

PRODUCCION ESCRITA: ESTRUCTURA FACTORIAL

MARÍA ESPERANZA HERRERA GARCÍA
INÉS MARTÍNEZ GARCÍA

Este estudio se planteó con la idea de comprobar de forma empírica si los indicadores lingüísticos obtenidos en investigaciones realizadas sobre comprensión lectora¹ aparecían también como significativos en el análisis de la producción escrita.

Si bien la investigación psicológica ha demostrado en numerosas ocasiones la relación existente entre comprensión y producción en el sentido de una maduración anterior de la primera con respecto a la segunda², podemos considerar a priori la existencia de indicadores que nos permitan medir ambos aspectos.

Para ello buscamos la estructura factorial que presentan los indicadores lingüísticos que vamos a considerar y comprobaremos si presenta los factores encontrados en comprensión lectora³.

Los indicadores lingüísticos que hemos utilizado se han obtenido a partir de diversas fuentes anteriores, en especial López Rodríguez⁴. Se trata de medidas indirectas altamente correlacionadas con la comprensión lectora y la complejidad sintáctica; algunos de ellos han sido reformulados⁵ para paliar el efecto que produce la longitud del texto (var. 13: Total de palabras) en las variables que hacen referencia a aspectos de vocabulario⁶, debido a la gran variación que presenta la misma en las muestras obtenidas de lenguaje escrito.

Los indicadores son los siguientes:

Var. 1: «Pronombres»: Número total de pronombres del texto obtenido según la fórmula:

$$\frac{\text{número de pronom. 1 y 2 persona}}{\text{número total de palabras texto}} \times 1000$$

1. BELTRÁN DE TRENA, R. (1983): «El proceso de comprensión lectora». *El sistema educativo hoy*, pp. 403-414. CINAIE, Buenos Aires.

2. BENEDICT, H. (1979): «Early lexical development: comprensión and production». *Journal of child language*, n. 6, pp. 183-200. PERAITA, H. (1986): «Desfase entre la producción y comprensión del léxico...». *Infancia y aprendizaje*, n. 35-36, pp. 11-24.

3. BELTRÁN DE TENA, R. o.c. p. 412.

4. LÓPEZ RODRÍGUEZ, N. (1981): *Fórmulas de legibilidad de la lengua castellana*. Tesis doctoral, Universidad de Valencia y —, (1982): *Cómo valorar textos escolares*. Cincel-Kapelusz, Madrid.

5. HERRERA GARCÍA, M. E. (1985): *Variables lingüísticas y rendimiento escolar...* Memoria de Licenciatura, universidad de Salamanca. MARTÍNEZ GARCÍA, I. (1985): *Variables lingüísticas y rendimiento escolar...* Memoria de Licenciatura, Universidad de Salamanca.

6. MULLER, Ch. (1973) *Estadística lingüística*. Gredos, Madrid.

Var. 2: «Numerales»: Número total de numerales del texto obtenido según la fórmula:

$$\frac{\text{Número de numerales}}{\text{Número total de palabras}} \times 1000$$

Var. 3: «Desviación típica de palabras»: medida del grado de dispersión de la longitud de las palabras.
Desviación típica \times Total de palabras ^{1/2}

Var. 4: «Vocabulario Común»: Medida de la correspondencia del Vocabulario obtenido con referencia al Vocabulario Común de García Hoz ⁷ a través de la fórmula:

$$\frac{\text{Total de palabras} - \text{Palabras del Diccionario}}{\text{Log. [Palabras total del texto]}}$$

Var. 5: «Palabras/frase»: Proporción de palabras por frase según la fórmula:

$$\frac{\text{Número de palabras totales del texto}}{\text{Número de frases}}$$

Var. 6: «Total de puntos»: Proporción de puntos en el texto según la fórmula:

$$\frac{\text{Número de puntos total del texto}}{\text{Número palabras totales del texto}} \times 1000$$

Var. 7: «Total de puntos y aparte»: Proporción de los puntos y aparte en el texto según la fórmula:

$$\frac{\text{Número total de puntos y aparte}}{\text{Número total de palabras del texto}} \times 1000$$

Var. 8: «Comas»: Proporción de comas en el texto según la fórmula:

$$\frac{\text{Número de comas totales del texto}}{\text{Número de palabras totales del texto}} \times 1000$$

Var. 9: «Media + Desviación típica»: Longitud máxima de palabra obtenida a través de la fórmula:

$$[X + 2,58 \text{ D.T. de palabras}] \times \text{Total de palabras}^{1/2}$$

Var. 10: «Puntos y coma»: Proporción de puntos y coma en el texto:

$$\frac{\text{Número de puntos y coma total del texto}}{\text{Número total de palabras texto}} \times 1000$$

7. GARCÍA HOZ, V. (1953): *Vocabulario usual, común y fundamental*. C.S.I.C., Madrid.

Var. 11: «Nombres propios»: Proporción de nombres en el texto teniendo en cuenta que han de ser palabras escritas con mayúscula y que no vayan detrás de un punto:

$$\frac{\text{Número de mayúsculas} - \text{total de puntos}}{\text{Número total de palabras del texto}} \times 1000$$

Var. 12: «Tasa de Redundancia Modificada»: Modificada de la de López Rodríguez (1982) y utilizándola para el lenguaje codificado (lexemas), según la fórmula:

$$\frac{\text{Número de palabras distintas del texto}}{\text{Log. [Número total de palabras del texto]}}$$

Var. 13: «Total de palabras»: Número total de palabras que aparecen en el texto del sujeto.

