



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

---

---

FACULTAD DE CIENCIAS DE LA COMPUTACIÓN

Modelo semántico para la comprensión lectora

Tesis que para obtener el grado de  
Doctora en Ingeniería del Lenguaje y del Conocimiento

PRESENTA:  
M.C. Meliza Contreras González

DIRECTORA: Dra. Mireya Tovar Vidal

CO-DIRECTOR: Dr. Guillermo De Ita Luna



Junio, 2021

# Dedicatoria

Agradezco a todos los que participaron en alcanzar este sueño, extrañaré por siempre a los que se adelantaron a un espacio mejor y apoyaré a los que requieran construir los suyos.



# Agradecimientos

Gracias a CONACYT y a la Dra. Mireya Tovar Vidal por su apoyo mediante el proyecto CB 2015-257357-H.

Gracias por el apoyo al Dr. Aurelio López López durante mi estancia en el INAOE en 2018.

Gracias por el apoyo a la Dra. Rosa María Valdovinos Rosas durante mi estancia en la UAEM Toluca en 2019.

Gracias por el apoyo y las retroalimentaciones de mi comité tutorial:

Dra. Mireya Tovar Vidal Asesora

Dr. Guillermo De Ita Luna Coasesor

Dra. Claudia Zepeda Cortés

Dr. Aurelio López Lopez

Dr. José Alejandro Reyes Ortiz



# Resumen

La comprensión lectora requiere estrategias precisas para apoyar al lector sobre el entendimiento de los tipos de texto que existen. La lectura en un idioma ajeno al nativo implica un mayor esfuerzo y requiere que el lector genere una representación mental, que le permita inferir sin tener que hacer la traducción específica de su idioma nativo al idioma extranjero.

En este trabajo se propuso un modelo semántico para discriminar las respuestas correctas de acuerdo a la estructura que poseen los exámenes de acreditación tipo TOEFL.

En la primera etapa se realizó el análisis de un corpus de 300 pasajes para identificar los tipos de textos, sus características, relaciones semánticas y temáticas.

En la segunda etapa se procedió a etiquetar en un formato xsd-xml los textos, para estructurar los elementos que caracterizaban los pasajes, posteriormente se realizó una representación mediante la lógica de predicados y el cálculo de situaciones.

En la tercera etapa se identificaron las estructuras semánticas del cuerpo del pasaje y los bloques cognitivos para determinar los patrones que permitan evaluar las preguntas para determinar las respuestas correctas.

En la cuarta etapa se plantearon reglas de inferencia para determinar la respuesta correcta y se generaron casos de estudio de cada pasaje para verificar el funcionamiento de las mismas.



# Índice de figuras

2.1. Ejemplo de especificación de reglas Castillo (1997) . . . . .	22
4.1. Ejemplo texto explicativo, fuente Universidad de Chuvanan Chuvanan- University (2018) . . . . .	34
4.2. Estructura Semántica Línea de Tiempo . . . . .	38
4.3. Estructura Semántica Línea de Acción del Agente . . . . .	38
4.4. Merónimos . . . . .	40
4.5. Co-Hipónimos . . . . .	40
4.6. Grafo . . . . .	41
4.7. Bloques cognitivo de textos narrativos . . . . .	45
4.8. Bloques cognitivo de textos expositivo . . . . .	46
4.9. Bloques cognitivo de textos narrativos . . . . .	46
4.10. Arquitectura General del Modelo Semántico . . . . .	49
4.11. Ejemplo de especificación de reglas Castillo (1997). . . . .	51
4.12. Ejemplo de especificación de reglas de bloques Cognitivos . . . . .	52
4.13. Ejemplo de especificación de reglas de respuesta . . . . .	53
5.1. Representación XSD de los pasajes. Primera parte. . . . .	58
5.2. Representación XSD de los pasajes. Segunda parte. . . . .	59
5.3. Ejemplo de uso de la representación en xml de un texto expositivo sobre las bacterias. Primera parte. . . . .	60
5.4. Representación en xml del texto expositivo de las bacterias. Segunda parte	60
5.5. Tercera parte de la representación del texto expositivo incluyendo el bloque de pregunta-respuesta . . . . .	61
5.6. Cuarta parte de la representación XML del texto expositivo relacionado con el bloque de pregunta-respuesta . . . . .	61
5.7. Quinta y última parte de la representación xml del texto expositivo . .	62

5.8. Ejemplo de texto expositivo, fuente Universidad de Chuvanan Chuvanan-University (2018) . . . . .	64
5.9. Aplicación de la regla de bloques cognitivos para textos expositivos . .	65
5.10. Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta incorrecta . . . . .	65
5.11. Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta correcta . . . . .	66
5.12. Ejemplo texto narrativo, fuente Universidad de Chuvanan Chuvanan-University (2018) . . . . .	69
5.13. Aplicación de la regla de bloques cognitivos para textos narrativos . . .	70
5.14. Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta correcta . . . . .	71
5.15. Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta incorrecta . . . . .	72
5.16. Ejemplo texto argumentativo, fuente Universidad de Chuvanan Chuvanan-University (2018) . . . . .	74
5.17. Aplicación de la regla de bloques cognitivos para textos argumentativos	75
5.18. Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta correcta . . . . .	76
5.19. Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta incorrecta . . . . .	76

# Índice de tablas

2.1. Tipos de Roles semánticos . . . . .	12
2.2. Clasificación de Inferencias . . . . .	18
4.1. Representación de las relaciones semánticas del tópico de los virus. . . . .	35
4.2. Representación de las relaciones semánticas del tópico de los virus. . . . .	36
4.3. Tipos de Patrones para textos narrativos . . . . .	42
4.4. Tipos de patrones para textos expositivos . . . . .	43
4.5. Tipos de Patrones para textos argumentativos . . . . .	44



# Índice general

1. Introducción	1
1.1. Planteamiento del problema . . . . .	1
1.1.1. Objetivos de la investigación . . . . .	2
1.1.2. Justificación de la investigación . . . . .	2
1.1.3. Pregunta de investigación . . . . .	3
1.2. Aportaciones de la investigación . . . . .	3
1.3. Organización de la tesis . . . . .	3
2. Marco teórico	5
2.1. Comprensión lectora . . . . .	6
2.2. Etapas de la comprensión lectora . . . . .	7
2.3. Relaciones Semánticas . . . . .	10
2.3.1. Entrañamiento . . . . .	11
2.4. Relaciones temáticas . . . . .	12
2.5. Gramáticas de valencias de Tesniere . . . . .	12
2.6. Patrones cognitivos . . . . .	14
2.7. Procesos inferenciales . . . . .	17
2.8. Cálculo de situaciones . . . . .	17
2.9. Sistemas basados en reglas . . . . .	22
3. Estado del arte	25
3.1. Modelos cognitivos . . . . .	25
3.1.1. Modelos de eventos . . . . .	28
3.1.2. Modelos contextuales . . . . .	28
3.2. Aplicaciones . . . . .	29
4. Modelo Semántico	31
4.1. Generación de Corpus de pasajes . . . . .	31

---

4.2.	Representación de pasajes . . . . .	32
4.3.	Determinación de Estructuras Semánticas . . . . .	37
4.3.1.	Estructuras semánticas para textos narrativos . . . . .	37
4.3.2.	Estructuras semánticas para textos explicativos . . . . .	39
4.3.3.	Estructuras semánticas para textos argumentativos . . . . .	41
4.4.	Extracción de patrones . . . . .	42
4.5.	Bloques cognitivos . . . . .	44
4.5.1.	Bloques cognitivos de textos narrativos . . . . .	45
4.5.2.	Bloques cognitivos de textos expositivos . . . . .	45
4.5.3.	Bloques cognitivos de textos argumentativos . . . . .	46
4.6.	Construcción de la arquitectura . . . . .	47
4.7.	Fases del modelo . . . . .	47
4.8.	Estrategia de solución . . . . .	48
4.8.1.	Solución basada en reglas . . . . .	50
5.	Resultados . . . . .	57
5.1.	Etiquetado . . . . .	57
5.2.	Casos de Estudio . . . . .	62
5.2.1.	Caso: textos expositivos . . . . .	62
5.2.2.	Caso: textos narrativos . . . . .	68
5.2.3.	Caso: textos argumentativos . . . . .	72
6.	Conclusiones y Trabajo Futuro . . . . .	79
7.	Publicaciones . . . . .	81

# Capítulo 1

## Introducción

La lectura es un proceso importante para la adquisición de conocimiento, actualmente ya existen audiolibros, resumidores e indexadores para los textos que son herramientas útiles para procesarlos. También la neurociencia, estudia 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 certera, como mencionan Kintsch & Dijk (1978).

Los lectores de textos en otros idiomas, distintos al nativo, tienen dos retos: por un lado; el primero es el proceso de traducción a su lengua nativa y el segundo es el mapeo de 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 comprensión lectora. Así en esta última sección, el lector construye un modelo de representación de conocimiento para responder a las preguntas que evalúan su nivel de comprensión, así como la aplicación de procesos inferenciales para comprender el significado del texto. Por otro lado, aunque se cuenten con herramientas de recuperación de información, los exámenes de acreditación de distintas disciplinas o idiomas no permiten su uso, puesto que se evalúan las habilidades del lector.

### 1.1. Planteamiento del problema

Actualmente, las Instituciones de Educación Superior plantean mecanismos de acreditación que evalúan el nivel de asimilación de conocimiento de los estudiantes, sobre todo en los procesos de razonamiento verbal y en el análisis de problemáticas en base a la información existente. Estas habilidades se sostienen en el desarrollo de la

capacidad inferencial, de acuerdo con los estímulos con los que cuente el lector; esto sugiere que una de las causas de estas deficiencias, es precisamente, la carencia de entrenamiento permanente desde la escolaridad básica, así como la falta de capacidad a la hora de integrar la información entre oraciones y complementarla con conocimientos anteriores, Peronard et al. (1997). La generación de modelos cognitivos ha permitido entender los procesos del pensamiento. Y si estos modelos se toman como base para la construcción de herramientas tecnológicas, los lectores se verán beneficiados para mejorar su eficacia en los procesos de acreditación.

El presente trabajo de investigación plantea la generación de un modelo semántico 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 esta sección.

### 1.1.1. Objetivos de la investigación

El objetivo general de este trabajo es diseñar un modelo semántico para la comprensión lectora de pasajes en el idioma inglés, con el propósito de realizar procesos de inferencia lógica. Los objetivos específicos se describen a continuación .

1. Diseñar un modelo a partir de las relaciones semánticas propias de la comprensión lectora
2. Proporcionar formalización teórica del modelo semántico para la comprensión lectora.
3. Proponer sistemas de deducción que permitan realizar inferencia sobre un texto.

### 1.1.2. Justificación de la investigación

Si bien existen simuladores de TOEFL para la comprensión lectora, estos sólo se enfocan en evaluar el contenido y arrojar los resultados para acreditar las habilidades de la lengua extranjera, pero carecen de herramientas de apoyo que brinden estrategias cuando los resultados no sean favorables. Esto se debe a la falta de modelos formales que permitan fundamentar: la discriminación de las relaciones semánticas, los esquemas

de comportamiento del proceso de comprensión, y la generación de inferencias ante cuestionamientos complejos.

Por lo que resulta pertinente abordar desde la perspectiva de la Ingeniería del Lenguaje y del Conocimiento, una solución formal considerando la versatilidad de la lógica de primer orden, el cálculo de situaciones y la semántica para modelar la información de los pasajes de este tipo de pruebas en una base de conocimiento.

### 1.1.3. Pregunta de investigación

Por lo que surge el siguiente cuestionamiento: ¿A partir de un modelo semántico es posible deducir relaciones semánticas e inferenciales entre partes de los pasajes?

## 1.2. Aportaciones de la investigación

El alcance de esta tesis radica, en proponer un modelo teórico que permita reconocer las relaciones semánticas e inferenciales propias de los textos de la comprensión lectora en los exámenes de acreditación del examen TOEFL. Así como establecer la formalización teórica del modelo semántico y finalmente, aplicar el modelo mediante un sistema de deducción que permita realizar inferencia sobre un texto específico.

## 1.3. Organización de la tesis

La tesis está conformada de la siguiente manera: en el capítulo 2 se define el marco teórico sobre las teorías de comprensión lectora, relaciones semánticas, gramáticas de Tesniere, procesos inferenciales que incluyen el cálculo de situaciones y los sistemas basados en reglas. En el capítulo 3 se menciona el estado del arte de los modelos cognitivos existentes y las herramientas de apoyo para la comprensión lectora, mientras que en el capítulo 4 se plantea como se realizó el modelo semántico, en el capítulo 5 se muestran los resultados obtenidos, en el capítulo 6 se mencionan las conclusiones y el trabajo a futuro a realizar y finalmente la última sección menciona la publicaciones realizadas e indexadas en scopus producto de esta trabajo.



# Capítulo 2

## Marco teórico

Leer es una actividad muy compleja, la mayoría de los procesos mentales que están implicados en la actividad de leer se han automatizado y se llevan a cabo con mucha rapidez (milésimas de segundo), por lo que resultan inaccesibles a la conciencia reflexiva del sujeto lector, Johnson-Laird (2006), García (1993).

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 (2008). 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. En la siguiente sección, se muestran los conceptos que fueron base para la elaboración de la propuesta, como son: comprensión lectora, relaciones semánticas y temáticas, gramáticas de valencias, patrones cognitivos, procesos inferenciales, cálculo de situaciones, reglas de asociación y sistemas basados en reglas.

## 2.1. Comprensión lectora

Un texto (o un discurso) es un conjunto de oraciones o enunciados relacionados y coherentes entre sí. Se suele emplear el término texto para referirse al lenguaje escrito, y discurso cuando se trata de lenguaje oral, aunque en otras ocasiones se emplea uno u otro término indistintamente, García (1993).

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 unidos de una forma lógica y consistente, Viramonte (2008).

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 (2008).

Los escritores pueden hacer más coherente un texto incorporando marcadores del discurso que hacen más claras las conexiones y transiciones, lo que permite informar al lector sobre la relación de un párrafo o segmento con otros.

Existen relaciones entre términos, oraciones incluso entre párrafos que representan causa - efecto, todo- parte, mientras otras relaciones son sólo conceptuales y se apoyan en un carácter jerárquico del léxico, incluso hay relaciones que se originan en el desarrollo del texto en sí ,Viramonte (2008).

Para Snow (2001) 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 esta 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 a García (1993) los procesos de comprensión lectora se pueden diferenciar en: decodificación, comprensión literal, comprensión inferencial y metacompreensión.

Decodificar significa descifrar un código; en este caso se trata de dar un significado a las letras impresas. Se admiten dos procesos decodificadores: asociar la palabra escrita

con el significado disponible en la memoria del sujeto. Recodificar implica transformar las letras impresas en sílabas y en sonidos para activar así el significado.

La comprensión literal consiste en combinar el significado de varias palabras de forma apropiada para formar proposiciones y se atiene a la información explícitamente reflejada en el texto, García (1993).

La comprensión inferencial proporciona una comprensión más profunda del texto y va más allá de lo explicitado en el mismo. El lector, mediante inferencias, elabora una representación mental más integrada y esquemática a partir de la información expresada en el texto y de sus conocimientos previos, García (1993).

La metacompreensión es la conciencia y control que el lector tiene de su proceso de comprensión. Consiste en establecer unas metas para la lectura, comprobar si se están alcanzando y rectificar oportunamente en su caso. Abarca procesos de planificación, supervisión y evaluación, García (1993).

## 2.2. Etapas de la comprensión lectora

Para comprender el significado se requiere un estado mental que se denomina representación mental, para lograr dicha representación se requiere realizar procesos de alto nivel cognoscitivo tales como relaciones, síntesis, interpretaciones e inferencias Amostoy (2002).

La inferencia es una relación entre dos situaciones o eventos. Este proceso es básico para adquirir ciertos tipos de pensamiento como el inductivo, el deductivo y el analógico.

Existen tres niveles en el procesamiento de la información: literal, inferencial y crítico. En primer lugar figura el nivel literal o concreto y las actividades que intervienen en este nivel se refieren a la recepción e identificación de la información. Respecto a esta actividad tenemos los verbos observar, identificar, discriminar, emparejar, recordar, ordenar.

El siguiente nivel es el inferencial donde los alumnos demuestran en qué forma aplican la información que recibieron y sus verbos asociados son: inferir, comparar, explicar, predecir, resumir, analizar, generalizar, interpretar, sacar conclusiones, reconocer las relaciones lógicas y parafrasear.

La capacidad de inferir o hacer una inferencia 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 al alcance y se transforma utilizandola de una manera nueva. Requiere la intervención de las habilidades de

pensamiento más sofisticadas, en este nivel los alumnos comienzan realmente a procesar y utilizar la información. Por ejemplo, extender el significado de las afirmaciones, explicar que es algo o qué es lo que hace, describir un objeto, resumir, hacer predicciones, identificar los puntos de vista personal y de los demás y resolver problemas (patrones cognitivos).

De acuerdo a 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 y 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 al alcance y se transforma utilizándola de una manera nueva. Requiere la intervención de las habilidades de pensamiento más sofisticadas. otro ejemplo, es extender el significado de las afirmaciones, explicar que es algo o qué es lo que hace, describir un objeto, resumir, hacer predicciones, identificar los puntos de vista personal y de los demás y resolver problemas, lo que corresponde a generar patrones cognitivos.

