	N° de palabras	% del total	X por respuesta	N° palabras diferentes	% palabras del texto	N° palabras retenidas
aa1a	691	121.2	6.6	49	70.9	531
aa2a	1537	269.5	10.5	55	35.8	1029
aa3a	2113	370.5	6.7	54	25.6	1543
aa4a	77	13.5	3.3	12	155.8	56
aa5a	649	113.8	6.2	45	69.3	490
aa6a	636	111.5	3.1	33	51.9	516
GLOBAL	5703	1000.0	6.4			4165

A partir de aquí, a través de etapas diferentes, podemos obtener información puntual respecto a palabras, frases breves y textos largos, que han contribuido a la construcción de estas seis clases.

La **Tabla 3** nos muestra las palabras más significativas junto al número de apariciones que han tenido a lo largo de los textos.

Clase N° 1	AA1A	Clase N° 2	AA2A	Clase N° 3	AA3A	Clase N° 4	AA4A
—un	81	—de	61	—y	133	—divertirme	23
—a	30	—y	57	—cosas	111	—y	14
—me	29	—es	52	—aprender	98	—aprender	9
—veces	28	—en	50	—es	97		
—es	27	—el	46	—muy	91		
—aburrido	25	—un	41	—porque	78		
—algo	25	—la	41	—leer	65		
—divertido	25			—importante	57		
		Clase N° 5	AA5A	Clase N° 6	AA6A		
		—una	105	—muy	135		
		—cosa	43	—divertido	118		
		—diversión	38	—importante	63		
		—y	36	—y	46		
		—que	33	—interesante	33		
		—me	21	—es	29		

Los segmentos breves más representativos de cada clase se muestran en el **Cuadro 6**, estando formados por agrupaciones de palabras, que en nuestro caso, incluyen entre dos y siete.

Cuadro 6	
Distribución de las frases breves más significativas de cada clase	
Texto N° 1 . AA1A. Clase 1/6	Texto N° 2. AA2A. Clase 2/6
1. un gusto	1. pasar el tiempo
2. un aburrimiento	2. entrar en el mundo de la fantasía
3. un aburrimiento	3. pasar aventuras
4. un hobby	4. como pasar aventuras imaginarias en la frente
5. un rollo	5. comprender la lectura
6. un pasatiempo	6. soñar con la imaginación
7. pues leerme un cuento	7. meterme en la aventura del libro
8. un aburrimiento	8. pasar el rato bien, sin aburrirte
9. un rollo	9. entras en alguna aventura
10. joby	10. pasar el tiempo bien
Texto N° 3. AA3A. Clase 3/6	Texto N° 4. AA4A. Clase 4/6
1. aprendo cosas buenas	1. divertirme
2. aprender buenas cosas	2. divertirme
3. aprender cosas	3. divertirme
4. aprender cosas	4. divertirme
5. aprender bastante: cosas	5. divertirme
6. descubrir cosas nuevas	6. divertirme
7. aprender cosas nuevas	7. divertirme
8. aprender nuevas cosas	8. divertirme y distraerme
9. aprender cosas nuevas	9. divertirme y aprender
10. aprender	10. divertirme y aprender
Texto N° 5. AA5A. Clase 5/6	Texto N° 6. AA6A. Clase 6/6
1. una tontería	1. divertido
2. una actividad del colegio	2. divertido
3. una distracción	3. divertido
4. una distracción	4. divertido
5. una historia fantástica	5. divertido
6. una ilusión	6. divertido, aventurero, estudioso
7. una diversión	7. divertido
8. una diversión	8. divertido
9. una diversión	9. bastante divertido
10. una diversión	10. divertido

Algunos textos utilizan muy pocas palabras (p.e. 4 y 6 con **divertirme, divertido**); otros a pesar de poseer cierta variedad lexical (p.e, 1 y 3) engloban un solo concepto; en el 5 observamos una cierta gradación conceptual que va desde **una tontería** hasta **una diversión**. El 2 está constituido por segmentos más amplios y formados, enfocados hacia el vivir la aventura.

Podemos obtener las frases más extensas a partir de otra etapa, cuyos resultados presentamos, de manera parcial, en el **Cuadro 8**.

Cuadro 8		
Distribución de las frases entre los diferentes textos		
Texto N° 1.	AA1A	Clase 1/6
1. "Un poco aburrido, pero a veces leo, en mi casa antes cada día me leía un capítulo de un libro llamado: dos años de vacaciones. "		
2. "Bonito pero como leo un poco mal no me gusta, pero cuando leo un poco más rápido me gusta."		
3. "Es como un cosa que es fácil pero que a veces es un poco complicado, pero es muy chulo."		
4. "Un poco aburrido pero a veces me divierto cuando estoy aburrido."		
.....		
10. "Es muy guay pero leo con alguna dificultad pero ahora ya leo porque como leo mal pues voy a leer . . .".		
Texto N° 2.	AA2A	Clase 2/6
1. "Depende, si el libro es de aventuras o de historia; si es de historia no me gusta si es de aventuras me encanta."		