Además hemos incluido una variable que hemos denominado «Rendimiento» (variable n.º 14), obtenida a partir de la media aritmética de las dos últimas evaluaciones de los alumnos en las materias de lengua, Sociales, Naturales, Matemáticas e Idioma.

Desarrollo de la investigación

La muestra utilizada está constituida por 354 alumnos del último curso de E.G.B. de la provincia de Salamanca.

Se les proporcionó a los niños como estímulo para la realización de una redacción una tarjeta postal, con objeto de proporcionar un tema más o menos cerrado que neutralizara el efecto diversificador que pudieran ejercer diferentes temas, y considerando que un estímulo icónico podría ser más cerrado que un estímulo verbal. El tiempo de duración de la prueba fue de diez minutos.

Las redacciones fueron introducidas en el ordenador respetando la redacción de los niños, con la excepción de los errores ortográficos que fueron corregidos debido a que el proceso de análisis exigía la comparación de los vocablos⁸ utilizados por los niños con los diccionarios utilizados⁹. En aquellas redacciones en las que no existía el punto final se añadía, por ser necesario para el cómputo correcto de algunas de las variables¹⁰. En un segundo proceso se procedió al despojo de las palabras para obtener los vocablos de cada sujeto. Una vez preparados los textos se los sometió al análisis para obtener los indicadores lingüísticos ya descritos. Con estos datos se realizó un tratamiento estadístico consistente en un análisis factorial con rotación varimax para la obtención de los factores.

8. Distinción entre palabras y vocablos: MULLER, Ch.: o. c. pp. 230 y ss.

9. Se han utilizado los siguientes: pronombres, numerales y Vocabulario Común de García Hoz.

10. Variable n.º 5, variable n.º 6 y variable n.º 7.

La matriz de correlaciones obtenida fue la siguiente:

FACTORES SIN ROTAR

FACTOR 1

1	-.1348
2	.0434
3	.5851
4	.7578
5	.0545
6	.0389
7	-.1693
8	-.0309
9	.9704
10	.0163
11	-.2254
12	.9234
13	.9239
14	.2425

FACTOR 2

1	-.1111
2	.1757
3	-.0161
4	.0545
5	.4483
6	-.1157
7	-.3605
8	.237
9	-.0361
10	-.0355
11	-.003
12	-.121
13	-.0474
14	-.0515

Raíz latente: 3.7292228
Saturación mínima: .1368

Raíz latente: .48983662
Saturación mínima: .142

FACTORES ROTADOS

FACTOR 1

1	-.1278
2	.0331
3	.5857
4	.7547
5	.0453
6	.0395
7	-.1636
8	-.0313
9	.9708
10	.0174
11	-.2242
12	.9271
13	.9222
14	.2479

FACTOR 2

1	.0227
2	-.0202
3	.0047
4	.1495
5	.3927
6	-.1464
7	-.3634
8	.3232
9	-.022
10	-.0199
11	.0232
12	-.0569
13	-.0917
14	.0703

Estos resultados nos muestran una agrupación de las variables en dos factores, uno de ellos relacionado con la riqueza de léxico: el primero, que hemos denominado de Vocabulario debido al mayor peso de las variables relacionadas con la misma; y el segundo denominado Fraseológico, en el que tienen el mayor peso las variables relacionadas con aspectos formales.

Aparece claramente un primer factor al que está asociada la mayoría de la varianza. Está constituido por cinco variables ¹¹ relacionadas con aspectos de vocabulario, que son: el total de palabras empleadas (tasa de producción del sujeto), las palabras que no pertenecen al Vocabulario Común de García de Hoz (lo cual nos indica que una mayor productividad del sujeto va a ir asociado a un vocabulario más rico y menos usual); también se encuentra en este factor la variable que mide la tasa de redundancia ¹²; junto a éstas, aparecen variables que tienen que ver con el tamaño de las palabras, son las variables n.º 9 (Media + 2,58 D.T.) y n.º 3 (Desviación típica de palabras), la primera con un peso muy superior al de la segunda.

La tasa de redundancia (TTR) aparece asociada en sentido positivo con el total de palabras, lo cual puede ser debido a que las muestras de los sujetos no son lo suficientemente grandes para agotar las opciones de vocablos que tienen los mismos y por esto, en general, un mayor número de palabras se va a corresponder con un mayor número de vocablos. Esto también puede tener relación con el hecho de que la mayoría de las redacciones sean de tipo descriptivo y ninguna agota todos los elementos que aparecen en el estímulo visual presentado, por lo que los únicos elementos de redundancia van a ser elementos sin carga semántica.