Para resolver inferencias 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?, ¿donde ocurrió el problema? Amostoy (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 Amostoy (2002).

En el proceso de comprensión, las preguntas son activadores 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, Amostoy (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 Amostoy (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.

Las preguntas que conectan lo que se dice en el texto con la experiencia previa, lo que propone (Kinstch,2001), en su modelo situacional, requieren establecer relaciones entre lo dado en el texto y el contenido de otro texto o algún suceso del pasado presente o futuro.

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.

Como herramienta de apoyo para la construcción del modelo mental, las relaciones semánticas permiten estructurar el conocimiento, por lo que se mencionan sus características en la siguiente sección.

### 2.3. 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, rosa pertenecen al campo semántico de las flores 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, 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, 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, Garrido (2010).

En el caso de que no se requiera un contexto, sino contar sólo con un vocabulario de términos que permita determinar hipónimos e hiperónimos, en este caso se puede emplear el entañamiento para expresar estas relaciones, como se muestra en Zenteno (2000).

### 2.3.1. Entrañamiento

De acuerdo a Zenteno (2000), el entrañamiento se identifica también como 'inferencia', según lo planteado por Kempson (1977). En cuanto al proceso inferencial, el entrañamiento es ampliamente utilizado por los lingüistas, tanto para explicar relaciones de sentido en el plano léxico como en el caso de: la hiponimia, la hiperonimia, la sinonimia, para dar cuenta de relaciones de sentido en el nivel oracional.

Por su carácter ambivalente, el entrañamiento puede ser definido, lógicamente, en términos de reglas válidas de inferencia o, semánticamente, en términos de la asignación de verdad o falsedad de proposiciones relacionadas: "(una proposición)  $p$  semánticamente entraña (una proposición)  $q$  sí y sólo si en cada situación en que  $p$  es verdadera  $q$  es también verdadera (o: en todos los mundos en que  $p$  es verdadera,  $q$  es verdadera)", Zenteno (2000).

Si el entrañamiento se maneja como una relación lógica entre proposiciones expresadas mediante oraciones, esta noción ha permitido relacionar sistemáticamente (con referencia al cálculo de predicados y sus relaciones, como la simetría, transitividad, y reflexividad) nociones tales como la hiponimia, la sinonimia, la antonimia, la oposición relacionada y la contradicción. Así, por ejemplo, Palmer señala que los hipónimos, predicados que establecen una relación de sentido tal, que el significado de uno está incluido en el de otro, envuelven entrañamiento: por ejemplo, rosa entraña flor. Los lexemas que se asocian mediante una relación hiponímica pueden también establecer transitividad: rosa entraña flor, y flor entraña ser vivo.

Existen otros tipos de entrañamiento entre proposiciones, por ejemplo en las proposiciones siguientes: David mató a Goliat, Goliat murió. La relación es válida considerando  $\text{Matar}(\text{David}, \text{Goliat})$  puede asegurarse que si esta proposición es verdadera implica que  $\text{Morir}(\text{Goliat})$  también lo es aunque no existan relaciones de hiponimia, sino más bien una causa con un efecto.

Dentro de las inferencias lingüística, Kempson define el entrañamiento como una inferencia lógico-semántica establecida entre proposiciones. Generalmente, el entrañamiento es presentado como una relación de interdependencia, en cuanto el valor de verdad de una proposición se deriva del valor de otra y puede ser inferido independientemente del contexto de enunciación. Por tanto, cualquier oración  $p$  implicará  $q$  si cuando  $p$  es verdadera,  $q$  también debe serlo, León & Pérez (2003).

Tabla 2.1: Tipos de Roles semánticos

Rol semántico	Descripción
Agente	El iniciador responsable de una acción
Paciente	La persona o cosa a la que la acción se aplica
Instrumento	El intermediario a través del cual una acción es ejecutada
Meta	La entidad hacia donde la acción es dirigida
Causa	La fuerza que provoca un cambio de estado
Ubicación	La ubicación de una acción o estado
Temporal	El tiempo en el que la acción o estado ocurre

## 2.4. Relaciones temáticas

Existen otras relaciones llamadas roles semánticos o relaciones temáticas, enfocadas a explicar tanto la estructura sintáctica como el significado de las palabras. Este enfoque se basa en el rol que juega el sustantivo respecto al verbo, es decir se realiza un mapeo del tipo de contribución: estado, acción o situación en el que participa el referente Garrido (2010). Los roles más comunes aparecen en la Tabla 2.4.

Pozo (2008) menciona que en la propuesta de Narayanan, los roles semánticos tienen una contribución particular en los sistemas de búsquedas de respuesta, con dos roles: complementario y como núcleo fundamental. En el caso de los roles como complementos se basa en el reconocimiento de entidades nombradas mientras que en los roles como núcleo plantea la resolución de preguntas complejas respecto a los aspectos temporales y causales de eventos complejos haciendo división en preguntas más sencillas, haciendo uso de información semántica y mecanismos de inferencia. Con este objetivo, Pozo (2008) plantea un uso doble de los roles semánticos, inicialmente utilizar las estructuras predicado argumento de la pregunta para determinar el modelo del tema de la misma y por otro utilizar los marcos semánticos relacionados con la pregunta y los pasajes relevantes con el fin de determinar el tipo de respuesta y extraer las respuestas esperadas, lo que implica el uso de inferencias complejas sobre estructuras de evento y de causas derivadas manualmente de estos marcos.

## 2.5. Gramáticas de valencias de Tesniere

Morales (2017) menciona la importancia de analizar el uso del verbo dentro de oraciones elaboradas a partir de dos sustantivos, evaluando en cada una que se cumplan los

parámetros semánticos y sintácticos que proporcionarán coherencia y cohesión al enunciado.

La Gramática Funcional, Morales (2017) es la primera aproximación lingüística que comenzó a darle relevancia a los factores semánticos dentro de la perspectiva sintáctica, e introdujo la noción de valencia o argumento. El concepto de valencia implica la relación entre un elemento léxico central y otros elementos léxicos que dependen de él y lo complementan. A través de éstos, se establece una relación entre el predicado verbal y los otros elementos que componen la predicación.

La teoría de las valencias se puede aplicar a cualquier categoría léxica no obstante, por su mayor complejidad se le ha dado más importancia a la valencia verbal. Esta aproximación propone que el núcleo predicativo es el elemento central en la constitución de las unidades predicativas y rechaza la primacía del sujeto, dándole al verbo la relevancia principal como el elemento central en todas las lenguas Lezcano (1977), como se maneja en la gramática de valencias de Tesnière.

Como ya se mencionó en apartados anteriores, el verbo tiene una doble función: desde el punto de vista semántico, expresa una relación entre grupos de objetos y fenómenos, y desde el punto de vista sintáctico realiza funciones gramaticales como organizador de la oración y portador de determinadas restricciones.

Por otro lado, las estructuras de los argumentos “activan marcos de procesamiento sintácticos y semánticos, que van más allá de la denominación concreta de una acción. Los argumentos del verbo se refieren a los roles de participación que cada acción necesita para ser posible y pueden identificarse aunque el verbo no esté conjugado. Tradicionalmente, se clasifican como transitivos e intransitivos Morales (2017), en donde los intransitivos requieren sólo un argumento; el agente: La mamá duerme, y los transitivos requieren, por lo menos, dos: un sujeto y un objeto, Ana lava los platos. Bajo esta perspectiva se han llegado a identificar hasta 8 tipos de argumentos: agente, tema, objeto, fuente, meta, beneficiario, complemento copulativo y objeto + verbo + inespecífico, Lezcano (1977).

De esta manera, los verbos se describen según el número de actantes o valencias y se proponen verbos cerovalentes, aquellos que no tienen ningún complemento, ni agente, como el verbo llover; verbos monovalentes, con un complemento por ejemplo nacer; verbos bivalentes que aparte del sujeto tienen otro complemento como recordar; verbos trivalentes, los que toman como valencias el sujeto y otros dos complementos como es el caso de prometer.

Por lo que se requiere abordar la teoría de los procesos inferenciales para así identificar las reglas implícitas en la comprensión lectora, considerando al verbo como actor

principal en las inferencias, por lo que se mencionarán a continuación.

## 2.6. Patrones cognitivos

El término cognitivo se refiere a la facilidad humana para pensar y razonar, esta fascinación sobre cómo se piensa, se razona y se resuelven problemas ha sido estudiado por años.

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; o simplemente, en el ámbito de lo cotidiano, la información incidental acerca de hechos o eventos del mundo que rodea al individuo. 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 Amostoy (2002). 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 Amostoy (2002).

El proceso es un operador intelectual capaz de transformar un estímulo externo en una representación mental, o en una acción motora. Los procesos son conceptos; cada proceso tiene un significado que lleva implícito la acción que lo caracteriza, la cual es ejecutada siguiendo el procedimiento que corresponde.

La práctica de procedimientos, bajo condiciones controladas, genera las habilidades de pensamiento. El proceso existe por sí mismo, independientemente de la persona que lo ejecuta, el procedimiento proviene de la operacionalización del proceso y la habilidad es una facultad de la persona, cuyo desarrollo requiere de un aprendizaje sistemático y deliberado Amostoy (2002).

Los procesos de pensamiento también pueden agruparse y ordenarse de acuerdo a sus niveles de complejidad y abstracción como sigue: procesos básicos, constituidos por seis operaciones elementales: observación, comparación, relación, clasificación simple, ordenamiento y clasificación jerárquica y tres procesos integradores: análisis, síntesis y evaluación, estos procesos son pilares fundamentales sobre los cuales se apoyan la construcción y la organización del conocimiento y el razonamiento. También existen los procesos superiores que son estructuras procedimentales complejas de alto nivel de abstracción como: los procesos directivos de planificación, supervisión, evaluación y retroalimentación, ejecutivos, de adquisición de conocimiento y discernimiento Amostoy

(2002).

Los niveles de procesamiento están secuenciados; cada nivel, a partir del primero, sirve de base para la construcción de los niveles que le siguen, El conocimiento, tanto semántico como procedimental, además, puede caracterizarse por el tipo de procesamiento a que da lugar. El conocimiento semántico o conceptual se refiere a dos categorías: a) la especificación de la esencia del concepto a través de las características esenciales de la clase que lo define o de la categoría conceptual correspondiente, y b) el conocimiento acerca del concepto -significación, importancia, utilidad, origen, razón de ser, trascendencia- del conocimiento. El primer caso se refiere al conocimiento de la esencia del concepto, mientras que el segundo se refiere al metaconocimiento Amostoy (2002). El conocimiento semántico o conceptual da origen a la formación de imágenes o representaciones mentales, bien sea del concepto o acerca del concepto.

El metaconocimiento puede referirse a la persona, a la tarea o a la estrategia, implica estar consciente acerca de las potencialidades y limitaciones de la persona; conocer el grado de dificultad o de complejidad de una tarea; o considerar la trascendencia de los actos que se realizan son ejemplos ilustrativos de metaconocimiento. Puede afirmarse que el metaconocimiento es una condición indispensable para el desarrollo de la habilidad de aprender a aprender.

El conocimiento procedimental da lugar a dos tipos de facultades: a) habilidad cognoscitiva, que se refiere simplemente al hábito de aplicar los pasos de un procedimiento para lograr un propósito determinado, y b) habilidades metacognoscitivas, que involucran la aplicación, en forma natural y espontánea, de cuatro de los procesos mentales superiores : planificación, supervisión, evaluación y retroalimentación, Amostoy (2002). La habilidad es la facultad de aplicar el conocimiento procedimental y puede referirse a la aplicación directa del proceso o a la evaluación y mejora de lo que se piensa y se hace.

Definir un concepto implica la observación directa o indirecta de distintos ejemplares pertenecientes a la clase que se desea definir, luego, la comparación de los ejemplos previamente observados, la identificación de las características esenciales que definen la categoría conceptual correspondiente, y finalmente, la observación de contraejemplos para discriminar el concepto de otros que se diferencien en al menos una característica. Desarrollar una habilidad implica la superación de la siguiente secuencia de etapas: conocimiento y comprensión de la operación mental que define el proceso; concientización de los pasos que conforman la definición operacional del proceso; aplicación, transferencia del proceso a variedad de situaciones y contextos; generalización de la aplicación del procedimiento; y evaluación y mejora continua del procedimiento. Para lograr la habi-

lidad de aplicar el proceso de manera. El análisis de relaciones entre pares de palabras efectiva es necesario practicarlo hasta lograr el hábito de utilizarlo, en forma natural y espontánea, en variedad de situaciones y contextos, adaptándolo de acuerdo a los requerimientos de la tarea Amostoy (2002).

Por otro lado, los moldes mentales o patrones cognitivos según Hernández-Guanir (2010) son los modos habituales con los que una persona se enfrenta cognoscitiva y afectivamente a la realidad y con los que interpreta y valora su relación con ella. También se consideran “constructos” derivados de la interacción entre las disposiciones genéticas y el aprendizaje que posibilita los distintos entornos. Así, cuanto más se ejecutan o se utilizan, más posibilidades habrá de que se vuelvan a aplicar en el futuro y en otras situaciones. Hernández-Guanir (2010) ejemplifica esta idea usando como símil la escena del agua sobre el terreno. Es decir, una vez que el agua abra un cauce en la tierra, posteriormente, será más probable que el agua vuelva a transitar por dicho cauce y, a su vez, cada vez que pase, ahondará más en él y lo hará más pronunciado.

El cuestionario MOLDES, Hernández-Guanir (2010) consta de 87 preguntas que describen comportamientos y modos de pensar ante diferentes situaciones; el sujeto debe indicar su grado de acuerdo entre estas descripciones y su forma de ser.

Los resultados del test se articulan de acuerdo a una estructura jerárquica formada por treinta moldes que responden a estrategias cognitivo-emocionales relacionadas con las funciones de anticipar, reaccionar, encauzar, evaluar, explicar o resolver situaciones que afectan a la identidad o a los intereses personales ( implicación directa, hiperanálisis, hipercontrol anticipatorio, afrontamiento borroso, anticipación aversiva, evaluación selectiva negativa, hostilgenia, inflación-decepción, focalizador en la carencia, imantación por lo imposible, anticipación devaluativa, disociación emocional, reclusión, oblicuidad cognitiva, anticipación esfuerzo y costo, precisión y supervisión, atribución falta de esfuerzo, justificación de los fallos, desplazamiento emocional, atribución al temperamento, atribución social del éxito, atribución mágica, atribución internalista del éxito, atribución a las estrategias, autoconvicción volitiva, automotivación proactiva, control emocional anticipatorio, anticipación constructiva previsor, evaluación positiva y transformación rentabilizadora), por lo probablemente la construcción del modelo mental durante cuestionamientos puede arrastrar dichos patrones, por lo que en cualquier diseño interactivo para apoyar la comprensión lectora, requerirá contemplar en su diseño estrategias para minimizar esos efectos adversos y favorecer la estructuración de un modelo mental adecuado para ese tipo de evaluaciones. A continuación se muestran los elementos característicos de la comprensión lectora.

## 2.7. Procesos inferenciales

Una inferencia se considera una conclusión u opinión a la que se llega tomando en cuenta evidencias o hechos conocidos. Desde el punto de vista de la lógica, una inferencia se define como el proceso de derivar consecuencias lógicas de las premisas asumidas, León & Pérez (2003).

Las inferencias cumplen diversas funciones en el procesamiento del discurso, permiten la identificación de referentes, desambiguar y/o completar significados semánticos, así como sustituir vacíos cuando la información no está disponible de forma explícita. La noción de inferencia lingüística la han definido como: cualquier conclusión que un lector razonablemente deriva a partir de una oración León & Pérez (2003). Otra categorización las divide en inferencias proposicionales y pragmáticas. Las primeras se basan en conocimiento lingüístico, y se derivan del contenido semántico de las proposiciones explícitas del texto. Algunos autores las llaman inferencias lógicas y son consideradas como necesariamente verdaderas.

Existe varios tipos de inferencia que son aplicadas al realizarse el proceso de lectura, como se menciona en el trabajo de León & Pérez (2003) y Gutiérrez-Calvo (1999) que caracterizan a las inferencias de acuerdo a diversos criterios y se presentan en la Tabla 2.2. En esta tabla se mencionan las inferencias de acuerdo a su criterio, el nombre por el que se conoce y el alcance de la misma.

A partir de esta clasificación de inferencias, se pueden modelar reglas que abarquen a cada una de ellas; estas reglas apoyarían a los mecanismos para la deducción de respuestas en exámenes de acreditación como en el caso del TOEFL.

Las inferencias permiten al lector construir representación de significado para después plantear cuestionamientos para responder preguntas específicas sobre los textos. Por lo que resulta importante revisar formalismos como el cálculo de situaciones, por lo que se describe a continuación.

## 2.8. Cálculo de situaciones

Para abordar el cálculo de situaciones, primero se definirán los elementos preliminares de la lógica de primer orden, como se menciona a continuación.