Si hasta ahora solamente poseíamos una única información (p.e. el concepto **aburrido**) podemos a partir de aquí establecer la gradación de ese aburrimiento (p.e. **un poco aburrido, pero a veces leo...** o **... a veces me**

divierte cuando estoy aburrido). Es posible observar que el texto 4 es muy pobre, y está centrado alrededor de **divertirme** y otro verbo (en muchas ocasiones **aprender**); el texto 5 sigue misma tónica aunque con mayor amplitud; el texto 6 define los adjetivos que incluye la lectura, para la mayoría ésta es un acto **importante, interesante** y hasta para algunos **entretenido**. El texto 3 está formado por todos aquellos que consideran la lectura como una forma de aprendizaje. El primero se centra en el aburrimiento, desde todas sus facetas posibles. Finalmente, el texto 2 profundiza en la tipología de los libros, la importancia del libro de aventuras, e incluso algunos temas aislados como leer mejor en clase o las faltas de ortografía.

Considerando estas mismas etapas pero teniendo en cuenta la variable sexo obtenemos un total de 849 respuestas (437 niños, 412 de niñas). El análisis efectuado con ellas (**Tabla 4**) nos demuestra que a pesar de que la media de palabras por texto es menor en las niñas que en los niños, en conjunto obtenemos más palabras de las niñas.

Identificador	Nº de palabras	% del total	X por respuesta	Nº palabras diferentes	% palabras del texto	Nº palabras retenidas
Sx01	2544	463.6	10.7	63	22.4	1934
Sx02	3059	536.4	7.0	59	19.3	2231
Global	5703	1000.0	6.4			4165

Las palabras más frecuentes son en los niños: **y** (131), **muy** (126), **es** (100), **divertido** (99), **una** (79), **importante** (70), **aprender** (59), **porque** (57), **me** (50); y en las niñas: **y** (175), **muy** (54), **es** (126), **divertido** (116), **que** (86), **me** (79), **porque** (67), **aprender** (66), **leer** (64), **una** (58). Encontramos algunas diferencias concretas: **libros** o **leer** aparecen más en las niñas, mientras que en **estudiar** o **aburrido** dominan los niños.

El sexo masculino se caracteriza por considerar que la lectura es una tarea bastante aburrida (el 80%), a pesar de encontrar algún caso aislado que la considere divertida. En las niñas encontramos una gran dispersión conceptual aunque en un ambiente positivo: para la mayoría es divertido (**divertido, entretenerme...**) y para las restantes instructivo (**aprendo cosas, te enseñan cosas...**).

En las frases largas no encontramos demasiadas diferencias ya que existen los defensores de la lectura como acto divertido y los que no, sin olvidar una cierta visión prospectiva en muchos de los casos (**mejorar la escuela, comprender los anuncios de los periódicos...**); las niñas parecen acusar más el aspecto lúdico de la lectura, sin olvidar la importancia que posee como fuente de aprendizaje.

Conclusiones

Podemos concentrar las conclusiones obtenidas a partir del análisis en las siguientes:

- **Los niños y niñas observan la lectura como una actividad positiva.** Un elevado porcentaje de alumnos contempla la lectura de manera positiva, sin especificar demasiados detalles.

Encontramos algunos casos de indiferencia, y una minoría que no la considera demasiado agradable hasta el punto de llegar a definirla como una tarea aburrida. Niños y niñas concuerdan en la importancia de saber leer bien, tanto para hoy como para mañana.

- **La lectura engloba una doble faceta: intelectual y lúdica.** La mayoría de los niños y niñas que forman la muestra tienen en cuenta esta doble faceta, a pesar de proporcionarnos opiniones muy diversas: para unos son dos campos totalmente separados, porque cuando se lee en la escuela se está llevando a cabo una tarea única y exclusivamente intelectual, bastante alejada de pasarlo bien; para otros, sin embargo, existe una relación estrecha entre ambos, ya que podemos leer desde un punto de vista intelectual, pero a su vez pasarlo bien, como también podemos leer por placer y, simultáneamente ir aprendiendo.

A pesar de la diversidad de opiniones siempre se acaba en el mundo del aprendizaje; sin embargo, los niños muestran un mayor realismo que las niñas.

- **El concepto general de lectura varía en función del sexo.** La opinión que poseen niños y niñas sobre la lectura es en general diferente; por un lado las niñas la consideran más positiva que los niños, calificándola como divertida e instructiva, mientras que los niños reflejan cierto aburrimiento.

- **Las niñas expresan los conceptos con más brevedad que los niños.** La media de palabras por texto es menor en las niñas que en los niños, a pesar de que aquellas que tienen ganas de escribir lo hacen de manera más extensa que los niños repitiendo reiteradamente una misma opinión.

Referencias bibliográficas

- Baudelot, Ch. (1987) **Reussir sa vie**. Nantes, Université de Nantes.
Fenelon, J.P. (1981) **Qu'est-ce que l'analyse des données**. Paris, Lefonen.
Lebart, L. (1988) **Analyse statistique des réponses libres dans les enquetes par sondage**. Paris, Centre National de la Recherche Scientifique.
Lebart, L. y A. Morineau (1985) **S.P.A.D.T. Systeme portable pour l'analyse des données**. Paris, Cisia.
Lebart, L.; A. Morineau y J.P. Fenelon (1985) **Tratamiento estadístico de datos**. Barcelona, Marcombo.
Lebart, L. y A. Salem (1988) **Analyse statistique des données textuelles**. Paris, Bordas.