

Determinación de la comprensibilidad de materiales de lectura por medio de variables lingüísticas

Nelson Rodríguez Trujillo*

Uno de los problemas de la educación popular y de la comunicación masiva es el de determinar hasta qué punto un material de lectura es de interés o puede ser comprendido por un grupo de personas. Este problema se conoce bajo el término de "lecturabilidad", y se refiere a cómo encontrar la congruencia entre los intereses y habilidades del lector, por un lado, y las características estilísticas, el contenido y la complejidad del material por el otro. En la medida en que se logre esa congruencia, en esa medida se logrará obtener satisfacción o información de la lectura.

El término "lecturabilidad" incluye cuatro términos altamente interrelacionados: la **legibilidad**, la **interestabilidad**, la **fluidez lectora** y la **comprensibilidad**.

- La **legibilidad** se refiere a las características físicas del material que condicionan el proceso perceptivo, es decir, la textura y color del papel, el tipo de letras utilizado, el color e intensidad de la tinta.
- La **interestabilidad** se refiere a propósito que mueve al lector a buscar el material, así como persistir en su lectura.
- La **fluidez** se refiere a la velocidad con que se realiza la lectura.
- La **comprensibilidad** se refiere a la complejidad lingüística y estilística del material que puede variar haciéndolo más fácil o más difícil (Bormuth, 1974).

Los conceptos de la lecturabilidad son de interés para muchas personas. Un ejemplo lo constituye el maestro en el salón de clases que debe buscar una selección específica para que sus alumnos practiquen los conceptos de lengua explicados recientemente. Otro ejemplo lo constituyen los autores y editores de libros infantiles, que deben tomar en consideración los intereses y las habilidades lingüísticas y lectoras de su público lector. Para el lingüista, el interés se centra más que todo en los conceptos de la comprensibilidad y la forma en que puede determinarse la dificultad de comprensión de los materiales.

La medición de la comprensibilidad de los materiales de lectura puede realizarse por varios métodos. Uno de ellos es por medio de juicios de estimación realizados por expertos, quienes basados en su experiencia personal pueden ordenar un conjunto de materiales según su dificultad relativa o clasificarlos por categorías. Un segundo método es por medio de pruebas, generalmente de opciones múltiples, que se desarrollan sobre cada uno de los materiales y se les administran a los sujetos después que han leído los materiales. Una tercera posibilidad la constituyen las fórmulas de comprensibilidad, que son especialmente adecuada dado el gran volumen de materiales que actualmente sale al mercado y a la variedad de público al cual van dirigidos.

* Doctor **Nelson Rodríguez Trujillo**: Psicólogo, especialista en lectura. Dirige el Departamento de Investigaciones del Banco del Libro, en Venezuela.

La fórmula de comprensibilidad consiste en una fórmula matemática, generada por métodos de regresión múltiple, que incluye ciertas características lingüísticas del material cuya dificultad se pretende evaluar. Para ser utilizada, simplemente se sustituyen los valores de las variables lingüísticas en la fórmula, se resuelve la ecuación, y se obtiene un resultado, que debidamente interpretado nos permite decir el nivel de dificultad del material. Esta metodología tiene la ventaja de ser mucho más exacta que los juicios de estimación y mucho más económica que las pruebas de opciones múltiples, ya que debe ser calculada sólo una vez, eliminándose así la necesidad de desarrollar nuevas pruebas.

El procedimiento de desarrollo de la fórmula de comprensibilidad es el siguiente:

