

# Caracterización de complejidad semántica en pasajes tipo TOEFL

## Characterization of semantic complexity in TOEFL passages

Meliza Contreras-González<sup>1</sup> , Mireya Tovar Vidal<sup>1</sup> , Guillermo De Ita Luna<sup>1</sup> 

<sup>1</sup>Benemérita Universidad Autónoma de Puebla, Puebla, México.  
[meliza.contreras@correo.buap.mx](mailto:meliza.contreras@correo.buap.mx), [mtovar@cs.buap.mx](mailto:mtovar@cs.buap.mx), [deita@cs.buap.mx](mailto:deita@cs.buap.mx)

(Recibido: 28 julio 2021; aceptado: 18 septiembre 2021; Publicado en Internet: 1 diciembre 2021)

**Resumen.** La comprensión lectora es un proceso que implica la realización de modelos mentales, así como el almacenamiento de recuerdos en la memoria a largo plazo, si a esto le agregamos que muchos estudiantes requieren acreditar otros idiomas, implica realizar más procesos cognitivos. Con los modelos mentales se pueden analizar las estructuras internas presentes, por ejemplo, en la sección de comprensión lectora del examen TOEFL donde se involucran tres tipos de textos: narrativos, expositivos, y argumentativos. En este trabajo se modelan semánticamente los tres tipos de textos como estructuras, donde las relaciones semánticas son el insumo para la construcción de estas, así como la complejidad semántica dependiendo del tipo de texto a revisar. Al contar con dichas estructuras los estudiantes pueden ubicar gráficamente los elementos que caracterizan a cada tipo de texto, así como se da una orientación para generar sistemas expertos para el entrenamiento de las actividades a desarrollar.

**Palabras clave:** Comprensión lectora, Estructura semántica, Relaciones semánticas.

**Abstract.** Reading comprehension is a process that involves the realization of mental models, as well as the storage of memories in long-term memory; if we add to this that many students need to accredit other languages, it involves performing more cognitive processes. With mental models, it is possible to analyze the internal structures present, for example, in the reading comprehension section of the TOEFL test where three types of texts are involved: narrative, expository and argumentative. In this work, the three types of texts are semantically modeled as structures, where semantic relations are the input for their construction, and the semantic complexity depending on text type to be reviewed. By having these structures, students can graphically locate the elements that characterize each type of text, and an orientation is given to generate expert systems for the training of the activities to be developed.

**Keywords:** Reading comprehension, Semantic structure, Semantic relationships.

**Tipo de artículo:** Artículo de investigación.

## 1 Introducción

La lectura es un proceso importante para la adquisición de conocimiento, por lo que desde el enfoque computacional se ha planteado el uso de herramientas como audiolibros, resumidores e indexadores para apoyar el proceso de comprensión lectora. También desde la perspectiva de la neurociencia, se estudian los procesos internos de la lectura que dan origen a las respuestas adecuadas ante cuestionamientos diarios. Por lo que el buen lector es capaz de construir una representación mental integrada del texto, que además sea coherente y certero, como menciona Kintsch & van Dijk (1978).

Los lectores de textos en otros idiomas, distintos al nativo, tienen dos retos: por un lado; hacer el proceso de traducción a su lengua nativa y posteriormente mapear la estructura del vocabulario de la lengua extranjera a su lengua materna.

En los exámenes de acreditación en el idioma inglés, se evalúan habilidades de gramática, comprensión auditiva y lectura de comprensión. Así en esta última sección, el lector construye un modelo de representación del conocimiento para responder a las preguntas que se presentan para evaluar su nivel de comprensión, así como la aplicación de procesos inferenciales para comprender el significado del texto.