Un lenguaje de primer orden con igualdad es especificado por dos conjuntos disjuntos de símbolos llamados el vocabulario del lenguaje, Reither (2001):

Símbolos lógicos: la interpretación de estos es fijada por las reglas de la lógica de primer

Tabla 2.2: Clasificación de Inferencias

Criterio	Nombre	Alcance
Grado de probabilidad	Lógica	se genera en sistemas de razonamiento formal
vs. Certeza	Pragmática	se basa en el conocimiento de las personas
Temporalidad	en-línea	sucede después de la lectura
	fuera de línea	sucede durante la lectura
Recursos cognitivos	Automáticos	realizados concientemente
	Estratégicos	producidos con mayor tiempo de procesamiento
Dirección	hacia adelante	busca antecedentes
	hacia atrás	predice información derivable
Necesidad de entendimiento	Obligatorio	apoyan la formación del modelo mental
	No necesario	generado adicionalmente al modelo mental
Coherencia	Local	en frases
	Global	conexión entre partes distantes del texto
Tipos de contenido	Información de la inferencia	se apega al contenido
	Pregunta	se desencadenan de acuerdo a la pregunta
Recursos de información	Textual	contenido explícito
	Basada en el conocimiento	construida en base a su experiencia
Nivel de representación	Estructura proposicional	basado en coherencia local
	Modelo Situacional	Contextualiza la información

orden. Sus elementos básicos son:

- Parentésis ( , ).
- Connectivos lógicos: ( $\supset, \sim$ ).
- Variables:  $x, y, z, \dots$
- Igualdad  $=$ .
- Cuantificadores :  $\forall$  y  $\exists$ .
- Predicados: Para cada  $n \geq 0$ , un conjunto de símbolos, llamados predicados  $n$ -arios.
- Funciones: Para cada  $n \geq 0$ , un conjunto de símbolos, llamadas funciones  $n$ -arias.

Términos, fórmulas atómicas, literales, fórmulas bien formadas son definidas usualmente como los conceptos de ocurrencias libres o ligadas de una variable en una fórmula. Una sentencia es una fórmula sin variables libres. Los símbolos  $\forall, \wedge, \exists$  son definidos para ser abreviaturas adecuadas de ocurrencias de una variable en una fórmula. Supongamos que con un conjunto no vacío  $I$ , cuyos miembros son los tipos mencionados predicados y funciones.

Símbolos lógicos: Como antes, excepto que para cada tipo  $i$ , hay infinitamente muchas variables  $x_1^i, x_2^i, \dots$  de tipo  $i$ . A cada término se le asigna un tipo único, como se indica a continuación:

1. Cualquier variable de tipo  $i$  es un término de tipo  $i$
2. Si  $t_1, \dots, t_n$  son términos del tipo  $i_1, \dots, i_n$  respectivamente y  $f$  es un símbolo de función del tipo  $\langle i_1, \dots, i_n \rangle$ , entonces  $f(t_1, \dots, t_n)$  es un término de tipo  $i_{n+1}$

Las fórmulas atómicas se definen como sigue:

1. Cuando  $t$  y  $t'$  son términos del mismo tipo,  $t = t'$  es una fórmula atómica.
2. Cuando  $P$  es un símbolo del predicado del tipo  $t \langle i_1, \dots, i_n \rangle$  y  $t_1, \dots, t_n$  son términos del tipo  $i_1, \dots, i_n$  respectivamente, entonces  $P(t_1, \dots, t_n)$  es una fórmula atómica.

El cálculo de situaciones es un formalismo lógico diseñado para la representación y razonamiento acerca de los dominios dinámicos. Fue propuesto por John McCarthy, McCarthy & Buvac (1994) en 1963. Baker menciona que permite representar escenarios cambiantes como un conjunto de fórmulas de lógica de primer orden, Reither (2001). Reiter define los elementos básicos del cálculo de la siguiente forma, Reither (2001): Las acciones se consideran todos los cambios del mundo. Una historia posible del mundo, formada de una secuencia de acciones es representada por un término de primer orden llamada situación. Un fluente es una propiedad que puede o no sostener una situación dada.

También es definida la función  $do(\alpha, s)$  que denota una situación sucesora a  $s$  para ejecutar la acción  $\alpha$ . Por ejemplo, el predicado  $cause(virus, disease)$  indica la acción que un virus causa una enfermedad, si se aplica la función  $do(cause(virus, disease), s)$  significa la situación  $s$  resultante de ejecutar la acción de que los virus causan una enfermedad a partir de una situación. En el cálculo de situaciones, las acciones son denotadas como símbolos de función y las situaciones son términos de primer orden. Con este esquema se puede emplear la lógica de primer orden para formalizar los efectos de varias acciones.

- Dos símbolos de función de la situación
  - Un símbolo constante  $S_0$ , denotando la situación inicial
  - Un símbolo de función binaria  $do : action \star situation \rightarrow situation$
- Un símbolo de predicado binario  $\sqsubset : situation \star situation$ , definiendo una relación de clases de situaciones.
- Por cada  $n \geq 0$ , varios símbolos de predicado infinitamente contables con aridad  $n$ , y tipos  $(action \cup object)^n$ . Estos se utilizan para denotar relaciones independientes de la situación como por ejemplo,  $virus(Rotavirus)$ .

Reither (2001) considera que esta representación permite modelar esquemas de tipo pregunta-respuesta, considerando plantear una secuencia de términos de acción y una fórmula  $G$ , ésta será verdadera de acuerdo a la ejecución de las acciones que contiene. Sin embargo, existe el problema llamado de proyección que permite analizar este enfoque y Reiter lo define de esta forma:

Suponga que  $D$  es una teoría de acción,  $\alpha_1, \dots, \alpha_n$  es una secuencia de términos acción, y  $G(s)$  es una fórmula sin variables libres  $s$ , cuyo término de situación es  $s$ . Determinar si:  $D \models G(do([\alpha_1, \dots, \alpha_n], S_0))$ .

Por ejemplo, una consulta de proyección para la secuencia de acciones  $Closest\_meaning(proven, showed)$  podría interpretarse: La palabra *proven* es el significado más cercano respecto a la palabra *showed* sobre el contexto de los virus

$$D \models is(do(Closest\_meaning(proven, showed), virus)).$$

Estas definiciones del cálculo de situaciones se puede extender para formalizar el contexto. McCarthy McCarthy & Buvac (1994) propone el término aserción de la forma  $assert(c, p)$ , en este caso la aserción indica que la proposición  $p$  bajo el contexto  $c$  puede ser evaluada o ejecutada. Por otro lado el examinar conversaciones con bloque query-answer en este modelo, se plantean dos tipos de preguntas : las proposicionales que son usadas para determinar si una proposición es falsa o verdadera, por lo que requieren una respuesta de tipo Yes o No, y las preguntas cualitativas son usadas para encontrar objetos que sostienen una fórmula. Para modelar el discurso se pueden plantear las funciones query y reply que son la representación central en estos discursos. Así la función query establece un contexto en el cual la respuesta a la pregunta será interpretada. Por ejemplo, si se tiene la proposición  $p$  es posible que tenga valor verdadero de acuerdo al contexto que se interprete. Así la función *reply* actualizará la información, es decir, sólo cambiará el estado epistémico del contexto del discurso.

Esto deriva una serie de axiomas al respecto:

- Axioma de Interpretación (proposicional): si  $\phi$  es una fórmula cerrada, entonces  $assert(query(K, \phi), \phi \equiv yes)$
- Axioma Marco (proposicional): si  $\phi$  es una fórmula cerrada, y la respuesta yes no ocurre en el contexto  $\Psi$  entonces  $assert(K, \Psi) \supset assert(query(K, \phi), \Psi)$
- Axioma de Interpretación (Cualitativo) si  $x$  es una variable libre en  $\phi$  entonces  $assert(query(K, \phi(X)), \phi(X) \equiv answer(X))$
- Axioma de Marco(Cualitativo) si  $x$  es una variable libre única free unique en  $\phi$  y  $answer$  no ocurre en el contexto  $\Psi$ , entonces  $assert(K, \Psi) \supset assert(query(K, \phi(X)), \Psi)$
- Axioma de Respuesta :  $assert(reply(K, \phi), \Psi) \equiv assert(K, \phi \supset \Psi)$

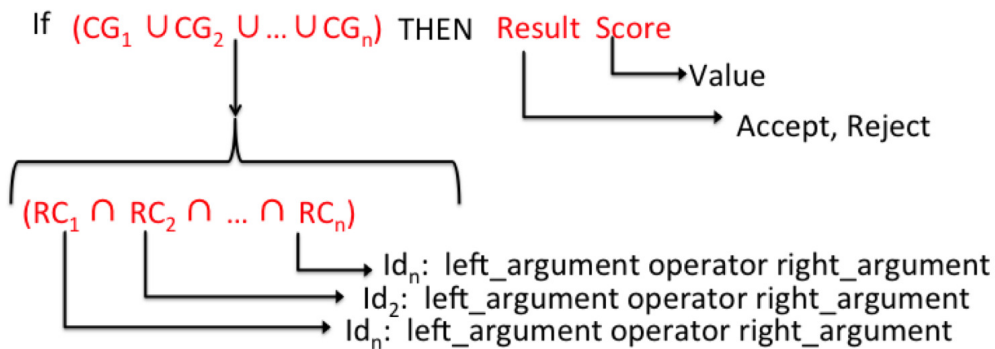


Figura 2.1: Ejemplo de especificación de reglas Castillo (1997)

## 2.9. Sistemas basados en reglas

Cuando se modela conocimiento se requiere analizar las propiedades y relaciones en un dominio específico, existen varios mecanismos para representar conocimiento, uno de ellos son las reglas que son una forma intuitiva de representar esquemas mediante una estructura que facilite la determinación a partir de una premisa y una conclusión. En el caso de los cuestionamientos de pregunta respuesta que se aplican en exámenes de acreditación tipo TOEFL, se requiere analizar qué patrones están vinculados con la pregunta y relacionados con la respuesta, para que a partir de este esquema y la información complementaria provista por los documentos, se pueda dar una solución acertada. La ventaja de la representación es que arroja un resultado determinista que otros esquemas de representación, que en su lugar hacen un manejo probabilístico. En esta sección se propone mediante reglas inferenciales el modelado de patrones para dar respuesta a cuestionamientos sobre textos académicos de tipo TOEFL

Un sistema experto puede definirse como un programa inteligente que utiliza conocimiento y procedimientos de inferencia para resolver problemas que son lo bastante complejos como para requerir expertos humanos en su resolución. Los problemas que se resuelven pueden ser clasificados según su naturaleza, en deterministas ó estocásticos Castillo (1997).

El primero grupo está conformado por aquellos problemas que se pueden plantear utilizando un conjunto de reglas, mientras el segundo grupo maneja situaciones inciertas y necesitan incorporar la medida intuitiva de incertidumbre que es la probabilidad, en nuestro caso abordamos el enfoque basado en reglas considerando que se requiere indicar la respuesta correcta de forma determinística Castillo (1997)

Las reglas en un sistema inteligente son consideradas una representación determinística

del conocimiento, a continuación en la Figura 2.1 se muestra el esquema general de representación de reglas (Castillo (1997)). Esta representación es usada considerando su cercanía al sentido común de los seres humanos, cómo se puede observar en el esquema inicial tenemos un antecedente y un consecuente o bien una premisa y una conclusión, en este caso se tiene un grupo de condiciones que están dentro del bloque condicional y en el resultado se dice si es aceptado o rechazado.

Dentro de los grupos de condición tenemos un conjunto de reglas que se van aplicando y dependiendo de esa aplicación se van descartando las posibles respuestas dado que tiene como operador principal un and si alguna de las reglas se inválida entonces en automático ya es inválido ese grupo de condiciones y entonces como el condicional tiene el operador o se requiere que al menos algunas de estas grupo de reglas condicionales sea válido para que se pueda dar el resultado. Las condiciones de reglas se caracterizan por definirse con un argumento a la izquierda un operador y un argumento a la derecha. A partir de estos elementos se requiere identificar la estructura para representar estas aserciones y generar inferencias cercanas a las que suceden en el modelo mental de comprensión lectora, por lo que en el siguiente capítulo se mostrará el estado del arte respecto la generación de inferencias en comprensión lectora.



# Capítulo 3

## Estado del arte

La representación de los procesos cognitivos en la comprensión lectora ha dado pie a teorías y modelos desde el punto de vista psicológico y computacional. El interés radica en explicar las habilidades, estrategias y capacidades del lector, tanto para mejorar su aprendizaje como para brindar herramientas que le sirvan para mejorar su rendimiento.

### 3.1. Modelos cognitivos

Los modelos cognitivos se caracterizan por estudiar cómo el ser humano conoce, piensa y recuerda, explora la capacidad de las mentes humanas para modificar y controlar la forma en que los estímulos afectan la conducta y sustenta al aprendizaje como un proceso donde se modifican significados de manera interna.

Esto lo logra integrando los mecanismos de memoria a corto y largo plazo con los de inferencia y aunque se realizan de forma automática, no todos los seres humanos lo realizan al mismo tiempo o de la misma forma. Como menciona Johnson-Laird (2006), para entender instrucciones en experimentos de razonamiento los estudiantes requieren entender los conceptos de premisa, conclusión e implicación para realizar una deducción correcta, incluso aunque se han realizado esfuerzos enseñando cursos de lógica, los resultados no han sido favorables, puesto que sólo se da un enfoque matemático y no se relaciona la asociación de significados de acuerdo a la experiencia de cada estudiante, por lo que se dificulta su desempeño en el proceso de comprensión lectora.

La comprensión lectora es el proceso simultáneo de extraer y construir significado y cuyos objetivos son: descifrar como las letras representan las palabras, abordar la traducción en forma precisa y eficiente de las letras a sonidos (extraer significado del texto), formular una representación de la información que esta siendo presentada, que

inevitablemente requiere la elaboración de nuevos significados y la integración de la información nueva con la antigua (construcción del significado) Snow (2001). Este último objetivo es el que presenta mayor interés tanto por psicólogos mediante la generación de modelos cognitivos, como por computólogos que han buscado mecanismos para explicar o emular los procesos del pensamiento con ayuda de la inteligencia artificial y el procesamiento del lenguaje natural.

Desde el punto de vista de la psicología cognitiva, interesada por la comprensión del discurso, se han venido desarrollando, desde hace un par de décadas, un sinnúmero de modelos teóricos que han tratado de explicar cómo se produce la comprensión, precisando como factores: el papel del conocimiento previo del lector, la realización de inferencias o la construcción de distintos niveles de representación mental que interactúan con las características del texto.

Los precursores de estos modelos mentales fueron Van Dijk y Kintsch, Kintsch & Dijk (1978) con su artículo en la revista *Psychological Review*, que explicaba en detalle el procesamiento cognitivo de un texto universitario de la psicología social. En este trabajo se buscaba comprender cómo se recuerdan los textos que se leen y postula que al leer un texto se trabaja con tres niveles de representación mental: el código de superficie, la base de texto y el modelo situacional. Dos conceptos clave en este proceso de recuerdo eran la macroestructura y la superestructura, las cuáles fueron propuestas en esa investigación. Esta teoría suponía que el procesamiento textual se hace por ciclos, debido a la limitada capacidad de la memoria de corto plazo después de la decodificación del código, y que de esta manera se construía gradualmente una representación del texto (texto base) en la memoria. Este texto base no solo consistía de una secuencia conectada de proposiciones, sino que también establecía una estructura jerárquica de macroproposiciones, que corresponden a los temas más importantes y menos importantes del texto deducidos (inferidos) por el lector.

El texto base es el resultado de las secuencias de proposiciones que se hacen coherentes por la repetición de argumentos. Las macroestructuras, por otro lado, se pueden definir como proposiciones de orden superior que incluyen proposiciones subyacentes. En otras palabras, las macroproposiciones están construidas con las microproposiciones de un texto, y son un resumen u otra estructura abstracta subyacente a un texto, por lo que ellas deben inferirse desde el texto. Así las micro y las macroproposiciones forman una macroestructura del texto, es decir, una estructura semántica que define el significado global de un texto. Sin embargo, estas estructuras requerían asociarse con un contexto asociado a la experiencia del lector, lo que formaba el modelo situacional, que es un modelo cognitivo de la situación reflejada en el texto que contiene material inferido.

Otro modelo propuesto es el de Miller y Kintsch, Kintsch (1988) que consistía de dos componentes: un programa de bloques y un programa de coherencia de microestructura por lo que hacían uso de un grafo de coherencia para la memoria de trabajo, pero no se consideró la realización de inferencias.

También en 1995 se propuso el modelo 3CAPS por Goldman, Varma y Cote, éste provee interacciones entre el procesamiento de texto, conocimiento a priori y procesos de lectura estratégicos Goldman & Varma (1995). Más adelante Kintsch en 1998, Kintsch (1988), propuso el modelo de Construcción- Integración considerando las redes de nodos y enlaces entre ellos, mapeando estas relaciones a una matriz de coherencia. Considerando la recuperación de la memoria para apoyar el proceso inferencial, Singer y Kintsch en 2001, Singer & Kintsch (2001), proponen una plataforma a partir del modelo de Construcción-Integración con el modelo de enlace memoria de Gillund y Shiffrin, en un esfuerzo por modelar los patrones complejos en los datos de memoria inferencial. Por otro lado Lin and Patel, Lin & Pantel (2001), usan el modelo Minipar para extraer reglas de inferencia del tipo si X resuelve Y entonces X trata con Y. Gildea y Jurafsky en 2002 Gildea & Jurafsky (2002), intentan etiquetar los roles semánticos en una sentencia. En su sistema un clasificador es entrenado mediante el uso de características como voz, tipo de gramática y predicados.

