

ESTIMACIÓN DEL VOCABULARIO TOTAL CONOCIDO POR UN ESCOLAR

A. Suárez

Universidad de Santiago

P. Meara

University College, Swansea

I. INTRODUCCIÓN

Entre los múltiples estudios cuantitativos referidos al vocabulario, existen aquellos que nos permiten estimar el número total de palabras «conocidas» por una determinada persona. Nos permiten afirmar, por ejemplo, de dos escolares de 10 años de edad, que uno conoce unas 5.000 palabras y el otro, unas 15.000, del total de palabras de un universo léxico prefijado.

El número total de palabras conocidas por una persona es una información mucho más rica que la obtenida mediante los tests habituales de vocabulario —con frecuencia, de elección múltiple—, que sólo nos dicen el lugar que ocupa el sujeto dentro del grupo normativo. La comparación de las diferencias del tamaño total del vocabulario en momentos sucesivos de la vida de una persona nos permite establecer el ritmo de crecimiento del vocabulario, dato del mayor interés, sobre todo para no caer en el error corriente de la subestimación del tamaño total del mismo (Seashore, 1949; Crystal, 1990) y en las indeseables consecuencias que de esta subestimación se derivan (Por ejemplo, la de sobrevalorar la importancia de la enseñanza directa de un vocabulario «tipo», como se hace en Casanova y Rivera, 1989, en lugar de insistir en los mecanismos indirectos de adquisición, que podríamos resumir con el constructo «volumen de experiencias lingüísticas» (Nagy & Anderson, 1984).

Ha habido diversos intentos de contestar a la pregunta de cuántas palabras conoce una determinada persona. Por un lado, tenemos estudios de tipo biográfico (Leopold, 1949; Nelson, 1973; Wagner, 1985, etc.), que registran las palabras que utiliza activamente el sujeto, actualmente mediante el uso de técnicas radio-microfónicas. Por otro lado, tenemos los estudios basados en cálculos inferenciales: se obtiene una muestra de los vocablos de un diccionario por el sistema de catas, se averigua cuántos de esos vocablos conoce el individuo y, finalmente, se hace la estimación inferencial de cuántas palabras conoce de las del diccionario de partida (Smith, 1926; Seashore & Eckerson, 1940; García Hoz, 1946; Diack, 1975, entre otros). Ambos tipos de estudios tienen limitaciones obvias.

II. NUESTRO TRABAJO

1. **Naturaleza de los tests.** Nuestra investigación se inscribe entre los trabajos que tratan de estimar el número total de vocablos conocidos por una persona a través de cálculos inferenciales. La aportación más novedosa de nuestro intento quizá consista en la naturaleza de los tests utilizados.

Cada test está formado por 40 palabras y 20 pseudopalabras. Como ilustración, presentamos los seis primeros estímulos de uno de los tests:

- | | | |
|-------------------|-------------------|-------------------------|
| 1 () narigudo | 2 () remaquir | 3 () encarecimiento |
| 4 () agenio | 5 () roscalde | 6 () regente |

Las 40 palabras pertenecen a una muestra aleatoria constituida por el 10% del total de unas 22.000 entradas que componen el Diccionario Escolar Etimológico (García Hoz, 1989, 17ª edición, actualizada y corregida). En esas 40 palabras están representadas proporcionalmente las palabras de cuatro estratos diferentes, correspondientes al «grado de dificultad de aprendizaje que presenta cada palabra», según el juicio emitido por tres profesores de E.G.B. Las pseudopalabras tienen características fonotácticas y grafotácticas similares a las palabras de la muestra y han sido generadas a partir de estas.

Al sujeto se le pide únicamente que escriba S (=sí) dentro del paréntesis si sabe lo que significa la palabra o que escriba N (=no) si no sabe lo que significa o no está seguro. Es una tarea motivadora para los sujetos, que realizan en unos tres minutos. Esta última característica permite aplicar varios de estos tests en el mismo tiempo que se administra un test convencional de vocabulario, lo que hace posible realizar las inferencias proporcionales a partir de más datos.

Las 20 pseudopalabras constituyen una especie de escala de honestidad, que, teniendo en cuenta la teoría de detección de señales, nos permite hacer una corrección del número de palabras realmente conocidas cuando el sujeto rodea como conocidas algunas pseudopalabras.

No podemos entrar aquí a fondo en el tema de **qué significa conocer una palabra** (AILA Review dedica un monográfico en 1989 al tema). Sin duda, nuestros tests sirven para averiguar, cuando menos, si el sujeto sabe distinguir lo que es y no es una palabra castellana, y es difícil imaginar un aspecto del conocimiento de una palabra más fundamental que este.

2. Hipótesis y diseño. Tratamos de someter a comprobación la hipótesis de que administrando a un sujeto un corto número de estos tests —independientemente de cuáles sean— siempre obtenemos la misma estimación de su vocabulario total. Si estimamos el tamaño del vocabulario a partir de un solo test, probablemente obtendremos una variación considerable entre las estimaciones de un test a otro; pero si lo estimamos a partir de la media obtenida en varios tests, reduciremos esta variación. La cuestión que tratamos de resolver es averiguar cuántos tests tenemos que administrar a un sujeto para obtener una estimación estable del tamaño de su vocabulario.