El presente trabajo de investigación plantea la generación de estructuras semánticas a partir de los pasajes disponibles en las secciones de comprensión lectora del examen TOEFL, a fin de establecer los mecanismos

de inferencia que permitan identificar la respuesta correcta de los cuestionamientos propios de este examen. Tomando en cuenta los modelos cognitivos existentes, resultará necesario formalizar la representación de conocimiento mediante relaciones semánticas e inferenciales propias de los tipos de documentos asociados a la comprensión lectora. El documento está estructurado de la siguiente forma: en la sección 2 se muestra el desarrollo del trabajo considerando: los conceptos preliminares de comprensión lectora y relaciones semánticas, la generación de las estructuras semánticas de acuerdo a los tipos de pasajes que son narrativo, expositivo o argumentativo y finalmente en la sección 3 se mencionan las conclusiones.

## 2 Desarrollo

En el caso de la lectura de textos tipo TOEFL, la interpretación se alcanza cuando además de verificarse la coherencia, se reconoce el contexto en el que está relacionado el texto. En la mayoría de los casos, a pesar de que la escritura es lineal, para acreditar la fase de comprensión lectora se requieren aplicar estrategias para realizar un recorrido no-lineal del texto; dado que involucra reorganizar el contenido, regresar a párrafos anteriores formando una estructura multidimensional cuyos elementos se relacionan entre sí, trascendiendo a su vecindad para descifrar las pistas y así responder a las preguntas en cuestión (Viramonte de Avalos, 2004). Sin embargo, los lectores se enfrentan al uso de mecanismos de inferencia cuando las preguntas distan de ser meramente conceptuales o factuales, e implican verificar su capacidad para inferir las relaciones causa-efecto, o incluso, se requiere hacer uso de conocimientos previos para deducir los resultados y generar conocimiento implícito.

Por otro lado, el conocimiento es un constructo que puede ser semántico o procedimental. El conocimiento semántico se define como la información acerca de hechos, conceptos, principios, reglas y planteamientos conceptuales y teóricos, que conforman una disciplina o un campo de estudio. El conocimiento procedimental es el resultado de la operacionalización de los procesos y se define como el conjunto ordenado de pasos o acciones que acompañan a un acto mental o una actividad motora. Este conocimiento sirve para generar cambios y/o transformaciones del conocimiento o de los estímulos del medio ambiente. Los procedimientos son los instrumentos o componentes dinámicos del conocimiento (Amestoy de Sánchez, 2002). Por lo que ambos conocimientos están implícitos en los procesos de comprensión lectora de los que se hablará en la siguiente subsección.

### 2.1 Comprensión lectora

Los textos permiten representar la realidad y para eso se requiere que el conjunto de ideas que se expresan estén vinculadas entre sí de manera coherente. Esta propiedad está presente, cuando las ideas de varios enunciados o párrafos están unidas de una forma lógica y consistente (Viramonte de Avalos, 2004).

La coherencia dentro de un texto permite a los lectores identificar su organización en una forma conocida. Por ejemplo, que cada párrafo tenga su propio enfoque o idea y que la idea de un párrafo de alguna manera fluya a o se relacione con la idea del próximo (Viramonte de Avalos, 2004).

Para Snow (2002) la comprensión lectora es el proceso simultáneo de extraer y construir significado y cuyo objetivo es:

1. Descifrar como las letras representan las palabras.
2. Abordar la traducción en forma precisa y eficiente de las letras a sonidos (extraer significado del texto).
3. Formular una representación de la información que está siendo presentada, que inevitablemente requiere la elaboración de nuevos significados y la integración de la información nueva con la antigua (construcción del significado).

De acuerdo con la lectura crítica dentro del método de lectura tenemos tres niveles. El primer nivel es literal y se le invita a extraer la información dada en el texto sin agregarle ningún valor interpretativo. Las habilidades que conducen este nivel de lectura son la observación, la comparación la clasificación, el análisis, la síntesis, y la evaluación.