Por lo que respecta al tamaño de las palabras hemos de indicar que la asociación encontrada es lógica ya que una mayor productividad (total de palabras) va a ir asociada a una mayor riqueza de léxico y ésta a su vez a la utilización de palabras menos usuales y por tanto, de mayor tamaño (de acuerdo con la Ley de Zipf ¹³). Esto también tendría relación con la variabilidad del tamaño de las palabras.

El segundo factor, mucho menos claro, con una raíz latente muy pequeña y con bajas saturaciones, está constituido por las variables: palabras por frase, puntos y aparte y comas; todas ellas miden aspectos relacionados con la complejidad estructural de la frase. Las dos primeras variables mantienen entre sí una relación inversa, lo cual puede ser debido a una inadecuada utilización de los signos de puntuación por parte de los niños del nivel estudiado: no utilizan los puntos y seguido para separar frases sino que utilizan los puntos y aparte, dando lugar así a párrafos unifrased. Además en este factor se incluye con un sentido inverso a los puntos y aparte la variable comas, lo cual confirma lo anteriormente dicho.

La variable rendimiento escolar aparece sólo en el primer factor y con una saturación muy baja, que nos indica una pequeña relación con aspectos relacionados con el vocabulario.

Los dos factores encontrados, en cuanto a su composición, coinciden con los dos factores encontrados en los estudios sobre comprensión lectora, aunque se observan diferencias de tipo cualitativo y cuantitativo.

Diferencias de tipo cuantitativo (una menor saturación de las variables) en el factor de Vocabulario, debidas a un menor dominio del mismo por parte de los sujetos, que produce una homogeneización del grupo, lo cual puede hacer disminuir el valor de la correlación.

En el segundo factor se observan diferencias cualitativas y cuantitativa: las primeras en relación con la no aparición de las variables total de puntos como discriminante,

11. Siguiendo a COMREY, A. (1985): *Manual de análisis factorial*. Cátedra, Madrid.

12. Según MULLER, Ch. (1973): o. c., el número de vocablos utilizados por el sujeto aumenta al aumentar el tamaño de su discurso hasta llegar a un punto característico de cada sujeto en el que la relación entre vocablos y las palabras se estabiliza.

13. ZIPF, G. K. (1949): *Human Behavior and de Principie o least Effort* Addison-Wesley, Reading (Mass).

y en la aparición de la variable comas en su lugar. Esto también puede ser debido al mismo problema que plateábamos con respecto al primer factor pero en este caso muy acentuado por el escaso dominio de los elementos de puntuación de los sujetos de la muestra, dominio que es escasísimo si se compara con el que poseen los autores que se tomaron como muestra para los estudios de comprensión lectora. Esta circunstancia introduce un componente de aleatoriedad en la puntuación que se observa de forma muy clara si se acude directamente a las redacciones de los niños.

Conclusiones

Los resultados del estudio realizado parecen confirmar claramente la existencia de un factor léxico o de vocabulario en el ámbito de la producción escrita que coincide plenamente con el hallado en los estudios sobre comprensión lectora¹⁴. No podemos decir lo mismo en relación al factor que hemos denominado fraseológico o sintáctico, el cual explica una mínima de la varianza. Esto puede ser debido, como ya hemos indicado, al escaso dominio de la puntuación que tienen los sujetos utilizados como muestra, y que imposibilita la discriminación intersujetos por parte de las variables utilizadas. Por el contrario, en el primer factor aparecen una serie de variables asociadas que van a permitir una discriminación entre sujetos. Esto puede ser debido a que los sujetos de 8.º de E.G.B. si parecen tener un control sobre los aspectos relacionados con el vocabulario: es decir, no parece existir dificultad en la transcripción del pensamiento al papel a nivel de vocabulario (actualización del léxico en vocabulario), aunque existen diferencias en cuanto a calidad y extensión del léxico; pero sí parecen tener dificultades por lo que se refiere a la estructuración formal de ese pensamiento (las frases están correctamente construídas a nivel morfosintáctico, pero mal puntuadas).

Esta intuiciones derivadas del análisis somero de los discursos de los sujetos requerirían para su constatación real una mayor profundización en su estudio.

MARÍA ESPERANZA HERRERA GARCÍA e INÉS MARTÍNEZ GARCÍA
Departamento Didáctica y Organización Escolar.
Facultad de Filosofía y CC. E.
Paseo Canalejas, 169.
37008 Salamanca

14. «El factor léxico tiene un peso definitorio en las condiciones de producción y de recepción del discurso en situación de aula»: RODRÍGUEZ DIEGUEZ, J. L. (1984): «Códigos lingüísticos y procesos de enseñanza» en *Educación y sociedad plural*, vol. 1, Actas del VIII Congreso de Pedagogía, Santiago. p. 132.