El diseño es de un factor con medidas repetidas y la técnica utilizada en el análisis de los datos, ANOVA. Podremos seguir manteniendo nuestra hipótesis en la medida que no hallemos diferencias significativas en las estimaciones medias obtenidas a partir de un cierto número de tests. Este resultado tiene que ir acompañado de altas correlaciones entre las medias de los grupos de tests en cuestión, para asegurarnos de que las estimaciones tampoco varían a nivel individual.

3. Administración de las pruebas. Se realizaron dos administraciones, ambas a alumnos de 6º de E.G.B., la primera en un Colegio Público de Alcobendas (Madrid) y la segunda en un Colegio subvencionado de Santiago de Compostela. En Alcobendas se administraron 5 de nuestros tests durante una sesión de la clase de Lengua y otros 5 una semana después. A todos los alumnos se le administraron los mismos tests y en el mismo orden. En Santiago, se administraron 4 tests en cada sesión y el orden de los tests para cada sujeto fue determinado al azar. En ambos casos entre las dos administraciones se aplicó a los escolares el subtest de vocabulario del test T.E.A.-1 (versión 1990).

4. Resultados. En la administración a los escolares de Alcobendas, la única diferencia no significativa obtenida —de las varias combinaciones comprobadas— se obtuvo cuando se compararon las estimaciones de los cinco tests administrados el primer día con las de los otros cinco tests aplicados una semana después: $F(1,105) = 1.61$. La correlación entre estos dos grupos de tests —«antes» y

«después»— fue $r = .7794$. Todas las correlaciones con el criterio (el subtest del T.E.A.-1 citado) resultaron positivas y significativas, incluso las de los tests considerados separadamente, que oscilaron entre $r = .25$ y $r = .51$, $p < .01$.

Al analizar los datos de la administración de Santiago, tampoco se obtuvieron resultados que siguieran una pauta clara: en algunos casos, la media en tres o cuatro tests resultó suficiente para obtener estimaciones del vocabulario total estables; en otras, no. Las correlaciones entre las medias de dos tests distintos van desde $r = .45$ a $r = .74$, $p < .001$. Las correlaciones entre cada uno de los tests y el criterio están entre $r = .20$ y $r = .43$, $p < .05$. La estimación de la media del vocabulario total para el grupo de alumnos de Santiago resultó ser $X = 10.306$, con una $s = 1.971$. Por supuesto, este vocabulario total ha de considerarse como un mínimo, ya que se trata de una estimación hecha a partir de un diccionario de sólo unas 22.000 entradas (Recordemos que el diccionario de la R.A.E. tiene cerca de 100.000 entradas).

III. CONCLUSIONES Y PROSPECTIVA

Los resultados obtenidos no constituyen una base empírica suficiente para poder sostener que nuestros tests sean instrumentos apropiados para estimar el vocabulario total conocido por un escolar. Sin embargo, existen varias razones que nos llevan a pensar que no debemos abandonar el intento: algunos de los resultados son esperanzadores, una simple inspección de los protocolos revela pautas claras en la mayoría de los sujetos, el intento en su conjunto parece obedecer a un planteamiento lógico, el propósito que perseguimos es decididamente relevante...

Nuestro programa inmediato de trabajo, guiado por la hipótesis explicitada anteriormente, tiene estos objetivos: mejorar las condiciones en que se realizan las pruebas y las instrucciones dadas a los sujetos (tratamos de disminuir la tasa de pseudopalabras rodeadas como conocidas); conseguir un modelo más satisfactorio para efectuar las correcciones de acuerdo al número de pseudopalabras rodeadas; administrar las pruebas a otros grupos de edad, etc.

BIBLIOGRAFÍA

- CASANOVA, M^a A., RIVERA, M. (1989): Vocabulario básico en la E.G.B., I-III. Madrid: M.E.C./ Espasa Calpe.
- CRYSTAL, D. (1989): Teaching Vocabulary: The Case for a Semantic Curriculum. En K. MOGFORD-BEVAN & J. SADLER: *Child Language Disability, V. II*. Clevedon, Avon: Multilingual Matters Ltd.
- DIACK, H. (1975): Standard Literacy Tests. St. Albans, Herts.: Hart— Davis Educational Ltd.
- GARCÍA HOZ, V. (1946): Evolución cuantitativa del vocabulario en escolares de nueve a dieciocho años. *Revista Española de Pedagogía*, 16, oct-dic., t. IV.
- JUSTICIA, J. F. (1984): Determinación y Análisis del Vocabulario del Niño de 6 a 10 años. Madrid: Memoria del Proyecto de Investigación 4.107/79, CAICYT.
- NAGY, W. E. & ANDERSON, R. C. (1984): How many words are there in printed school English? *Reading Research Quarterly*, v. XIX, 3.
- MEARA, P. & BUXTON, B. (1987): An alternative to multiple choice vocabulary tests. *Language testing*, 4, 2, 142-154.
- SEASHORE, R. H. & ECKERSON, L. D. (1940): The measurement of individual differences in general English vocabularies. *J. of E. Psychology*, 31, 14-38.