El segundo nivel es el inferencial, en el que se establecen relaciones que van más allá del contenido literal del texto, es decir, que se hacen inferencias acerca de lo leído. Dichas inferencias pueden ser

inductivas o deductivas. Este nivel de lectura además de los procesos mencionados, en el primer nivel requiere hacer uso de la decodificación con la inferencia, el inductivo, y el deductivo con el discernimiento y la identificación e interpretación de las temáticas de un tema.

El tercer nivel de lectura es el analógico que permite trasladar las relaciones extraídas de la lectura de un ámbito a otro. En este nivel, además de los procesos requeridos en los niveles anteriores se precisa interpretar las temáticas del escrito para establecer relaciones analógicas de diferentes índoles y emitir juicios de valor acerca de lo leído. La capacidad de inferir consiste en utilizar la información disponible para aplicarla o procesarla con miras a emplearla de una manera nueva o diferente, por ejemplo, basándose en determinados datos, se puede suponer cómo sucedieron ciertos hechos, considerando la información de alcance y se transforma utilizándola de una manera nueva. Requiere la intervención de las habilidades de pensamiento más sofisticadas.

Para resolver inferencia los estudiantes requieren desarrollar la habilidad de análisis que implica detectar en un texto narrativo: ¿qué es lo que pasa?, ¿cuál fue el orden en que se presentaron los acontecimientos?, ¿quién está involucrado?, ¿dónde ocurrió el problema? (Amestoy de Sánchez, 2002).

Por otro lado, los diagramas son herramientas de análisis que permiten visualizar las relaciones entre los diferentes personajes, conceptos e ideas que se manejan en un texto. Ayudan a lograr imágenes o representaciones mentales acerca de lo leído y facilitan la comprensión del tema (Amestoy de Sánchez, 2002).

En el proceso de comprensión, las preguntas son activadoras del pensamiento, estimulan la generación de ideas y facilitan la profundización de la lectura. Las preguntas pueden estimular la observación y el análisis. Tanto las preguntas, como los diagramas, se apoyan en los procesos del pensamiento. Dichos procesos, reciben el nombre de estrategias cognitivas y sirven para adquirir conocimiento. Amestoy de Sánchez (2002) indica una estrategia general para analizar la información de un texto:

1. Realiza una lectura general.
2. Lee nuevamente el texto y fórmula preguntas que te ayuden a comprenderlo.
3. Elabora diagramas o patrones de organización que te permitan visualizar las relaciones que se presentan en el texto.
4. Elabora preguntas para extraer más información.
5. Elabora una síntesis para revisar el proceso.

En el caso del segundo punto propuesto por Amestoy de Sánchez (2002), existen cuatro tipos de preguntas:

Preguntas que se responden a partir de información textual que corresponderían a preguntas factuales. Los procesos para responderlas implican la observación directa y el análisis de la información dada en el contenido del texto.

En el caso de las preguntas que requieren suponer algún dato o completar información del texto, por lo que el proceso vinculado sería la inferencia a partir de la información dada en el texto para responderlas.

En el caso de la recomendación de realizar diagramas o patrones, resulta útil considerando que un patrón de organización es una estructura que establece las relaciones entre los diferentes conceptos o características que se plantean en un texto, puesto que facilitan la visualización de los nexos que se plantean entre los diferentes conceptos. El patrón de organización más sencillo es el que se deriva de la observación que se puede representar mediante un mapa o diagrama de características.

A partir de estas directivas, se puede observar que los estudiantes pueden aplicar gran variedad de patrones cognitivos para desarrollar habilidades en los procesos de comprensión lectora.

## 2.2 Relaciones semánticas

La semántica es la parte de la lingüística que estudia el significado de las palabras, oraciones y expresiones del lenguaje. Todas las palabras que mantienen entre sí una relación de significado forman parte de un mismo campo semántico. Por ejemplo: clavel y rosa pertenecen al campo semántico de las flores (Campoy Garrido, 2010). Entre las palabras que forman un campo semántico pueden existir relaciones de hiponimia e hiperonimia, sinonimia y antonimia.