1. Se selecciona el **conjunto de materiales** tomando en cuenta muy especialmente la representatividad que tiene esa muestra en el universo total en que se empleará posteriormente la fórmula; es decir, las fórmulas son específicas para ciertos materiales, y no tienen necesariamente validez universal para todos los materiales.
2. Se definen las **variables lingüísticas** que se consideran más valiosas para predecir la dificultad.
3. Se define el **criterio** por medio del cual se determinará la dificultad de los materiales. Hasta hace poco se utilizaban las preguntas de opciones múltiples o los juicios de estimación, pero el procedimiento "*cloze*" ha obtenido cada vez más aceptación como criterio de dificultad. Este procedimiento consiste en la eliminación sistemática y mecánica de cada enésima palabra (en español se recomienda la eliminación de cada quinta palabra. Véase Rodríguez Trujillo, 1977) y su sustitución por una línea de longitud estándar. El *test* así producido se administra a una muestra de sujetos con las instrucciones de que reemplacen la palabra eliminada. Se asigna un punto por cada palabra correctamente sustituida. No se aceptan sinónimos porque se ha demostrado que dar crédito por la palabra exacta es un procedimiento más válido y confiable, además de ser más económico (Morles, 1975). El promedio de los puntajes obtenidos por los sujetos constituye un nivel de dificultad de la selección, es decir, el criterio a estimar.
4. Se hace un conteo de las variables lingüísticas para cada trozo, éstas constituyen las variables predictoras que entrarán en la fórmula para predecir el criterio.
5. Los resultados se someten a un programa de regresión múltiple, que escogerá las variables más valiosas para predecir la dificultad, así como los coeficientes correspondientes. Un ejemplo de una fórmula de lecturabilidad es la desarrollada por el Departamento de Investigaciones Educativas (Gutiérrez y colaboradores, 1972) del Ministerio de Educación de Venezuela.

$$C_p = 95,2 - 9,7 (L/P) - 0,35 (P/O)$$

Para precisar la dificultad de un determinado material, con esta fórmula, se cuenta el número de palabras (**P**), el número de letras (**L**) y el número de oraciones (**O**) (definidas como un elemento gráfico contenido entre dos puntos), luego se crean los índices y se aplica la fórmula. El criterio utilizado para esta fórmula fue el *test "cloze"*, aplicado a niños de 6º grado; por lo tanto, el resultado

indicará el porcentaje promedio de respuestas a un *test "cloze"*, basado en ese trozo, que obtendrían sujetos de 6º grado. Naturalmente que este resultado tiene poco sentido si no se conoce el significado de los puntajes *"cloze"*. Es decir, aún falta por saber a partir de qué punto una persona es capaz de asimilar la información contenida en el material, y a partir de qué punto el material es demasiado difícil. Para ello es necesario un sexto paso que es el de **desarrollar tests de habilidad lectora** que se le apliquen a los sujetos y que aporten una correspondencia entre cada uno de los niveles del *"cloze"* y las características del sujeto.

La alta confiabilidad y validez del procedimiento *"cloze"* demostrada en numerosos estudios tanto en inglés como en español (Rankin, 1957; Potter, 1968; Rodríguez Trujillo, 1977) así como la rigurosidad de los procedimientos de muestreo de los sujetos y los materiales, garantizan hoy en día la calidad de las fórmulas de comprensibilidad. Sin embargo, una fórmula es sólo tan buena como las variables lingüísticas que la determinan. Solamente en la medida en que el investigador pueda desenvolver variables directamente relacionadas con el proceso de la lectura, con las características estilísticas del material, será posible desarrollar las fórmulas que puedan predecir con validez y confiabilidad la dificultad de los materiales. Es por eso que la intervención del lingüista en el desarrollo de fórmulas de comprensibilidad es un requisito indispensable.

Hasta hace poco las fórmulas de comprensibilidad incluían solamente variables indirectamente relacionadas con el proceso psicolingüístico de la lectura (Klare, 1963). Dos grandes grupos de variables eran las más utilizadas: las constituidas por un factor de vocabulario y las constituidas por un factor de oración. Las variables de vocabulario más comunes son de dos tipos: la longitud de las palabras, y su frecuencia de aparición en las comunicaciones escritas. El uso de estas dos variables se justifica por el siguiente razonamiento. Las palabras que ocurren más frecuentemente en materiales de lectura tienden a ser más familiares a los lectores y, por tanto, más fáciles de leer. Por ende, basta con determinar en un material, la proporción de palabras frecuentes o infrecuentes (véase Rodríguez Bou, 1952, para una lista de frecuencia de palabras en español) para precisar la dificultad del material. Asimismo, las palabras que ocurren más frecuentemente en cualquier idioma tienden a ser más cortas que las que ocurren infrecuentemente, tal y como lo demostrara Zipf (1935) de acuerdo con la "ley del esfuerzo mínimo" (un ejemplo de ello es el acortamiento de la palabra "borrador" a "borra"; o de "señorita" a "seño" que han hecho los niños venezolanos). El problema con este tipo de variables es que se basan en suposiciones acerca de una relación existente entre la frecuencia de ocurrencia y/o la longitud de las palabras y el proceso subyacente en la lectura; no necesariamente existe una correspondencia unívoca entre esos procesos y las relaciones en que se basan las suposiciones son desconocidas.