El modelo de espacio de situaciones distribuido de Frank, Koppen, Noordman y Vonk, propuesto en 2003, Frank et al. (2003), simula como las inferencias basadas en conocimiento son realizadas durante la comprensión de historias. Por otro lado, el modelo paradigmático sintagmático (SP) de Dennis y Kintsch, en 2012, Dennis & Kintsch (2012) provee un mecanismo candidato para resolver el problema de mapeo de texto, tiene un enfoque orientado a la memoria y genera representaciones de conocimiento estructurado automáticamente a partir de un corpus.

El modelo de Collins en Burkert (1995), plantea para la representación de conocimiento una red con elementos llamados nodos tipados y nodos de ocurrencia, para la representación de relaciones asociativas que conectan los nodos unos con otros emplea el operador and y or. Las aristas se consideran relaciones binarias.

El modelo de Miller y Johnson Laird, en Burkert (1995) estuvieron interesados en la relación entre percepción y lenguaje, así como los procesos que permiten a los objetos y eventos ser asociados mediante etiquetas léxicas, la percepción es vista como una actividad de conceptualización referente a esquemas. En su modelo describen que un concepto léxico consiste de una etiqueta, una regla acerca del comportamiento sintáctico de la etiqueta.

Aún cuando se han propuesto varios modelos cognitivos, el modelo de Kintsch y Van

Dijk, cuenta con elementos interrelacionados que fusionan a la psicología cognitiva y a la lógica de predicados para el apoyo en el proceso de la comprensión lectora, lo que resulta interesante abordar desde el punto de vista del razonamiento no monotónico.

### 3.1.1. Modelos de eventos

Se dice que el contexto influencia las estructuras y procesamiento de texto y del habla, anteriormente los modelos mentales de la memoria episódica era limitado a sólo representaciones mentales de los eventos Dijk (2009). Los modelos de eventos representan la interpretación subjetiva del discurso mediante la generalización y la abstracción de la información representada y proveen la base para compartir el conocimiento y son construidos de la información derivada del discurso.

Respecto al modelado de relaciones respecto el tiempo, en ? utilizan el concepto de circunstancia: como un factor externo que afecta un agente en concreto pueden ser permanentes o puntuales, también se considera una condición o característica no esencial de tiempo lugar y modo que rodea a una persona o cosa y que influye en ellas, con este concepto desarrollan minería de datos para encontrar patrones de interés en una base de datos.

Por otro lado Woszezenki et al. (2018), trabaja modelado de relaciones mediante ontologías, pero proponen dentro de su arquitectura el concepto de asociación, para descubrir conocimiento.

Estas tendencias, han dado pie a formalizar aún más el modelo semántico para textos narrativos.

### 3.1.2. Modelos contextuales

Los modelos de contexto Dijk (2009) son un caso especial de un tipo más general que son los modelos de experiencia que toman sentido de los episodios de la vida diaria y se consideran subjetivos. Estos modelos constan de varias clases de información analógica y proposicional considerando categorías como tiempo, ubicación, circunstancia, participantes en varios roles metas y varios tipos de actividades como sus propiedades.

Otros modelos existentes como el modelo de Paisaje y el modelo de paisaje con análisis semántico latente Yeari & van den Broek (2011), se basan en los picos y crestas cuando el cerebro accede a la memoria, para analizar el impacto del proceso de comprensión en el aprendizaje, en el caso del modelo con análisis semántico latente amplía el primero

considerando la relación de los términos mediante un análisis de la distancia entre los mismos.

## 3.2. Aplicaciones

Desde el punto de vista de las aplicaciones se cuenta con IStart que es propuesto por McNamara et al. (2004) que propone esta herramienta interactiva para apoyar a los estudiantes en el entrenamiento de lo que sería la lectura de comprensión. Toma como referencia el modelo de Kinstch y su aplicación cuenta con tres fases: una explicación introductoria que incluye definiciones y ejemplos una demostración de las técnicas en acción y la práctica correspondiente como última fase. La fase que más destaca es la fase de demostración donde cuenta con avatares para manejar los elementos necesarios para explicar lo que sería el monitorio de la comprensión la parafrasis, las inferencias elaborativas, hacer predicciones y hacer puentes entre inferencias, de manera que el texto es leído y es explicado sentencia en sentencia y el facilitador sería un personaje llamado Merlín que hace las explicaciones necesarias para que al estudiante le queden claras. También cuenta con un módulo adaptativo que permite dependiendo de la pregunta hacer que Merlín se adapte a elaborar preguntas más complejas, esta fase se considera una retroalimentación considerando que va guiando al estudiante sobre cuáles eran las estrategias que le permiten ir asociando la respuesta correcta. En la fase de práctica el estudiante intenta usar las estrategias de lectura aprendidas en la introducción y en la demostración, lo que puede observarse en los cuestionamientos es que son preguntas cerradas sobre las técnicas que usó para interpretar el texto y maneja retroalimentación y las preguntas están enfocadas únicamente a las estrategias usadas para dar la respuesta, sus algoritmos están basados en varios factores: las palabras del contenido que corresponden a la sentencia destino incluyendo pronombres adjetivos verbos y adverbios, por otro lado las palabras asociadas para que cada palabra del contenido se identifiquen manualmente en base a los protocolos recolectados y a los diccionarios y una prioridad de la longitud manualmente asignada a cada sentencia lo cual determina qué tan larga es la explicación que debe ser generada. Si bien esta herramienta es bastante interesante por qué ya está incorporando las recomendaciones de lectura de comprensión, son textos que no son complejos en el caso de nuestra propuesta requerimos procesar textos académicos que implican elementos aún más complejos que las preguntas de esta herramienta, pues se requiere apoyar al estudiante con la generación de inferencias para así facilitar la obtención de la respuesta requerida. Otra aplicación generada por IBM es Debater, Reither (2001) menciona que es una

plataforma de inteligencia artificial basada en la nube, para el apoyo de recursos compartidos en la toma de decisiones, reúne argumentos a favor y en contra de un tema específico. Trabaja de la siguiente forma: primero escucha a las personas debatir sobre un cierto tema, recopila la información y crea discursos basados en esa información.

La base de esta plataforma se conforma de varios investigadores, como Hou (2017), propone una clasificación de relaciones argumentativas usando una inferencia conjunta basada en distribuciones de probabilidad, su línea de investigación radica en lo que se llama minería de argumentación, uno de los enfoques que también trabaja son las redes lógicas markovianas y el modelo de inferencia combinada para generar argumentos, Reither (2001). Por otro lado, Reither (2001), indica que Bilu (2015), propone dentro del enfoque de la minería de argumentación analizar las estructuras argumentativas de manera que sea posible realizar la negación de conclusiones de acuerdo a las preguntas que abarquen de forma automática. Maneja como una primera metodología tener la oración y simplemente aplicar la negación como un agregado a la oración inicial sin hacer detección del contexto aunque se basa en los criterios de gramática, caridad, posición. Propone un algoritmo de acuerdo a las preguntas: cómo, porqué y cuándo, parte de generar un POS para saber en que posición debe de generar la negación y lo hace de manera automática

A partir de los modelos existentes se pueden generar nuevos modelos adecuados a las necesidades de la comprensión lectora en el idioma inglés, la metodología del diseño del modelo y sus características se mencionan en el siguiente capítulo.

# Capítulo 4

## Modelo Semántico

En este capítulo se presentan los procesos y elementos que permitieron generar el modelo semántico para determinar las respuestas correctas en los textos tipo TOEFL. Por lo que se plantea en el Algoritmo 1, los pasos para la generación del modelo semántico que serán revisados en cada sección de este capítulo.

---

### Algoritmo 1 Algoritmo de generación de modelo semántico

---

Entrada: ninguna

Salida: modelo semántico

- 1: Generar el corpus de pasajes.
  - 2: for cada pasaje  $P$  do
  - 3:   Crear la representación de  $P$  con lógica de predicados y cálculo de situaciones.
  - 4:   Determinar las estructuras semánticas de la sección del cuerpo del pasaje  $P$ .
  - 5:   Extraer patrones cognitivos de la sección de preguntas del pasaje  $P$ .
  - 6: end for
  - 7: Producir los bloques cognitivos a partir del corpus de pasajes y de los patrones cognitivos para determinar el tipo de texto.
  - 8: Especificar la arquitectura semántica para la inferencia de la respuesta correcta de los pasajes.
  - 9: Especificar modelo basado en reglas.
- 

### 4.1. Generación de Corpus de pasajes

Se realizó una investigación sobre las características del examen TOEFL en la sección de comprensión lectora Mahnke & C.B.Duffy (1993) que evalúa tres habilidades en la

lectura:

- 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 corresponde a uno de los siguientes 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.

Posteriormente se buscaron ejemplos de estos tipos de textos y se encontraron 300 pasajes en la Universidad de Chuvanan Chuvanan-University (2018). Cada pasaje posee una estructura con el cuerpo del pasaje, el bloque de preguntas con sus cuatro posibles respuestas y al final un apartado con el listado de respuestas correctas.

Posteriormente para la implantación del modelo se puede generar un crawler que es un rastreador web que busca textos que se encuentren en distintos sitios web considerando que ya se tiene la estructura clara y un corpus de ejemplo.

Una vez que se cuenta con el corpus, el siguiente paso es realizar la representación de pasajes como se menciona a continuación.

## 4.2. Representación de pasajes

Para representar los pasajes fue requerido analizar la lógica de predicados y el cálculo de situaciones para plantear un esquema que permitiera el uso de inferencias. Por lo que se realizaron los siguientes pasos sobre cada pasaje:

1. Identificar el tipo de texto del pasaje a partir de la definición.

2. Determinar las relaciones semánticas para cada tipo de documento, se identificaron relaciones tanto en el texto del pasaje como en el bloque pregunta-respuesta, reiterando que la menor complejidad le corresponde a los textos narrativos de acuerdo a Amostoy Amostoy (2002) , la intermedia a los expositivos y la mayor a los argumentativos.
  - a) En el caso de los textos explicativos, se encuentran presentes varias relaciones entre las que destacan la hiperonimia e hiponimia, así como la meronimia y la sinonimia.
  - b) En el caso de los textos narrativos, se identificaron relaciones temporales relacionando eventos y agentes.
  - c) En el caso de los textos argumentativos, se identificaron relaciones entre hechos, opiniones y evidencias para generar el proceso de inferencia apoyando o refutando las opiniones.
3. Generar predicados que modelen las relaciones semánticas.

Por ejemplo, en la Figura 4.1, se muestra un ejemplo de un texto explicativo del tema de los virus, disponible en el sitio web de la Universidad de Chuvanan Chuvanan-University (2018). La estructura del pasaje es la siguiente: en el primer párrafo el tema central es introducido, mientras que las secciones subsecuentes los componentes de los virus y la forma en que ellos atacan son descritos.

Se identificaron las relaciones semánticas como los hiperónimos e hipónimos que son encontrados en las sentencias: *Viruses are very simple pieces of organic material; They are parasites*, merónimos en las sentencias: *Viruses are composed only of nucleic acid, either DNA or RNA, enclosed in a coat of protein made up of simple structural units*, ( *Some viruses also contain carbohydrates and lipids*).

Finalmente después del cuerpo del pasaje las preguntas son colocadas, las cuales tienen diferentes niveles de complejidad desde la identificación del tema principal, sinonimia, meronimia.

Por tanto considerando el análisis, se puede construir esta representación usando estas relaciones y el cálculo de situaciones. En la Tabla 4.1, se presenta la base de conocimiento.

En la mayoría de los casos los predicados modelados son bivalentes o trivalentes, de acuerdo, a la gramática de Tesnière, como en el caso: *derive(virus, poison)*, *emanate(noxious stench, swamps)*, en otros casos los predicados necesitan ser mode-

## Topic: Virus

The term 'virus' is derived from the Latin word for poison, or slime. It was originally applied to the noxious stench emanating from swamps that was thought to cause a variety of diseases in the centuries before microbes were discovered and specifically linked to illness. But it was not until almost the end of the nineteenth century that a true virus was proven to be the cause of a disease.

The nature of viruses made them impossible to detect for many years even after bacteria had been discovered and studied. Not only are viruses too small to be seen with a light microscope, they also cannot be detected through their biological activity, except as it occurs in conjunction with other organisms. In fact, viruses show no traces of biological activity by themselves. Unlike bacteria, they are not living agents in the strictest sense. Viruses are very simple pieces of organic material composed only of nucleic acid, either DNA or RNA, enclosed in a coat of protein made up of simple structural units. (Some viruses also contain carbohydrates and lipids.) They are parasites, requiring human, animal, or plant cells to live. The virus replicates by attaching to a cell and injecting its nucleic acid. Once inside the cell, the DNA or RNA that contains the virus' genetic information takes over the cell's biological machinery, and the cell begins to manufacture viral proteins rather than its own.

1. Which of the following is the best title for the passage.

(A) New Developments in Viral Research (B) Exploring the Causes of Disease (C) DNA: Nature's Building Block (D) Understanding Viruses

2. Before microbes were discovered It was believed that some diseases were caused by

(A) germ-carrying insects (B) certain strains of bacteria (C) foul odors released from swamps (D) slimy creatures living near swamps

3. The word "proven" in line 4 is closest meaning to which of the following.

(A) Shown (B) Feared (C) Imagined (D) Considered

4. The word "nature" in line 6 is closest in meaning to which of the following?

(A) Self-sufficiency (B) Shapes (C) Characteristics (D) Speed

5. All of the following may be components of a virus EXCEPT

(A) RNA (B) plant cells (C) carbohydrates (D) a coat of protein

6. The author implies that bacteria were investigated earlier than viruses because

(A) bacteria are easier to detect (B) bacteria are harder to eradicate (C) viruses are extremely poisonous (D) viruses are found only in hot climates

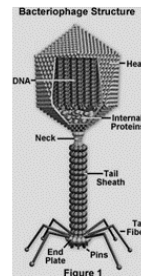


Figura 4.1: Ejemplo texto explicativo, fuente Universidad de Chuvanan Chuvanan-University (2018)

Tabla 4.1: Representación de las relaciones semánticas del tópico de los virus.

Oraciones	Aserciones
The term virus is derived from the Latin word for poison or slime.	<i>derive(virus, poison).</i> <i>derive(virus, slime).</i>
It was originally applied to the noxious stench emanating from swamps that was thought to cause a variety of diseases in the centuries before microbes were discovered and specifically linked to illness.	<i>emanate(noxiousstench, swamps).</i> <i>cause(diseases, virus)</i>
But it was not until almost almost the end of the nineteenth century that a true virus was proven to be the cause of a disease.	<i>cause(diseases, emanate(noxiousstench, swamps))</i> <i>prove(caused(diseases, virus))</i>
The nature of viruses made them impossible to detect for many years even after bacteria had been discovered	<i>detect(virus, hard).</i> <i>detect(bacteria, easy).</i>
Not only are viruses too small to be seen with a light microscope, they also cannot be detected through their biological activity, except as it occurs in conjunction with other organisms.	<i>isBiggerThan(bacterias, virus)</i> <i>detect(virus, and(virus, organisms)).</i>
In fact, viruses show no traces of biological activity by themselves.	<i>not(trace(virus, biologicalActivity))</i> <i>not(hyponym(LivingAgents, virus))</i>
Unlike bacteria, they are not living agents in the strictest sense	<i>hyponym(LivingAgents, bacterias)</i> <i>hyponym(materialOrganic, virus).</i>
Viruses are very simple pieces of organic material composed only of nucleic acid, either DNA or RNA, enclosed in a coat of protein made up of simple structural units.	<i>composed(organicMaterial, nucleicAcid)</i> <i>hyponym(nucleicAcid, DNA)</i> <i>hyponym(RNA, nucleicAcid)</i> <i>enclose(nucleicAcid, coatProtein)</i>
(Some viruses also contain carbohydrates and lipids.)	<i>contain(virus, carbohydrates)</i> <i>contain(virus, lipids)</i>



De forma similar se analizaron los otros dos tipos de texto, por lo que a partir de los elementos representados se puede determinar las estructuras semánticas que caracterizan a los mismo como se muestran en la siguiente sección.

### 4.3. Determinación de Estructuras Semánticas

De acuerdo a Dijk (2009), 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, se modelaron las estructuras semánticas de los textos narrativos, explicativos y finalmente los argumentativos.

#### 4.3.1. Estructuras semánticas para textos narrativos

En la Figura 4.2 se muestra la estructura propuesta para representar un texto narrativo, la estructura se generó a partir del análisis del corpus revisado para el examen de acreditación del TOEFL. Los documentos que integran el corpus 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. Estos elementos pueden representarse mediante aserciones de la forma *action(agent [recipient, detail,time])*, donde la acción(predicado) es un verbo que realiza un agente y dicha acción puede ser recibida por el remitente, o puede tener un detalle particular en un periodo de tiempo específico. Estas aserciones se basan en las relaciones y roles semánticos que permiten interpretar el contenido más significativo de los textos.