Así una palabra es hipónima de otra si su significado está englobado en ella. Por ejemplo, rosa es una flor. En sentido inverso se tiene las palabras hiperónimas cuyo significado engloba el significado de otras. Por ejemplo, flor es hiperónimo de rosa (Campoy Garrido, 2010). La sinonimia es un fenómeno semántico

por el cual un mismo concepto o idea puede ser expresado con dos o más palabras distintas. Las palabras sinónimas poseen, por lo tanto, un significado igual o muy parecido dentro de un mismo contexto. En el caso de la antonimia es un fenómeno semántico que se produce cuando dos palabras poseen un significado opuesto o contrario (Campoy Garrido, 2010).

Estas relaciones semánticas se encuentran presentes en los textos y en la mayoría de los casos, su identificación apoya en las estrategias inferenciales para responder las preguntas de comprensión lectora. Sin embargo, para mapear el significado de los textos a una base de conocimiento se requiere modelar estas relaciones desde dos aspectos que dependen del contexto (Campoy Garrido, 2010). A continuación, se menciona la metodología para el diseño de las estructuras semánticas.

### 2.3 Metodología

En el examen de acreditación de TOEFL se requiere demostrar tres habilidades en la lectura, MacMillan (2006):

- Identificación de hechos específicos que son mencionados en el pasaje o incluso son omitidos.
- Uso de correferencia en partes específicas del texto.
- Aplicación de inferencias de información implícita en el pasaje

Por cada prueba se evalúan alrededor de cinco a seis pasajes de una extensión de 400-500 palabras y por cada pasaje se evalúan de cinco a seis preguntas resultando al menos 50 preguntas por prueba. Los textos de los pasajes son académicos y son de tres tipos:

- *Expositivo*: el pasaje ofrece explicaciones sobre un tema.
- *Argumentativo*: el pasaje presenta un punto de vista sobre un tema y evidencias que apoyan ese punto de vista.
- *Histórico-narrativo*: el pasaje es un recuento de eventos pasados o sobre la vida de una persona.

De acuerdo con Amestoy de Sánchez (2002), los eventos son más fáciles de identificar para los seres humanos, por lo que a partir de esta observación y de los resultados del etiquetado consideramos modelar en estructuras semánticas los textos narrativos, posteriormente textos explicativos, y finalmente los textos argumentativos. Esto implica que la complejidad semántica va a depender del tipo de relaciones presentes en los textos que se vincula al grado de entendimiento de los lectores, así como en la implementación de un sistema experto. Por lo que se revisaron pasajes de los distintos tipos de textos, como los disponibles en Exam English Ltd., (2021), para revisar a partir de las relaciones semánticas encontradas las estructuras que representan a cada tipo de texto.

### 2.4 Estructuras semánticas para textos narrativos

En la [Figura 1](#) se muestra la estructura propuesta para representar un texto narrativo. Ésta se generó a partir del análisis del corpus revisado para el examen de acreditación del TOEFL. Este tipo de documentos se caracterizan por contar en su desarrollo con una secuencia de eventos en el tiempo, personajes con sus acciones más representativas para el tópico propuesto.

Sin embargo, si sólo se cuenta con la lista de aserciones como representación de conocimiento resulta insuficiente, considerando que para la generación de patrones cognitivos se requiere asociar significado entre las acciones, por lo que es requerido representar de forma gráfica las relaciones entre personajes o en la secuencia de eventos para así seguir un orden que permita a los estudiantes responder acertadamente de acuerdo con el periodo de tiempo. Esta estructura se propone considerando que la línea de tiempo permite ordenar una secuencia de eventos o de hitos sobre un tema, de tal forma que se visualice con claridad la relación temporal entre ellos.

Para elaborar una Línea de Tiempo, como se observa en la [Figura 1](#), se deben identificar los eventos y las fechas (iniciales y finales) en que estos ocurrieron; ubicar los eventos en orden cronológico; seleccionar los hitos más relevantes del tema estudiado para establecer los intervalos de tiempo más adecuados; agrupar los eventos similares; determinar la escala de visualización que se va a usar y, por último, organizar los eventos en forma de diagrama.