Las variables del factor de oración más utilizadas son la longitud de las oraciones y su complejidad estructural. El uso de la primera se basa en el mismo argumento utilizado para las variables del factor de vocabulario, es decir, se asume que las oraciones largas son necesariamente de mayor complejidad y, por lo tanto, más difíciles de comprender que las oraciones cortas. La complejidad estructural ha sido generalmente analizada en términos del número de oraciones complejas o compuestas, que aparecen en un determinado texto. Aquí también es

válida la crítica de que no existe una relación unívoca entre la complejidad estructural y la longitud de las oraciones.

Las fórmulas de comprensibilidad que se basan en variables lingüísticas tan débiles tienen una baja confiabilidad cuando se las somete a validación cruzada con otras muestras de materiales y/o sujetos. La fórmula experimental presentada anteriormente, logró explicar solamente el 68 por ciento de la variación total de la dificultad de la muestra en que se desarrolló. Fórmulas similares desarrolladas en inglés, tienen igualmente baja confiabilidad y validez. Esto es lógico esperarlo porque ninguna de las variables utilizadas en la fórmula puede considerarse como causativa de la dificultad, sino como meros indicadores.

De aquí surge la necesidad de utilizar, por un lado, una metodología que describa mejor las características de los materiales. Bormuth (1966,1969) encontró que la relación entre la frecuencia de ocurrencia de las palabras y su dificultad no es lineal, como se asumía generalmente, sino curvilínea; es necesario emplear por ello, correcciones polinomiales o programas de regresión múltiple que acepten relaciones curvilíneas.

Por otro lado es necesario el desarrollo y utilización de variables lingüísticas más directamente relacionadas con el proceso de la lectura y que puedan interpretarse como causativas de la dificultad de comprensión. Klare (1963) presenta un exhaustivo recuento de la historia de las fórmulas de comprensibilidad en inglés. Es interesante ver cómo en sus intentos por determinar la dificultad de comprensión de los materiales, los investigadores probaron todo tipo de variables predictoras, sin lograr elevar el techo de las correlaciones obtenidas más allá de 70. Bormuth (1966) considera que esto obedecía a dos causas principales: la falta de criterios adecuados y la falta de una teoría lingüística consistente que permitiese la generación de variables predictoras adecuadas. La introducción del procedimiento "*cloze*" en 1953, por W. Taylor, solucionó el primer problema. El desarrollo de la lingüística generativa transformacional y el consecuente desarrollo de la psicolingüística, ha creado el marco teórico para la solución del segundo. En aplicaciones experimentales (Bormuth, 1969), logró niveles sin precedencia de exactitud predictiva ($R = 94$ en validación cruzada) con fórmulas que incluían variables lingüísticas tales como el "Índice de profundidad" de Yngve (1962), o la "Frecuencia de Cláusulas Independientes" basado en los trabajos de Strickland (1962). Bormuth (1966) demostró asimismo que el aumento en la capacidad predictiva de sus fórmulas no se debía solamente a la mayor confiabilidad del *test "cloze"* usado como criterio, sino a las variables lingüísticas utilizadas, que describen más adecuadamente las características estilísticas del material.

En español solamente se conocen además de la fórmula a la que se hizo mención anteriormente, otras dos fórmulas: una debida a Spaulding y publicada en 1956 y otra publicada por Patterson en 1972 (tomado de Klare, 1974). Sin embargo, su aplicación ha sido muy limitada.

La escasa divulgación en países de habla hispana de los conceptos de la lecturabilidad, parece indicar una falta de conciencia acerca de la magnitud del problema que implica la publicación de materiales que no pueden ser comprendidos por los niños, jóvenes y adultos a los cuales van dirigidos. Esa falta

de conciencia contrasta con la queja constante de educadores y padres acerca de la incapacidad de las escuelas primarias y media de enseñar a leer adecuadamente.