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 al periodo de tiempo. Esta línea de acción de agente 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 sobre un tema particular, 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

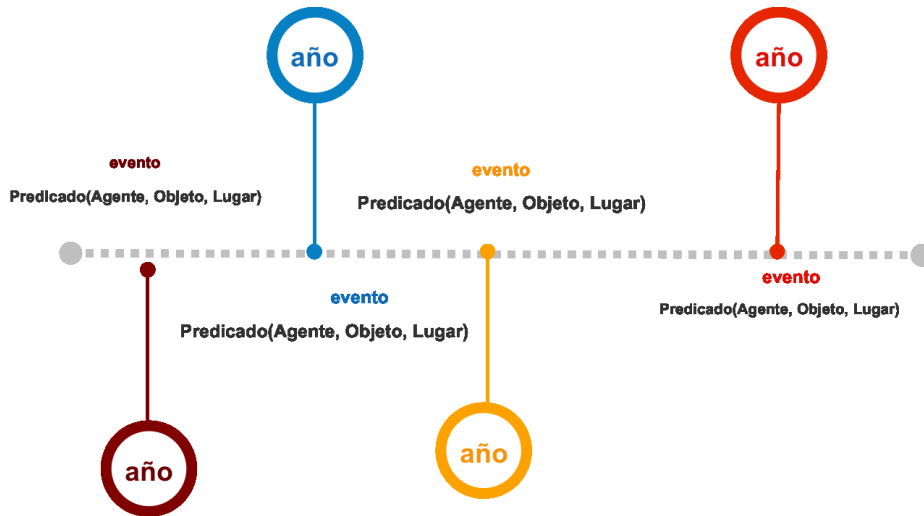


Figura 4.2: Estructura Semántica Línea de Tiempo

## Agente

### Estados

### Acciones

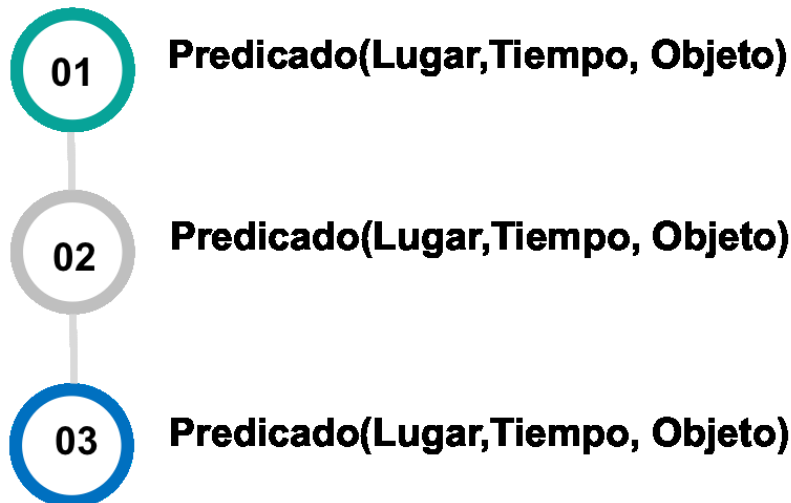


Figura 4.3: Estructura Semántica Línea de Acción del Agente

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, los resultados indicaron que las relaciones semánticas presentes corresponden a relaciones 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.

En la Figura 4.3 se muestra otra representación para los textos narrativos de significado con la estructura semántica de diagrama de estados de un agente. En este caso, un diagrama de estado especifica el comportamiento de una parte del sistema diseñado a través de transiciones de estados finitos. En la propuesta se define como la secuencia de aserciones que funjen como acciones de los agentes durante la narración y se ordenan de acuerdo a su ocurrencia en el texto.

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

#### 4.3.2. Estructuras semánticas para textos explicativos

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 4.4.

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 4.5.

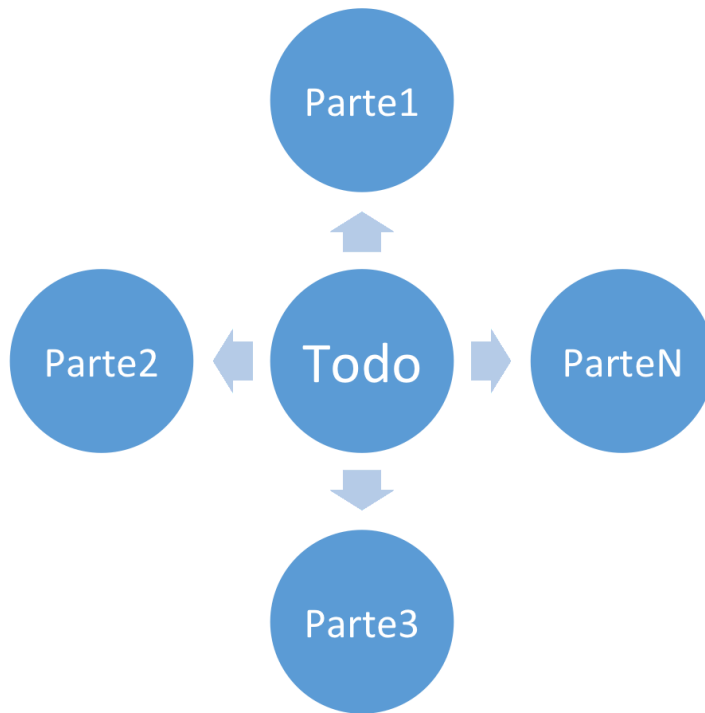


Figura 4.4: Merónimos

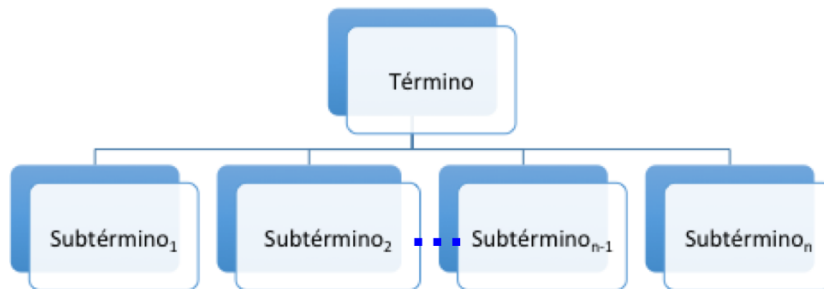


Figura 4.5: Co-Hipónimos

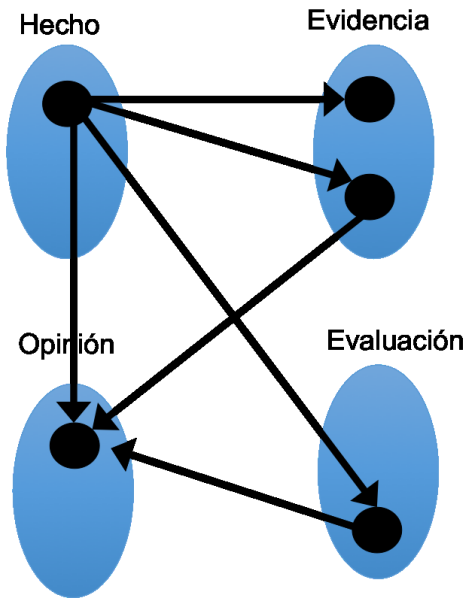


Figura 4.6: Grafo

### 4.3.3. Estructuras semánticas para textos argumentativos

En el caso de los textos argumentativos se propone una estructura de representación, como se muestra en la Figura 4.6, considerando que cuando se realiza una evaluación sobre una opinión, el lector requiere analizar los hechos, buscar a partir de éstos 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 es más complejo para realizar el proceso de comprensión lectora por la interpretación que requiere realizar el lector.

Una vez que se determinan las estructuras semánticas, a partir del cuerpo del pasaje, resulta indispensable caracterizar el otro componente del pasaje que es el bloque de preguntas y respuestas. Para verificar que los pasajes identificados como un tipo de texto en particular corresponda a lo identificado, al analizar esta sección se identificaron patrones cognitivos que permitieron determinar la respuesta correcta ante los cuestionamientos y se mencionan en la siguiente sección.

Tabla 4.3: Tipos de Patrones para textos narrativos

Pregunta	Factor	Respuesta
cuál	hecho, suceso circunstancia, actividad	causas, ventajas, circunstancias, razones, contraste, actividad, agente causas
qué	sucedio, tiempos (first, begin, early), sección, línea, párrafo,	lugares, acciones, razones, línea, párrafo, problemas, actitudes
porque	periodo, líneas, objetos fueron importantes, fue significativo	causas, acciones, contraste justificación, personas, lugar
cuándo	agente, circunstancias	circunstancias, fechas, acontecimiento, periodo
cómo	acción	acción
quién	persona	persona
cuánto	cantidad	cantidad

#### 4.4. Extracción de patrones

En este paso se determinan los patrones a partir del bloque de pregunta-respuesta, para determinar el tipo de inferencia requerida, así como será su aplicación para la determinación de la respuesta correcta.

En la Tabla 4.3 se muestran los patrones detectados en los textos narrativos. La tabla contiene el patrón representado como pregunta (primera pregunta) que depende de un factor (segunda columna) vinculado con las relaciones temporales asociadas a preguntas factuales, por ejemplo, cómo, cuál, qué, por qué, cuándo, cómo y quién que están ubicados en la relevancia de los eventos y de los personajes, cuya respuesta (tercera columna) corresponde al mismo campo semántico.

Tabla 4.4: Tipos de patrones para textos expositivos

Pregunta	Factor	Respuesta
cuál	clase, desarrollo, ejemplo, no es mencionado como elemento, principio describe, forma de, característica, son mencionados como componente	tipos parte todo
qué	se relaciona a, conceptos similares, compara con, hechos, inferido acerca de un componente, tipos, sentencia describe, es parte de	comparación, tipos, temas
incluye todo excepto	exclusión	elemento que no corresponde al parte todo

En la Tabla 4.4 se muestran ahora los patrones detectados de textos expositivos. Como se observa estos patrones están involucrados con propiedades, componentes, meronimia, hiponimia, hiperonimia y comparaciones asociados a preguntas factuales (primera columna) como: cuál, qué, incluye todo excepto, que están ubicados en la relevancia de la descripción de un tema que se representa mediante el factor (segunda columna), cuya respuesta esta relacionada con estas propiedades (tercera columna).

En la Tabla 4.5 se identifican los patrones para textos argumentativos. Los patrones están asociados con las preguntas (primera columna): qué y cómo, cuyos factores (segunda columna) se relacionan con hipótesis, opiniones, sentencias que dan soporte a ideas y estructuras del pasaje que están ubicados en la aceptación o refutación de una idea cuya respuesta (tercera columna) muestra una opinión, demostración o elemento de la estructura solicitadas.

Con los patrones identificados se pueden armar bloques cognitivos para caracterizar el contenido de cada tipo de texto, por lo que a continuación se mencionaran los bloques cognitivos que componen el bloque pregunta y respuesta de los pasajes.

Tabla 4.5: Tipos de Patrones para textos argumentativos

Pregunta	Factor	Respuesta
qué	hipótesis opinión sentencias apoyan se implica se infiere significado de la expresión	demostración opinión
cómo	se organiza	estructura

## 4.5. Bloques cognitivos

Los bloques cognitivos surgen a partir de los patrones identificados en el bloque pregunta respuesta, son elementos que se caracterizan de integrar elementos presentes en la pregunta de acuerdo a la complejidad cognitiva, por ejemplo no es la misma acción cognitiva reconocer una fecha que aplicar una inferencia para una causa y efecto, o deducir una línea del pasaje donde se esté estableciendo una opinión o refutando una argumentación, por tanto cada bloque cognitivo va a corresponder a una complejidad en particular de acuerdo al tipo de pasaje que se esté revisando.

En la revisión de los 300 pasajes del examen TOEFL, se observó que para medir las habilidades de los estudiantes de acuerdo al Marco Común Europeo, es requerido que el estudiante tenga la libertad de determinar qué pregunta se le facilita, por tanto para cada tipo de pasaje se tendrán preguntas muy simples que fácilmente se puede llegar a una inferencia local y posteriormente aumentan su complejidad de acuerdo al tipo de texto.

El examen está balanceado de manera que se pregunta sobre comprensión lectora de los tres tipos de pasaje a partir de cinco lecturas. Por lo que la estrategia propuesta es el análisis de los patrones derivados de las preguntas, para determinar la complejidad cognitiva de acuerdo a cada tipo de texto. Además de que cada tipo de texto cuenta con sus propios bloques.

A continuación se muestra la descripción de los bloques cognitivos correspondientes a cada tipo de texto:

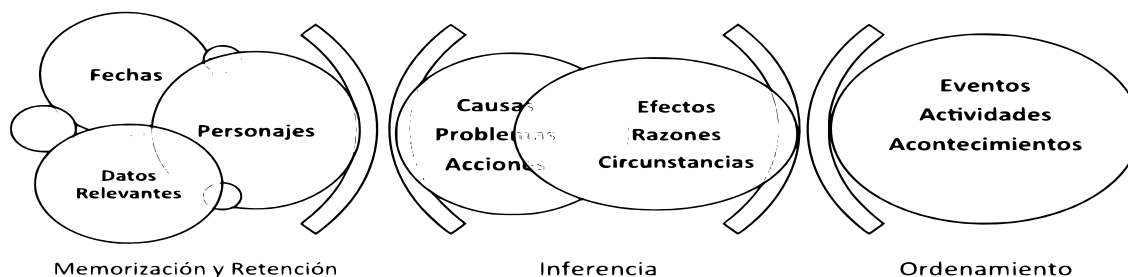


Figura 4.7: Bloques cognitivo de textos narrativos

#### 4.5.1. Bloques cognitivos de textos narrativos

En el caso de los textos narrativos se definieron los siguientes bloques cognitivos:

- memorización y retención de elementos históricos: que se evalúa cuando se solicitan fechas, personajes y datos relevantes y su alcance es solo la línea donde aparece la información para determinar la respuesta, las inferencias manejadas son las del tipo: coherencia, local, tipo de contenido y recursos de información que aparecen en la Tabla 2.2.
- inferencia: donde a partir de elementos históricos se identifican causas, problemas, acciones, efectos, razones y circunstancias en periodos de tiempo, debido a que requiere establecer el párrafo donde se encuentra la respuesta, si es que en la propia línea no existe información suficiente para deducirla; en este caso el tipo de inferencias relacionadas son la de dirección(hacia atrás, hacia adelante).
- ordenamiento: que se encuentra presente a partir de la identificación de eventos, actividades y acontecimientos durante todo el pasaje, en este caso, corresponde al nivel más complejo, considerando que requiere inferir la respuesta a partir de todos los párrafos del texto; las inferencias relacionadas al bloque son las de grado de probabilidad (lógicas) y recursos cognitivos(estratégicas).

#### 4.5.2. Bloques cognitivos de textos expositivos

En el caso de los textos expositivos se definieron los siguientes bloques cognitivos:

- memorización y retención: considerando que en el momento de la lectura se debe de identificar el elemento principal y sus características, en este se parte de lo que es un tema y los hechos relacionados.

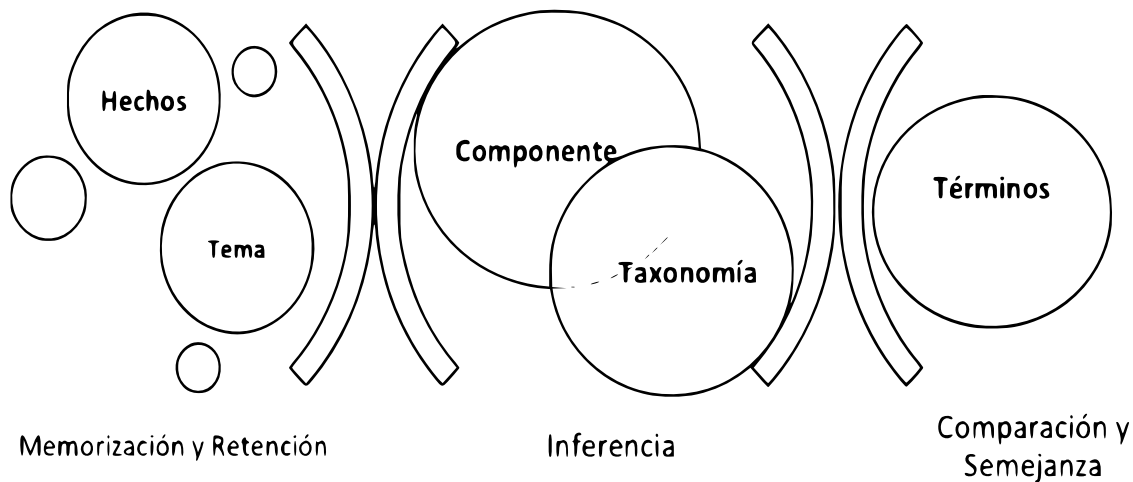


Figura 4.8: Bloques cognitivo de textos expositivo

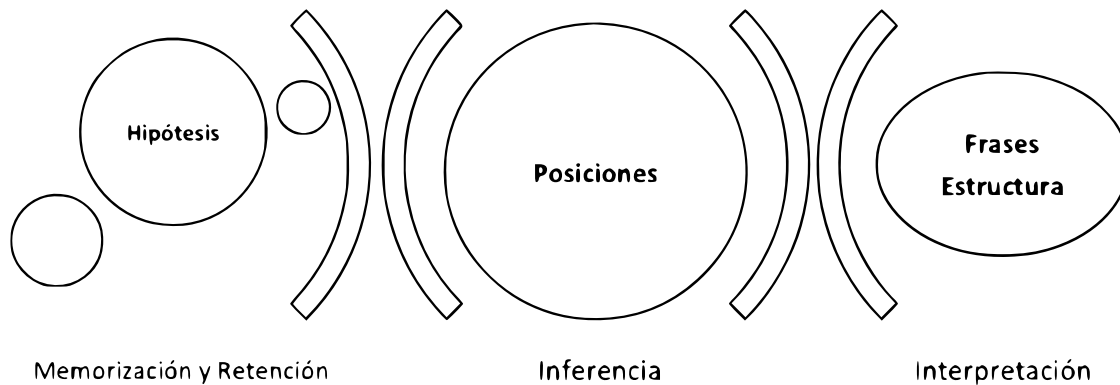


Figura 4.9: Bloques cognitivo de textos narrativos

- inferencia: se relaciona con la taxonomía entre conceptos y los componentes que forman parte de los mismos.
- comparación y semejanza, es el más complejo puesto que requiere tomar en cuenta el tema, los hechos, las clases, los componentes para generar una comparativa sobre las propiedades, elementos, procesos relacionados con dos términos o más.