Las Líneas de Tiempo son valiosas para organizar información en la que sean relevante los períodos de tiempo en los que se suceden acontecimientos o se realizan procedimientos. En el caso de los textos narrativos, las relaciones semánticas corresponden a relaciones de caso y de espacio-tiempo y causa propósito donde el tiempo genera una secuencia de hitos importantes. Por lo que usar una línea de tiempo es una estructura semántica útil para la comprensión lectora considerando que los eventos se procesan como aserciones, lo que apoya al proceso inferencial y al ser una representación gráfica favorece la asociación de significado entre los conceptos.

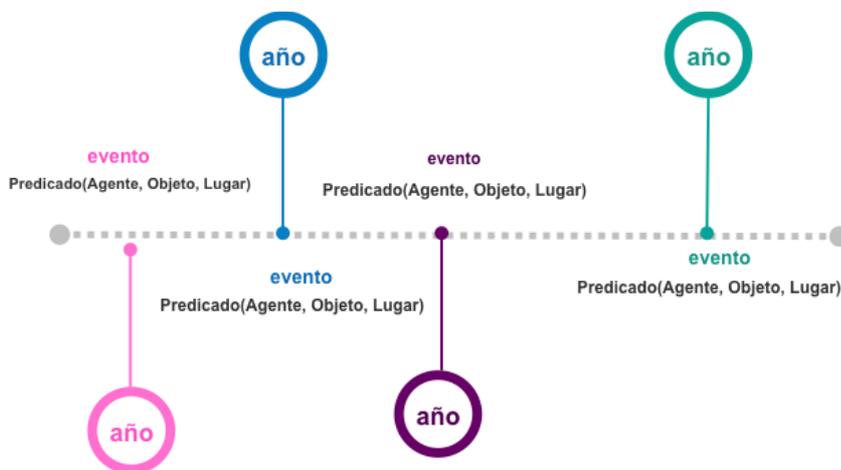


Figura 1. Estructura semántica línea de tiempo

En la Figura 2 se muestra la representación de significado con la estructura semántica de diagrama de estados de un agente. En este caso, un diagrama de estado se considera un diagrama de comportamiento usado para especificar el comportamiento de una parte del sistema diseñado a través de transiciones de estados finitos. En la propuesta se define como un diagrama de estado como la secuencia de aserciones que funcionan como acciones de los agentes durante la narración y se ordenan de acuerdo a su ocurrencia en el texto, como por ejemplo, el pasaje donde se menciona el nacimiento de un género musical o la historia de independencia de algún país.

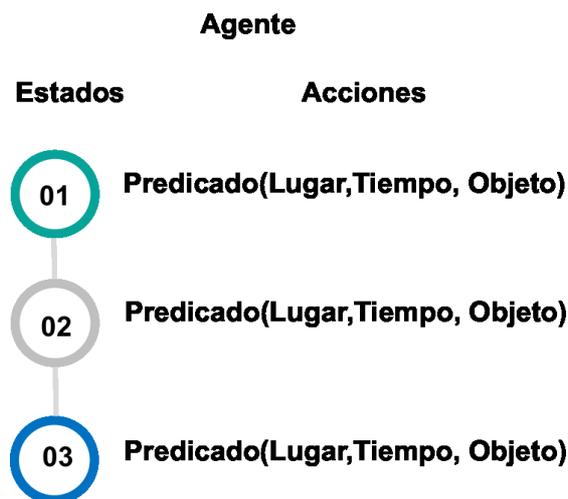


Figura 2. Estructura semántica línea de acción del agente

Esta estructura resulta útil en el proceso de comprensión lectora debido a que los textos narrativos emplean personajes y sus acciones para indicar momentos históricos importantes. Al contar con esta representación se puede generar una ruta de inferencia que incluso requiera deducción de acciones Agente

Paciente entre los personajes, lo que facilita interpretar las preguntas y deducir las respuestas con la información presente.