Las fórmulas de comprensibilidad obviamente que no constituyen por sí mismas la solución a este problema. Sin embargo, en su desarrollo se hará necesario un análisis cuidadoso de las variables lingüísticas que directa o indirectamente determinan la dificultad de comprensión de los materiales. Ese análisis serviría además para explicitar que debe existir una adecuación entre las habilidades lingüísticas de los materiales, y que es necesario un control de las últimas, hasta tanto los niños no ejerciten sus habilidades adecuadamente. Es posible que gran parte de la incapacidad de la escuela para enseñar a leer, radique en la falta de materiales adecuados, más que en los métodos utilizados por los maestros.

La adaptación del procedimiento "*cloze*" al español (Morles, 1975; Rodríguez Trujillo, 1977) como criterio para la medición de la dificultad de los materiales, y la popularización de la computadora de alta velocidad han puesto al alcance de los países latinoamericanos la posibilidad de desarrollar fórmulas de comprensibilidad relativamente a bajo costo. Sin embargo, hace falta el trabajo del lingüista, que haga estudios dirigidos a ayudar en el desarrollo de las variables adecuadas para esas fórmulas, y que orienten además a maestros y autores de materiales educativos hacia la necesidad de controlar la dificultad de los materiales de lectura. Es posible que ésta sea una de las más grandes contribuciones que pueda hacer la lingüística a la educación y a la comunicación de masas.

Referencias bibliográficas

- Bormuth, J.R. (1969) "Development of readability analyses." H.E.W. Proyecto DEC -3-7-07-52-03236. Chicago Universidad de Chicago, mimeografiado.
- Bormuth, J.R. (1966) "Readability, a new approach." **Reading Research Quarterly**, **3**, 79-132.
- Bormuth, J.R. (1974) "Literacy, Policy and the Emerging Technology of Readability". En **Visible Language**, **8 (2)**, 123-135.
- Gutiérrez de Polini, L.E. y colaboradores. "Investigación sobre lectura en Venezuela." Ponencia presentada ante las **Jornadas de Educación Primaria**. Caracas, Ministerio de Educación, mimeografiado.
- Klare, G.R. (1974-1975) "Assessing Readability." En **Reading Research Quarterly**, **1**, 62-102.
- Klare, G.R. (1963) **The Measurement of Readability**. Iowa State University Press.
- Morles, A. (1975) "The Scoring of '*Cloze*' Comprehension Tests in the Spanish Language." Tesis de Master no publicada. Chicago, University of Chicago.
- Patterson, F. (1972) **Cómo escribir para ser entendido**. El Paso, Texas, Casa Bautista de Publicaciones.
- Potter, T.C. (1968) **A Taxonomy of "*Cloze*" Research, Part I? Readability and Reading Comprehension**. Inglewood, California, Southwest Regional Lab.
- Rankin, E. (1957) "An Evaluation of the '*Cloze*' Procedure as a Technique for Measuring Reading Comprehension." University of Michigan, Disertación de Ph.D. no publicada.
- Rodríguez Bou, I. (1952) **Recuento de vocabulario español**. Río Piedras, Universidad de Puerto Rico.

- Rodríguez Trujillo, N. (1977) "Adaptation of the 'Cloze' Procedure to the Spanish Language." University of Chicago. Disertación de Ph.D. no publicada.
- Spaulding, S.A. (1956) "Spanish Readability Formula." En **Modern Language Journal**, diciembre, **4D**, 433-431.
- Strickland. R. (1962) "The Language of Elementary School Children". En **Bulletin of the School Education**. Indiana University, 34, 4.
- Taylor, W. (1953) "Cloze Procedure: a new Tool for Measuring Readability." En **Journalism Quarterly**, **3D**, 415-433.
- Yngve. V.M. (1962) "Computer Programs for Translation." En **Scientific American**, 206, 69-76.
- Zipf, G. (1935) **The psycho-biology of language**. Boston, Houghton, Mifflin & Co.