#### 4.5.3. Bloques cognitivos de textos argumentativos

En el caso de los textos argumentativos se definieron los siguientes bloques cognitivos:

- memorización y retención, donde se identifican las hipótesis que dan pie al texto.

- inferencia, donde se determinan las posiciones a favor o en contra sobre las hipótesis mencionadas
- interpretación requiere de las hipótesis, los hechos, las opiniones y las evidencias, para expresar de otra forma el contenido de acuerdo al contexto y en el caso más complejo definir una estructura a partir de los elementos planteados en la argumentación.

Por lo que surge la idea de plantear una arquitectura a partir de los bloques cognitivos que definen el tipo de texto que se está evaluando, las estructuras semánticas que conforman la riqueza de términos para hacer consultas y así aplicar el mecanismo de inferencia para identificar la respuesta correcta, por lo que en la siguiente sección se muestra su construcción.

## 4.6. Construcción de la arquitectura

A continuación en la Figura 4.10 se muestra la arquitectura general propuesta cuyas fases se mencionarán a continuación.

El modelo semántico planteado se compone de tres fases que forman una arquitectura general que surgió a partir de la necesidad de buscar una solución específica para la respuesta correcta mediante los sistemas basados en reglas. A continuación se describe a cada una de ellas.

## 4.7. Fases del modelo

En la Fase 1 de la Figura 4.10 se determinan los bloques cognitivos que caracterizan a partir de las preguntas el tipo de pasaje que corresponda, en algunos casos el pasaje puede tener información particular de los tres tipos de texto, por lo que es importante determinar, cuál es el tipo de pasaje sobre el cual se van a determinar las respuestas correctas a partir de los elementos semánticos.

Por lo que se requiere en esta fase que exista al menos un bloque cognitivo de cada elemento presente en la pregunta, de manera que si al menos las preguntas están correspondiendo a los bloques cognitivos, por lo tanto podemos validar que el pasaje corresponda al tipo de texto.

Para caracterizar el pasaje usamos el análisis de la pregunta (análisis de abajo hacia arriba), para así determinar los bloques cognitivos y de ahí enlazamos con lo que sería

el pasaje, empleamos este enfoque porque un lector inexperto hace lo contrario (análisis de arriba hacia abajo) por lo que podría clasificar el texto del pasaje en otra categoría que sea incorrecta.

En el caso del enfoque planteado estamos considerando que las preguntas están orientadas al tipo de pasaje que corresponda, por tanto los bloques cognitivos se consideran excluyentes de acuerdo al tipo de texto, por lo que al clasificar el tipo de texto se puede entrar a la Fase 2.

En el caso de la Fase 2 de la Figura 4.10 se procesa el pasaje a partir de las relaciones identificadas para establecer la semántica, esto ha generado un conjunto de estructuras semánticas propias para cada tipo de pasaje y por tanto para inferir la respuesta correcta debemos tener en el contexto los patrones que caracterizan cada tipo de texto, para así pasar a la Fase 3.

Las estructuras semánticas están construidas a partir de los elementos presentes en cada tipo de pasaje, si bien cómo se comentaba anteriormente pudiera darse el caso que hay algún elemento, por ejemplo dado un texto expositivo donde se mencione un elemento de un texto narrativo, en ese caso hay menor incidencia de sus elementos y como la fase 1 validó el tipo de texto, lo siguiente es trabajar los patrones respecto a las estructuras semánticas y la pregunta basada en el contexto. En el caso de Fase 3 de la Figura 4.10 se aplica un operador para determinar la respuesta correcta, cabe mencionar que en el corpus de pasajes se tiene la batería de respuestas, en este caso también se asume que cada pregunta de un pasaje siempre tiene cuatro posibles respuestas de las cuales, una siempre es la correcta, por tanto en esta fase se va a revisar cuál es la respuesta, más cercana al planteamiento de la pregunta a partir de su contexto y de las estructuras semánticas propias del pasaje.

## 4.8. Estrategia de solución

En este proceso se analiza la formalización de la estructura resultante de la arquitectura, para determinar la respuesta correcta de forma determinista.

En este enfoque se emplean reglas lógicas, planteando que a partir de las preguntas y los patrones encontrados se determine que sólo una respuesta es correcta, porque sólo existe una ruta que haga válido una de las respuestas.

A continuación se muestran un solución planteada para definir la respuesta correcta.

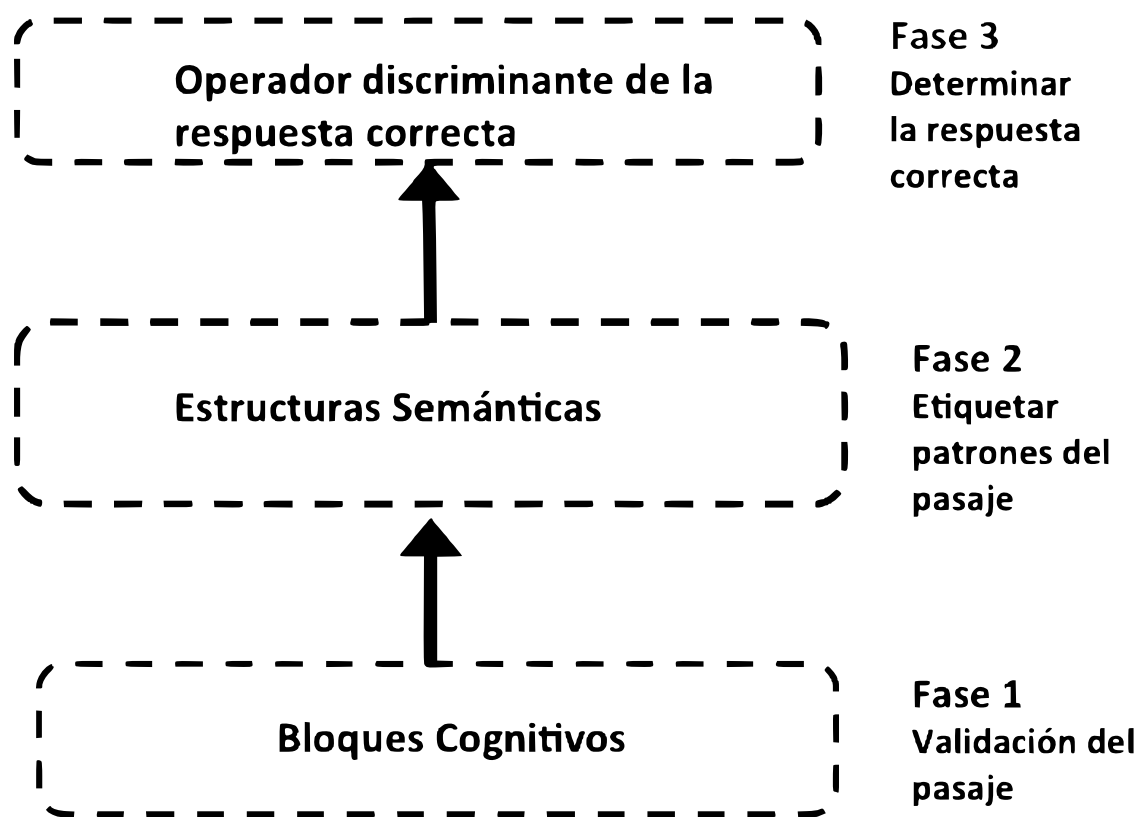


Figura 4.10: Arquitectura General del Modelo Semántico

#### 4.8.1. Solución basada en reglas

Cuando se modela conocimiento se requieren analizar las propiedades y relaciones en un dominio específico, existen varios mecanismos para representar conocimiento, uno de ellos son las reglas, que son una forma intuitiva de representar esquemas mediante una estructura que facilite la determinación a partir de una premisa y una conclusión. En el caso de los cuestionamientos de pregunta respuesta que se aplica en exámenes de acreditación tipo TOEFL, se requiere analizar qué patrones están vinculados con la pregunta y relacionados con la respuesta, para que a partir de este esquema y la información complementaria provista por los documentos, se pueda dar una solución acertada.

La ventaja de la representación por medio de reglas es que arroja un resultado determinista que otros esquemas de representación en su lugar hacen un manejo probabilístico. En esta sección se propone mediante reglas inferenciales el modelado de patrones para dar respuesta a cuestionamientos sobre textos académicos de tipo TOEFL.

Un sistema experto puede definirse como un programa inteligente que utiliza conocimiento y procedimientos de inferencia para resolver problemas que son lo bastante complejos como para requerir expertos humanos en su resolución. Los problemas que se resuelven pueden ser clasificados según su naturaleza, en deterministas ó estocásticos Castillo (1997).

El primero grupo está conformado por aquellos problemas que se pueden plantear utilizando un conjunto de reglas mientras el segundo grupo manejan situaciones inciertas y necesitan incorporar la incertidumbre que es la probabilidad, en nuestro caso abordamos el enfoque basado en reglas considerando que se requiere indicar la respuesta correcta de forma determinística Castillo (1997).

Las reglas en un sistema inteligente son consideradas una representación determinística del conocimiento, a continuación en la Figura 4.11 se muestra el esquema general de representación de reglas Castillo (1997). Esta representación es usada considerando que es muy cercana al sentido común de los seres humanos, cómo se puede observar en el esquema inicial tenemos un antecedente y un consecuente o bien una premisa y una conclusión; en este caso se tiene un grupo de condiciones denotados por  $CG_i$  que están dentro del bloque condicional y en el resultado se da una o se dice si es aceptado o rechazado.

Dentro de los  $CG_i$  se tienen un conjunto de reglas que se van aplicando y dependiendo de esa aplicación se van descartando las posibles respuestas, dado que tiene como operador

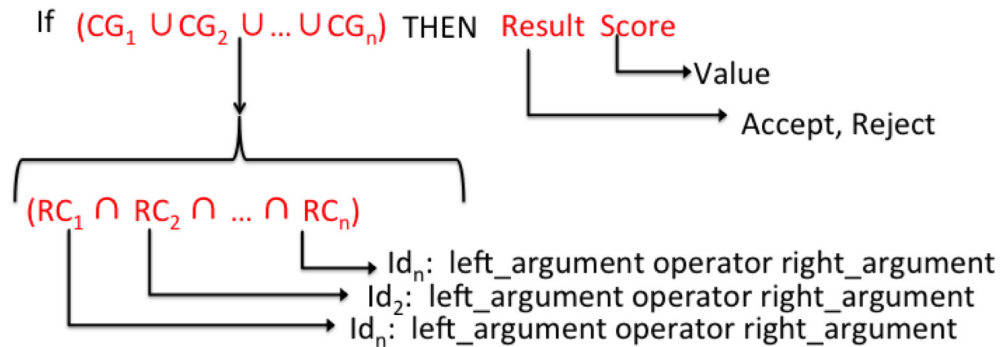


Figura 4.11: Ejemplo de especificación de reglas Castillo (1997).

principal el *and*, si alguna de las reglas se inválida entonces en automático ya es inválido ese grupo de condiciones, entonces como el condicional tiene el operador *or* se requiere que al menos algunas de estos grupos de reglas condicionales sea válido para que se pueda dar el resultado. Las condiciones de reglas  $RC_i$  se caracterizan por definirse con un argumento a la izquierda, un operador y un argumento a la derecha. Esto implica considerando las gramáticas de Tesniere que tendremos en esta notación una valencia de dos, considerando que tenemos dos argumentos.

Considerando estos elementos se determinó que para caracterizar un tipo de texto específico, y evitar ambigüedad en el caso de que un lector tuviera duda si es un pasaje explicativo narrativo o argumentativo, se requería la validación por parte de los bloques cognitivos para determinar el tipo de texto que se estaba abordando.

Una vez que se hace la verificación de que el texto corresponde de acuerdo a los bloques cognitivos, entonces se modela el bloque de pregunta de manera que se desmenucé en dos elementos: el contexto y el patrón.

En nuestro caso hemos hecho una adaptación de esta regla considerando los bloques cognitivos que caracterizan el bloque de pregunta respuesta de cada tipo de texto.

Cómo se observa en la Figura 4.12 se realiza una adaptación de los  $CG_i$  para aplicar el primer paso de la metodología, que es la determinación del tipo de texto, para eso en la sección de la premisa se ha colocado la validación de cada uno de los bloques cognitivos de acuerdo al tipo de texto; el operador que los conecta es un *or* exclusivo considerando que los pasajes son sólo de un tipo, por tanto sólo uno de los bloques cognitivos arrojará un resultado como verdadero y los demás serán falsos.

En el caso del tipo de texto que puede arrojar corresponde a los definidos anteriormente como expositivo, narrativo o argumentativo, también en este caso el resultado siempre va a dar el resultado de aceptado porque está categorizado en este tipo de textos que

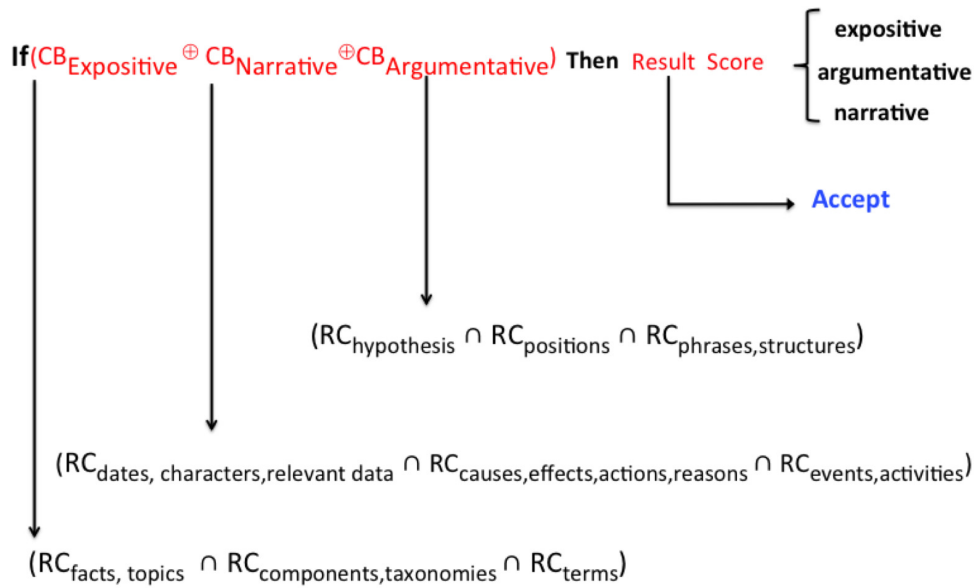


Figura 4.12: Ejemplo de especificación de reglas de bloques Cognitivos

siempre va a corresponder alguno de ellos. Se observa que cada uno de los bloques cognitivos tienen sus condiciones de regla  $RC_i$  de acuerdo a la definición de los bloques cognitivos.

En la sección inferior de la Figura 4.12 se muestran las condiciones para los tres tipos de textos, de acuerdo al que aplique. En el caso de los textos expositivos poseen las condiciones de regla de hechos y temas, también deben de estar presentes en las condiciones de reglas los componentes y las taxonomías y de la misma forma los términos, es decir debe de existir al menos una ocurrencia de cada uno de los elementos del bloque para que pueda corresponder a los textos expositivos.

En el caso de el bloque narrativo se requiere que tres condiciones de regla estén con valor verdadero, el primero corresponde a la presencia de fechas personajes o datos relevantes, también es requerido que exista evidencia de causas, efectos, acciones o razones y también de eventos y actividades que caracterizan a este tipo de texto.

Respecto los textos argumentativos se requiere de condiciones de regla que exista una hipótesis, las posiciones a favor o en contra y frases o estructuras.

Una vez que se ha validado el bloque cognitivo al cual pertenece el texto y retorne el tipo de texto encontrado, lo que se busca es la validación de la respuesta correcta a partir del contexto presentado por la pregunta, como se muestra en la Figura 4.13. Así se tendrán un grupo de condiciones con un bloque condicional relacionado con una respuesta en particular, en este caso se manejan cuatro posibles respuestas y lo que se

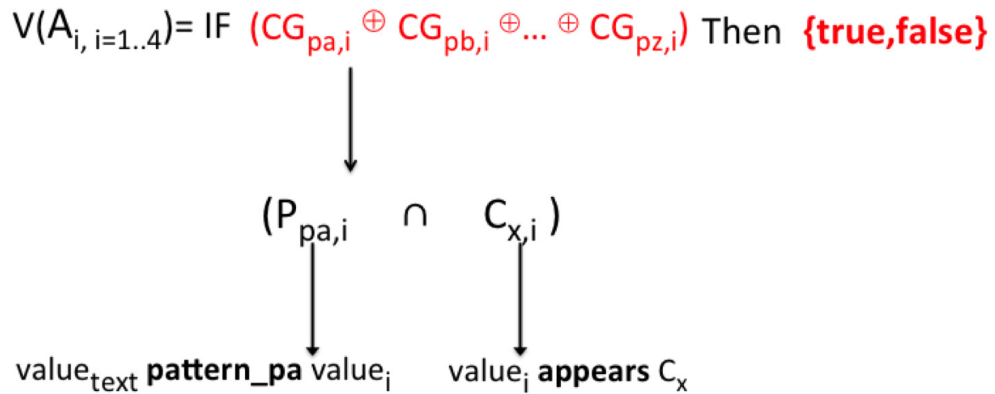


Figura 4.13: Ejemplo de especificación de reglas de respuesta

busca es que arroje un resultado verdadero o falso.