### 2.5 Estructuras semánticas para textos expositivos

En el caso de los textos explicativos aparecen dos estructuras en el corpus de forma frecuente entre los párrafos: las relaciones semánticas de hiponimia, hiperonimia, y meronimia. Considerando que se requiere una estructura gráfica para asimilar mejor el conocimiento se plantea que en este tipo de textos el lector requiere construir dos tipos de representaciones para abstraer la información relevante del texto. En el caso de la meronimia se determina un concepto central que se describe en el texto con base a los elementos que lo componen, por lo que la representación propuesta es una estructura donde a partir del todo se conectan los elementos que lo conforman como se observa en la [Figura 3](#), por ejemplo, cuando se mencionan temas de biología donde se considera un organismo y sus partes.

En el caso de la hiponimia e hiperonimia la estructura propuesta corresponde a una relación jerárquica tipo árbol donde el nodo padre es el concepto más abstracto y los nodos hijos son los conceptos más específicos, como aparece en la [Figura 4](#), como por ejemplo cuando se define un organismo y los tipos que se derivan de él.

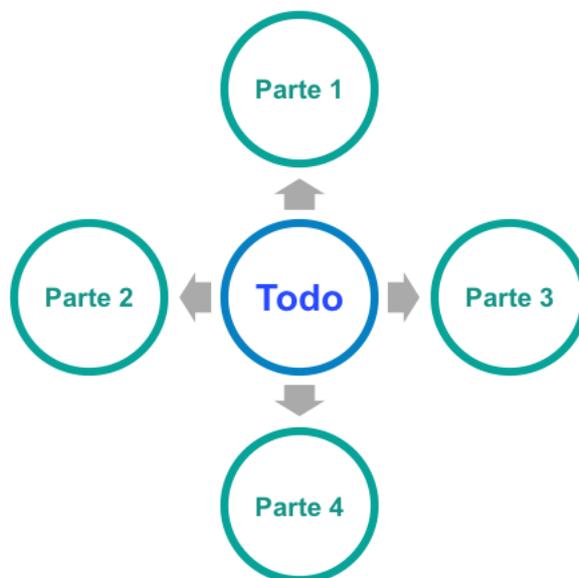


Figura 3. Merónimos

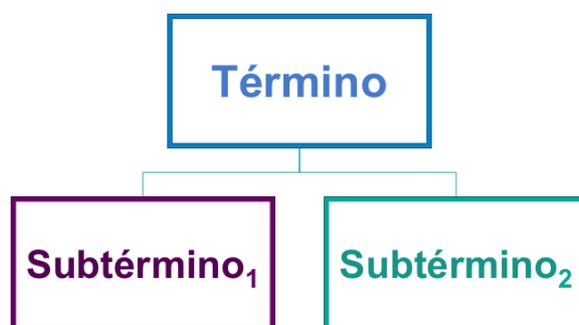


Figura 4. Hipónimos

## 2.6 Estructuras semánticas para textos argumentativos

En el caso de los textos argumentativos se propone una estructura de representación, como se muestra en la Figura 5 considerando que cuando se realiza una evaluación sobre una opinión el lector requiere analizar los hechos, buscar a partir de los hechos las evidencias sobre opiniones para así evaluar el valor de verdad de los juicios realizados por el autor, por lo que estos elementos son retomados en la estructura como conceptos que representan un conjunto por separado y dependiendo de su relación se establecen conexiones entre ellos. Cabe mencionar que este tipo de texto son más complejos para realizar el proceso de comprensión lectora por la interpretación que requiere realizar el autor, por ejemplo, si se cuestionan temas políticos donde se vinculen posiciones a favor o en contra del impacto de un movimiento socialista o capitalista.

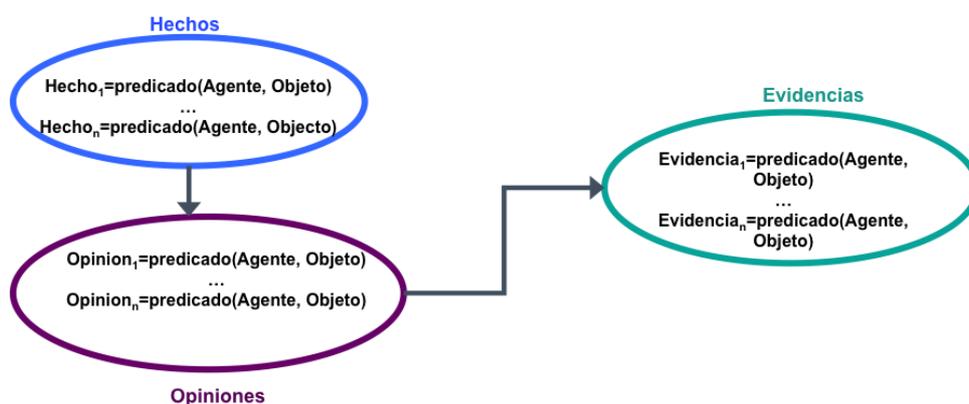


Figura 5. Grafo

## 3 Conclusiones

Teniendo en cuenta las habilidades que requieren los lectores para realizar los procesos de comprensión lectora, es fundamental contar con herramientas que los apoyen en la identificación de estrategias para resolver las cuestiones que dificultan la aprobación de sus evaluaciones.

A partir de las relaciones semánticas, los lectores pueden construir sus aseveraciones y representarlas en un orden secuencial con respecto a las estructuras semánticas: línea de tiempo, línea de acción del agente, merónimos, hipónimos, y grafos que favorecen la asociación de significado y les permite generar posibles rutas para responder correctamente el tipo de preguntas que aparecen en un examen de inglés como el TOEFL, así como les permite discriminar la complejidad semántica de cada pasaje de lectura por lo que facilita la creación de estrategias más precisas para abordar los textos de acuerdo al tiempo y dificultad.

Se propone como trabajo futuro el desarrollo de una herramienta web que permita a los lectores registrar estas estructuras semánticas y, a continuación, utilizar las aseveraciones creadas por un experto para generar rutas de inferencia que respondan correctamente a las preguntas, para posteriormente construir la arquitectura de un sistema experto que apoye el entrenamiento de los lectores sobre estos cuestionamientos en los exámenes de acreditación.

## Declaración de conflicto de intereses

Los autores declaran no tener conflicto de intereses con respecto a la investigación, autoría o publicación de este artículo.

## Financiación

Los autores recibieron apoyo financiero del Consejo Nacional de Ciencia y Tecnología (CONACYT) bajo el proyecto número 257357 de Ciencia Básica, registro 25547.

## ORCID iD

Meliza Contreras-González  <https://orcid.org/0000-0003-3375-4493>

Mireya Tovar-Vidal  <https://orcid.org/0000-0002-9086-7446>

Guillermo De Ita Luna  <https://orcid.org/0000-0001-7948-8253>

## Referencias

- Amestoy de Sánchez, M. (2002). La investigación sobre el desarrollo y la enseñanza de las habilidades de pensamiento. *Revista Electrónica de Investigación Educativa*, 4(1), 1–32.
- Campoy Garrido, N. (2010). Relaciones semánticas entre las palabras: hiponimia, sinonimia, polisemia, homonimia y antonimia. Los cambios de sentido. *Contribuciones a las Ciencias Sociales*, 1–6.
- Exam English Ltd. (2021). *TOEFL® reading test 4: Free practice exercises from Exam English*.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5). <https://doi.org/10.1037/0033-295X.85.5.363>
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Rand Corporation.
- Viramonte de Avalos, M. (2004). *Comprensión lectora*. Colihue.