Así la respuesta correcta será la única que tenga el valor verdadero y los demás distractores requieren tener el valor falso, no hay resultados parcialmente correctos dada esta regla determinista.

En el caso de la premisa lo que se trabaja es un grupo de condiciones de acuerdo a un patrón que coincida con la opción de respuesta, por esa causa la notación se considera como el patrón *pa* de la *pa* a la *pz* dependiendo cuantos patrones se tengan, para una respuesta *i* específica de las cuatro posibles, como se observa en la Figura 4.13 .

En cada grupo de condición se verifican dos aspectos que deben existir al mismo tiempo: el primero es determinar si el valor de la respuesta corresponde a uno de los patrones que se esté fijando respecto al texto que aparece en el pasaje y también se debe cumplir que la respuesta o el distractor esté relacionado con el contexto. Esta regla se establece de esta forma porque puede darse el caso de que el distractor corresponda al campo semántico. Sin embargo no haya evidencia de dicho contenido en el texto del pasaje lo que está relacionado con el contexto de la pregunta, por tanto si no existe en el contexto puede descartarse.

Al hacer la verificación en la premisa solamente uno de los patrones será válido, por tanto sólo una vez el patrón corresponderá al valor que corresponda al campo semántico que coincida con la definición que esté dentro del pasaje, una vez que se aplica el operador lógico and entonces podemos decidir si el patrón está caracterizando a la respuesta.

A continuación se muestra en el Algoritmo 2 el procedimiento para deducir la respuesta correcta. En este caso, la entrada es un pasaje tipo TOEFL en primera instancia y como salida las respuestas correctas. Es necesario manejar la base de conocimiento de

los pasajes a una transformación de predicado para trabajar las reglas para tener un argumento a la izquierda, una propiedad de la base de conocimiento, y un argumento a la derecha. Las preguntas se procesan para aplicar la regla del bloque cognitivo. Así, indican el tipo de texto que corresponderá mediante una selección múltiple, centrándose en los elementos que tiene el bloque pregunta-respuesta en qué tipo de texto corresponderá. Si es narrativo, expositivo argumentativo, como se menciona en las reglas, entonces para cada una de las preguntas, tenemos las cuatro respuestas.

Para cada respuesta, se aplica la regla interna del patrón y el contexto. Aquí cabe mencionar que estos patrones están precisamente relacionados con las relaciones semánticas, una vez que hemos localizado de qué tipo de patrón se trata, comprobamos que también corresponda con el contexto presente en el pasaje. Entonces la respuesta bajo esta condiciones es la correcta; todas las demás, en particular, no se vincularán a un patrón o contexto particular.

Considerando la construcción del modelo descrita a continuación en el siguiente capítulo se mostrarán ejemplos de la aplicación del funcionamiento del modelo y sus resultados relacionados.

---

**Algoritmo 2** Algoritmo determinista para inferir las respuestas correctas en textos tipo TOEFL
 

---

Entrada: Pasaje TOEFL

Salida: respuestas

- 1: Generar la base de conocimiento a partir del pasaje y transformarlo a predicados de la forma *argumento\_izquierdo* propiedadKB *argumento\_derecho*.
  - 2: Procesar las preguntas a predicados de la forma *argumento\_izquierdo* propiedadKB *argumento\_derecho* para obtener el bloque cognitivo .
  - 3: Aplicar la regla para determinar el tipo de texto a partir de  $Q_z$ Propiedad
  - 4: switch ( $Q_z$ Propiedades):
  - 5: case ( $RC_{hechos,temas} \cap RC_{componentes,taxonomias} \cap RC_{terminos}$ ):
  - 6: return Expositivo
  - 7: case: ( $RC_{fechas,personajes,relevantdata} \cap RC_{causas,efectos,acciones,razones} \cap RC_{eventos,actividades}$ ):
  - 8: return Narrativo
  - 9: case: ( $RC_{hipotesis} \cap RC_{posiciones} \cap RC_{frases,estructuras}$ ):
  - 10: return Argumentativo
  - 11: for cada pregunta  $Q_z$ ,  $z = 1 \dots n$  do
  - 12:   for cada respuesta,  $A_i$  of  $Q_z$  do
  - 13:     Aplicar la regla de acuerdo al patrón y al contexto.
  - 14:     ban=false
  - 15:     switch( $A_i$ ):
  - 16:       case ( $p_{a,i}(A_i)$ ): ban=true
  - 17:       case ( $p_{b,i}(A_i)$ ):ban=true
  - 18:       ...
  - 19:       case ( $p_{z,i}(A_i)$ ):ban=true
  - 20:       if (*ban* and  $C_{x,i}$ ) then
  - 21:          return "  $A_i$  es la respuesta a la pregunta  $Q_z$ ."
  - 22:       end if
  - 23:   end for
  - 24: end for
-



# Capítulo 5

## Resultados

En este capítulo se mostrarán los resultados que se generaron durante la elaboración del modelo, enfocándose en primera instancia en un etiquetado mediante xsd y xml para identificar las secciones del pasaje y las relaciones semánticas, posteriormente se trabajó con una representación en lógica de predicados aplicándose las reglas de bloques cognitivos y de respuesta.

### 5.1. Etiquetado

Uno de los resultados obtenidos fue el etiquetado del corpus recuperado de Chuvanan-University (2018). El total de pasajes incluidos en este corpus es de 300, de los cuáles 140 son narrativos, 20 argumentativos y 140 expositivos.

El etiquetado fue llevado por 2 personas certificadas en TOEFL nivel C1 y C2. Posteriormente, el almacenamiento de este corpus etiquetado fue realizado en un archivo xsd, en las Figuras 5.1 y 5.2 se muestra la estructura, donde se definió para cada pasaje su ubicación, tipo de documento, tema, área, párrafos, relaciones semánticas, contexto, y sus aserciones.

Posteriormente esta estructura es llenada, en las Figuras 5.3, 5.4, 5.5, 5.6 y 5.7 se muestra un ejemplo de un texto expositivo, donde se coloca la descripción del pasaje, el área, se separan los párrafos, sus sentencias y las relaciones semánticas dentro de las mismas, (ver Figuras 5.3, 5.4, 5.5). En las Figuras 5.5, 5.6 y 5.7 aparece la definición del bloque pregunta-respuesta donde se representan también las relaciones semánticas en la respuesta.

Las relaciones semánticas encontradas se definieron en el capítulo de modelo semántico, por lo que esta representación sirvió de base para la generación de las estructuras

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<xs:schema xmlns:xs=http://www.w3.org/2001/XMLSchema elementFormDefault="qualified">
  <xs:element name="passage">
    <xs:complexType><xs:sequence>
      <xs:element name="description" minOccurs="1" maxOccurs="1">
        <xs:complexType><xs:sequence><xs:element name="idPassage" type="xs:integer"/>
          <xs:element name="DocumentType"><xs:simpleType>
            <xs:restriction base="xs:string"><xs:enumeration value="Expository" />
              <xs:enumeration value="Narrative" /><xs:enumeration value="Argumentative" />
            </xs:restriction></xs:simpleType></xs:element>
            <xs:element name="area" type="xs:string" />
            <xs:element name="topic" type="xs:string" /></xs:sequence>
          </xs:complexType></xs:element>
      <xs:element name="paragraph" minOccurs="1" maxOccurs="unbounded"/>
        <xs:complexType><xs:sequence>
          <xs:element name="idParagraph" type="xs:integer"/>
          <xs:element name="paragraph" type="xs:string">
            <xs:complexType><xs:sequence>
              <xs:element name="sentence" minOccurs="1" maxOccurs="unbounded">
                <xs:complexType><xs:sequence>
                  <xs:element name="SemanticRelation"><xs:simpleType>
                    <xs:restriction base="xs:string">
                      <xs:enumeration value="ClassInclusion/Hyponim" />
                      <xs:enumeration value="ClassInclusion/Hyperonim" />
                      <xs:enumeration value="PartWhole/ObjectComponent" />
                      <xs:enumeration value="PartWhole/CollectionMember" />
                      <xs:enumeration value="PartWhole/CollectionMember" />
                      <xs:enumeration value="PartWhole/ActivityStage" />
                      <xs:enumeration value="Similar/Synonym" />
                      <xs:enumeration value="CausePurpose/CauseEffect" />
                      <xs:enumeration value="CausePurpose/AgentGoal" />
                      <xs:enumeration value="CausePurpose/InstrumentGoal" />
                      <xs:enumeration value="CaseRelation/AgentInstrument" />
                      <xs:enumeration value="SpaceTime/LocationAction" />
                    </xs:restriction>
                  </xs:element>
                </xs:sequence>
              </xs:complexType>
            </xs:sequence>
          </xs:element>
        </xs:complexType>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>

```

Figura 5.1: Representación XSD de los pasajes. Primera parte.



```

Input XML
1 <?xml version="1.0" encoding="utf-8"?>
2 <passage>
3 <description><idPassage>41</idPassage>
4 <documentType>expository</documentType>
5 <area>biology</area>
6 <topic>bacterias</topic>
7 </description>
8 <paragraphs>
9 <paragraph>
10 <id_paragraph>1</id_paragraph>
11 <paragraphText>Bacteria are extremely small living things. While we measure our own sizes in inches or centimeters,
    bacterial size is measured in microns. One micron is a thousandth of a millimeter a pinhead is about a millimeter
    across. Rod shaped bacteria are usually from two to four microns long, while rounded ones are generally one micron
    in diameter. Thus if you enlarged a founded bacterium a thousand times, it would be just about the size of a pinhead
    . An adult human magnified by the same amount would be over a mile(1.6 kilometers) tall.
12 </paragraphText>
13 <sentences>
14 <sentence>
15 <sentenceText>Bacteria are extremely small living things.</sentenceText>
16 <semanticRelation>ClassInclusion/Hyponym</semanticRelation>
17 <context>no</context>
18 <assertion>hyponym(living things,bacterias)</assertion>
19 </sentence>
20 </sentences>
21 <sentenceText>Rod shaped bacteria are usually from two to four microns long, while rounded ones are generally one micron in
    diameter. </sentenceText>
22 <semanticRelation>ClassInclusion/Hyponym</semanticRelation>

```

Figura 5.3: Ejemplo de uso de la representación en xml de un texto expositivo sobre las bacterias. Primera parte.

```

22 <semanticRelation>ClassInclusion/Hyponym</semanticRelation>
23 <context>no</context>
24 <assertion>hyponym(bacterias,rodShapedBacteria)</assertion>
25 </sentence>
26 </sentences>
27 </paragraph>
28 <paragraph><id_paragraph>2</id_paragraph>
29 <paragraphText> Even with an ordinary microscope, you must look closely to see bacteria. Using a magnification of 100 times,
    one finds that bacteria are barely visible as tiny rods or dots. One cannot make out anything of their structure. Using
    special stains, one can see that some. Others have only one flagellum. The flagella rotate, pushing the bacteria though
    the water. Many bacteria lack flagella and cannot move about by their own power while others can glide along over
    surfaces by some little understood mechanism.
30 </paragraphText>
31 <sentences>
32 <sentence>
33 <sentenceText> Using a magnification of 100 times, one finds that bacteria are barely visible as tiny rods or dots
    .</sentenceText>
34 <semanticRelation>Similar/Synonym</semanticRelation>
35 <context>no</context>
36 <assertion>synonym(bacteria, tinyRods)</assertion>
37 <assertion>synonym(bacteria, dots)</assertion>
38 </sentence>
39 </sentences>
40 </paragraph>
41 <paragraph><id_paragraph>3</id_paragraph>
42 <paragraphText> From the bacterial point of view, the world is a very different place from what it is to humans. To a
    bacterium water is as thick as molasses is to us. Bacteria are so small that they are influenced by the movements of the

```

Figura 5.4: Representación en xml del texto expositivo de las bacterias. Segunda parte

```

42 - <paragraphText> From the bacterial point of view, the world is a very different place from what it is to humans. To a
    bacterium water is as thick as molasses is to us. Bacteria are so small that they are influenced by the movements of the
    chemical molecules around them. Bacteria under the microscope, even those with no flagella, often bounce about in the
    water. This is because they collide with the water molecules and are pushed this way and that. Molecules move so rapidly
    that within a tenth of a second the molecules around a bacterium have all been replaced by new ones even bacteria
    without flagella are thus constantly exposed to a changing environment.
43 - </paragraphText>
44 - <sentences>
45 - </sentences>
46 - </paragraph>
47 - </paragraphs>
48 - <questionAnswerBlock>
49 - <questions>
50 - <question>
51 - <questionText>According to the passage, someone who examines bacteria using only a microscope that magnifies 100 times would
    see </questionText>
52 - <answers>
53 - <answer>
54 - <answerText>tiny dots</answerText>
55 - <semanticRelation>Similar/Synonym</semanticRelation>
56 - <context>no</context>
57 - <assertion>synonym(bacterias, tinyDots)</assertion>
58 - </answer>
59 - <answer>
60 - <answerText>small "hairs"</answerText>
61 - <semanticRelation>Similar/Synonym</semanticRelation>
62 - <context>no</context>

```

Figura 5.5: Tercera parte de la representación del texto expositivo incluyendo el bloque de pregunta-respuesta

```

62 - <context>no</context>
63 - <assertion>no</assertion>
64 - </answer>
65 - <answer>
66 - <answerText>large rods</answerText>
67 - <semanticRelation>Similar/Synonym</semanticRelation>
68 - <context>no</context>
69 - <assertion>no</assertion>
70 - </answer>
71 - <answer>
72 - <answerText>detailed structures</answerText>
73 - <semanticRelation>Similar/Synonym</semanticRelation>
74 - <context>no</context>
75 - <assertion>no</assertion>
76 - </answer>
77 - </answers>
78 - </question>
79 - <question>
80 - <questionText>The relationship between a bacterium and its flagella is most nearly analogous to which of the following?
81 - </questionText>
82 - <answers>
83 - <answer>
84 - <answerText> A rider jumping on a horse's back</answerText>
85 - <semanticRelation>Similar/Synonym</semanticRelation>
86 - <context>no</context>
87 - <assertion>no</assertion>
88 - </answer>

```

Figura 5.6: Cuarta parte de la representación XML del texto expositivo relacionado con el bloque de pregunta-respuesta

```

88 </answer>
89 <answer>
90 <answerText>A ball being hit by a bat</answerText>
91 <semanticRelation>Similar/Synonym</semanticRelation>
92 <context>no</context>
93 <assertion>no</assertion>
94 </answer>
95 <answer>
96 <answerText> A boat powered by a motor</answerText>
97 <semanticRelation>Similar/Synonym</semanticRelation>
98 <context>movement</context>
99 <assertion>synonym(bacteria,boat)</assertion>
100 <assertion>synonym(flagella,motor)</assertion>
101 </answer>
102 <answer>
103 <answerText>A door closed by a gust of wind</answerText>
104 <semanticRelation>Similar/Synonym</semanticRelation>
105 <context>component</context>
106 <assertion>no</assertion>
107 </answer>
108 </answers>
109 </question>
110 </questions>
111 </questionAnswerBlock>
112 </passage>

```

Figura 5.7: Quinta y última parte de la representación xml del texto expositivo

semántica que aparecen en el capítulo mencionado.

Sin embargo, como el objetivo de la propuesta era desarrollar desde cero el modelo semántico se optó por trabajar en la representación de predicados, por lo que posteriormente puede emplearse ésta en otros trabajos.

## 5.2. Casos de Estudio

A partir de la propuesta por medio de reglas inferenciales para la discriminación de la respuesta correcta, se realizaron pruebas para determinar si el Algoritmo 2 contemplaba la solución para todos los tipos de pasajes, a continuación se muestra un ejemplo de cada caso.

### 5.2.1. Caso: textos expositivos

En este caso, en la Figura 5.8 se tiene un ejemplo de un texto expositivo, cuyo tema principal son los virus. Aplicando el Algoritmo 2, se procesa tanto el cuerpo como el bloque de pregunta respuesta del pasaje (pasos 1 y 2).

Después se aplica el paso 3 del Algoritmo 2, que internamente funciona como se muestra el proceso en la Figura 5.9; por ejemplo se toman del texto de la Figura 5.8, tres preguntas para validar los bloques (paso 4 del Algoritmo 2), *The author implies that bacteria were investigated earlier than viruses because* (pregunta 6), resultando el predicado manejado en el paso 2 del Algoritmo 2, de la forma *argumento izquierdo*

*propiedadKB argumento derecho*, en este caso *wereInvestigatedBacterias belongs fact*, corresponde a lo que sería la determinación de un hecho, porque está dando una pauta de que se realizó una investigación, por tanto corresponde a la verificación de la condición de reglas de tema y hechos (paso 5 del algoritmo).

Posteriormente se encuentra la pregunta 5: *All the following may be components of a virus EXCEPT?*, que corresponde a los componentes, en este caso, a partir de los componentes cuál no pertenece a los virus, cuyo predicado generado en el paso 2 del Algoritmo 2, es *parts notBelongs virus*. Además en este caso, aparece un operador negativo, el *not*, pero en realidad lo que se está evaluando es la relación de pertenencia, por tanto también está presente esta regla de componentes y taxonomías como se observa en la Figura 5.9, aplicando el paso 5 del Algoritmo 2.

Finalmente se encuentra una de las preguntas que hace comparación entre términos, la pregunta 3: *The word proven in line 4 is closest meaning to which of the following*, en este caso se busca el significado más cercano del verbo *proven*, por lo que implica la comparación de acuerdo a la sinonimia con otros términos que pudieran aparecer, cuyo predicado generado en el paso 2 del Algoritmo 2, es *proven is closestMeaning*, en este caso se cumplen los tres bloques cognitivos de los textos expositivos de acuerdo a la regla de la Figura 5.9.

Es decir en el Algoritmo 2, se han ejecutado los pasos: 4 y 5 para seleccionar el bloque y el 6 que retorna al cumplirse la regla que el texto es expositivo. Los pasos del 7 al 10 del Algoritmo 2, no se ejecutan por ser un *switch* la estructura condicional empleada. Internamente la regla de los bloques cognitivos (ver Figura 5.9) evaluará los bloques requeridos para el texto expositivo  $CB_{Expositive}$  como se mencionó anteriormente, al estar las condiciones de regla conectadas por el operador *and* y cumplirse los tres bloques retornará verdadero en  $CB_{Expositive}$ , por tanto la regla es verdadera.

En el caso de los  $CB$  para textos narrativos y argumentativos que se conectan con un *or* exclusivo son falsos, porque no hay elementos de sus bloques, por tanto al ser sólo válida la información del bloque expositivo, los demás  $CB$  son falsos. Así, la regla arroja que es un texto expositivo y que por tanto es aceptada.

Posteriormente se ejecuta el paso 11 del Algoritmo 2, cuyo propósito es generar las respuestas correctas para cada pregunta. En este caso se seleccionó la pregunta 5 del texto de la Figura 5.8, para ejemplificar el funcionamiento.

En el paso 12 del Algoritmo 2, se validan las respuestas para la pregunta *All the following may be components of a virus EXCEPT?*(Todos los siguientes son componentes de un virus excepto?), que corresponde a excepto, que es una negación. Por lo que se ejecuta el paso 13 de este algoritmo, que aplica de forma interna la regla

## Topic: Virus

The term 'virus' is derived from the Latin word for poison, or slime. It was originally applied to the noxious stench emanating from swamps that was thought to cause a variety of diseases in the centuries before microbes were discovered and specifically linked to illness. But it was not until almost the end of the nineteenth century that a true virus was proven to be the cause of a disease.

The nature of viruses made them impossible to detect for many years even after bacteria had been discovered and studied. Not only are viruses too small to be seen with a light microscope, they also cannot be detected through their biological activity, except as it occurs in conjunction with other organisms. In fact, viruses show no traces of biological activity by themselves. Unlike bacteria, they are not living agents in the strictest sense. Viruses are very simple pieces of organic material composed only of nucleic acid, either DNA or RNA, enclosed in a coat of protein made up of simple structural units. (Some viruses also contain carbohydrates and lipids.) They are parasites, requiring human, animal, or plant cells to live. The virus replicates by attaching to a cell and injecting its nucleic acid. Once inside the cell, the DNA or RNA that contains the virus' genetic information takes over the cell's biological machinery, and the cell begins to manufacture viral proteins rather than its own.

1. Which of the following is the best title for the passage.

(A) New Developments in Viral Research (B) Exploring the Causes of Disease (C) DNA: Nature's Building Block (D) Understanding Viruses

2. Before microbes were discovered It was believed that some diseases were caused by

(A) germ-carrying insects (B) certain strains of bacteria (C) foul odors released from swamps (D) slimy creatures living near swamps

3. The word "proven" in line 4 is closest meaning to which of the following.

(A) Shown (B) Feared (C) Imagined (D) Considered

4. The word "nature" in line 6 is closest in meaning to which of the following?

(A) Self-sufficiency (B) Shapes (C) Characteristics (D) Speed

5. All of the following may be components of a virus EXCEPT

(A) RNA (B) plant cells (C) carbohydrates (D) a coat of protein

6. The author implies that bacteria were investigated earlier than viruses because

(A) bacteria are easier to detect (B) bacteria are harder to eradicate (C) viruses are extremely poisonous (D) viruses are found only in hot climates

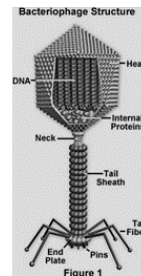


Figura 5.8: Ejemplo de texto expositivo, fuente Universidad de Chuvanan Chuvanan-University (2018)

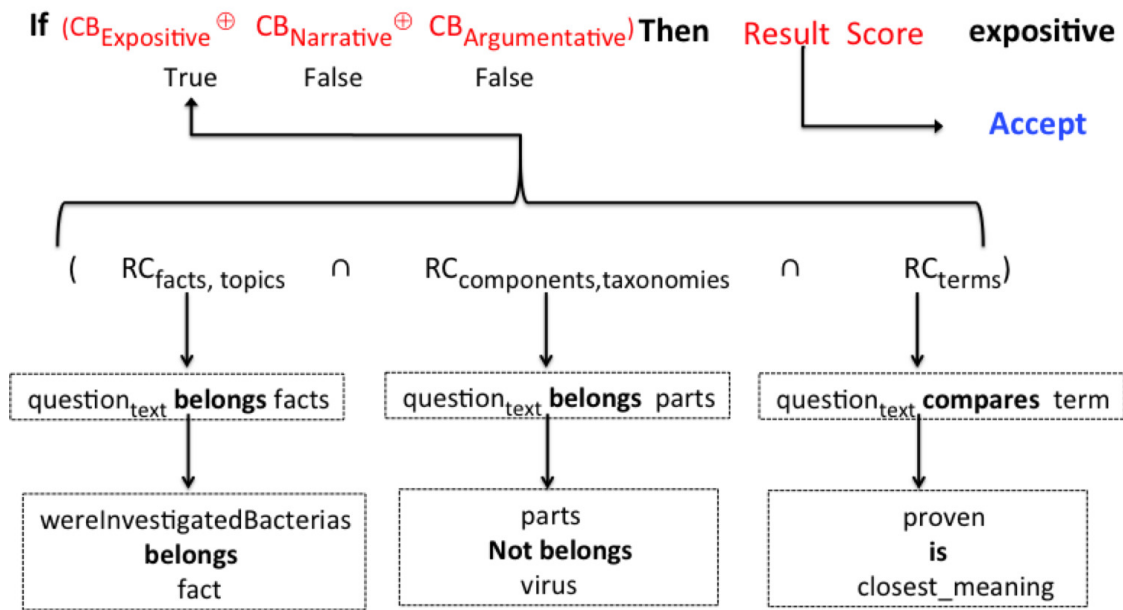


Figura 5.9: Aplicación de la regla de bloques cognitivos para textos expositivos

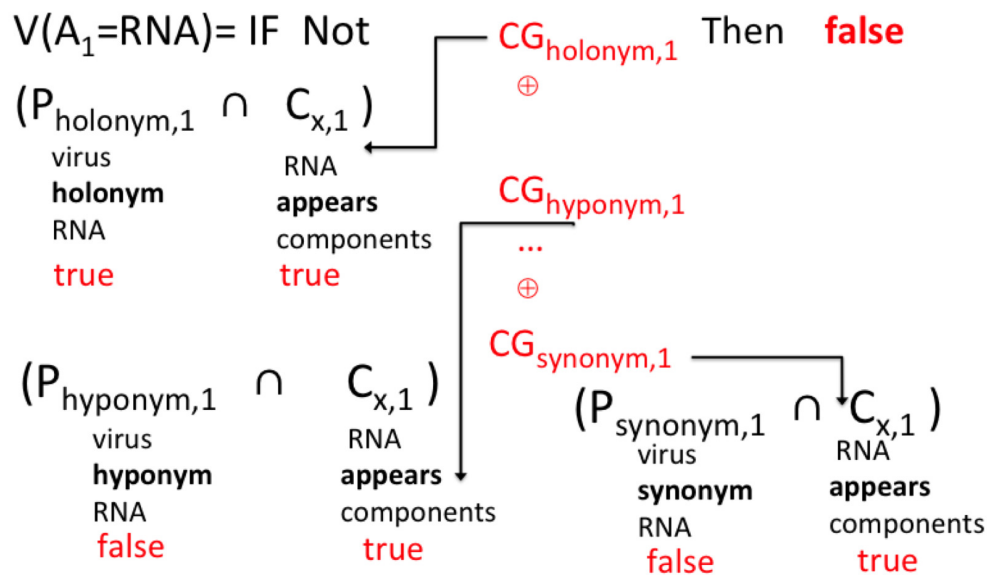


Figura 5.10: Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta incorrecta

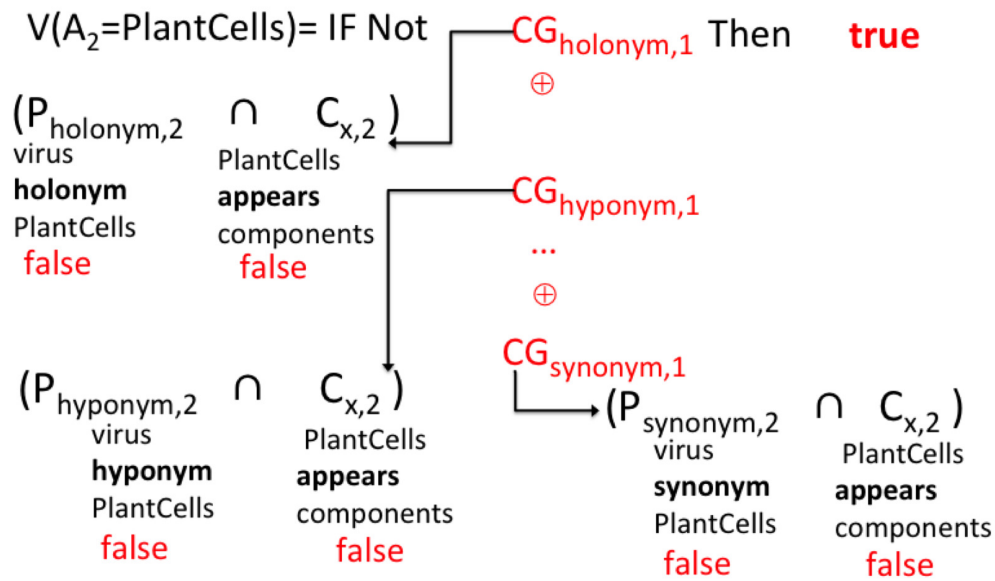


Figura 5.11: Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta correcta

de respuesta que aparece en la Figura 4.13 .

En el caso de la regla de respuesta mencionada en la Figura 4.13 como hay un operador *not*, entonces dentro del condicional requerimos que no se cumpla la condición interna, en este caso vamos a tener descritos los patrones expositivos, de  $CG_{pa,i}$  al  $CG_{pz,i}$  que son todos los posibles patrones encontrados de acuerdo al bloque pregunta respuesta que se realizó en el análisis del corpus, por lo que se ejecuta del paso 14 al 19 del Algoritmo 2, donde se utilizará una bandera con valor falso para que se encienda cuando se encuentre el patrón correcto de acuerdo a uno de los casos de la sentencia *switch*.

Como se observa en las Figuras 5.10 y 5.11, se realiza la aplicación de la regla de respuesta anteriormente mencionada, se evalúan los patrones holónimos, hipónimos, sinónimos, merónimos entre otros. En este caso, se muestran la aplicación de estos patrones para ilustrar el funcionamiento de la obtención de la respuesta correcta y una respuesta incorrecta, considerando que siempre se tiene una respuesta correcta y 3 incorrectas en cada pregunta.

En la Figura 5.10, se inicia el análisis con una de las respuestas incorrectas, en este caso  $A_1 = RNA$ , se aplica la regla de condición, como se observa en el antecedente de la regla, los  $CG_{TipoPatron}$  se aplican al mismo tiempo, por lo que se tendrá que probar si esa respuesta corresponde a uno de los patrones y al contexto para cada  $CG_{TipoPatron}$ . Por lo que se verifica, si en el predicado el argumento izquierdo virus corresponde con el

operador *holonym* respecto al argumento derecho que es RNA, en este caso es verdadero puesto que hay evidencia dentro del pasaje que está descrito como un componente, por lo que también se valida el contexto que tiene como argumento izquierdo RNA, el operador *appears* y como argumento derecho *components* al ser los dos elementos verdaderos la regla de condición es verdadera pues ambos elementos están conectados mediante el operador *and*.

Por otro lado, en cuestión de los otros patrones, aunque el contexto es verdadero para los dos casos, la regla se hace falsa pues en el caso de los hipónimos es falso, porque la relación de virus con RNA sólo es holonimia y no hiponimia, como se muestra en la Figura 5.10. Finalmente en cuestión de la sinonimia no se cumple la propiedad de que el argumento izquierdo virus sea sinónimo (operador *synonym*) del argumento derecho RNA, por tanto en este caso resultan falsos y el único verdadero es la relación de holonimia, por tanto la respuesta RNA resultará con valor verdadero, puesto que en un *or* exclusivo, se regresa verdadero si sólo un elemento es verdadero y los demás son falsos, pero al negarlo recordando que se busca el componente que no cumple arrojará falso, ver (Figura 5.10).

Este proceso se realiza de la misma manera con las otras respuestas que si corresponden con componentes de los virus y también arrojarán falso, por la negación del excepto.

En la Figura 5.11, se valida la respuesta  $A_2 = Plant\ Cells$ , ahora se verifica si el argumento izquierdo virus corresponde con el operador *holonym* respecto al argumento derecho que es *Plant Cells*, en este caso es falso, además de que también es falso el contexto definido con el argumento izquierdo *Plant Cells*, el operador *appears* y el argumento derecho *components*, puesto que *Plant Cells* no aparece como componente en el pasaje.

Con el resto de las patrones: hipónimos, sinónimos también *Plant Cells* no corresponde a ellos, así como en todos el contexto también es falso. Eso implica que en cada regla de condición todos los resultados serán falsos y aplicando el *or* exclusivo, como todos tienen el mismo resultado la evaluación resulta falsa y por tanto la negación resulta verdadera, por lo que  $A_2 = Plant\ Cells$  es la respuesta correcta al no formar parte de los virus. Esto se ilustra en los pasos 20 y 21 del Algoritmo 2, en el paso 20 al activarse la bandera implica que se encontró un patrón que correspondía con la respuesta (en este caso la negación, es decir el elemento que no es un componente de los virus), por lo que en el paso 21 se retorna la respuesta correcta.

Este proceso se realiza de la misma manera con las otras respuestas incorrectas donde sucederá el mismo proceso (repetición del paso 12 del Algoritmo 2).

De la misma forma se abordan las otras preguntas del pasaje, verificando los patrones

para encontrar la respuesta correcta (repetición del paso 11 del Algoritmo 2).

### 5.2.2. Caso: textos narrativos

En este caso, en la Figura 5.12 se tiene un ejemplo de un texto narrativo, cuyo tema principal es el jazz. Aplicando el Algoritmo 2, se procesa en los pasos 1 y 2 el pasaje. Después se aplica el paso 3 del Algoritmo 2, que internamente funciona como se muestra en la Figura 5.13, que es el proceso de evaluación de la regla de los bloques cognitivos. Se toman del texto de la Figura 5.12 tres preguntas para la validación de los bloques (paso 4 del Algoritmo 2), la primera es la pregunta 4: *The Nickname Duke belonged to which of the following bandleader?*, sobre un líder de banda que se llama Duke, resultando el predicado manejado en el paso 2 del Algoritmo 2, de la forma *argumento izquierdo propiedadKB argumento derecho*, en este caso *nickName belongs bandleaders*, lo que involucra a un personaje, por tanto corresponde a la verificación de la condición de reglas de personas y fechas (paso 7 del Algoritmo 2).

Posteriormente se encuentra la pregunta 1: *it can be inferred from the passage that Louis Armstrong went to Chicago for which of the following reasons*, que corresponde a causas, cuál fue la causa que un personaje se fuera a Chicago, cuyo predicado generado es: *wentChicago belongs causes*, por tanto también está presente la regla de causas, problemas y acciones (paso 7 del Algoritmo 2).

Finalmente se encuentra una de las preguntas que relaciona eventos, la pregunta 5: *The passage supports which of the following conclusions*, en este caso se interpreta sobre cuál es la idea que apoya la conclusión, que implica ubicar el evento significativo mencionado en el pasaje, cuyo predicado resultante es: *supportConclusion belongs event*. Por lo que se cumplen los tres bloques cognitivos de los textos narrativos.

Es decir en el Algoritmo 2, se han ejecutado los pasos: 4 y 7 para seleccionar el bloque y el 8 que retorna al cumplirse la regla que el texto es expositivo. Los pasos 5,6, 9 y 10 del Algoritmo 2, no se ejecutan por ser un *switch* la estructura condicional empleada. Internamente la regla de los bloques cognitivos (ver Figura 5.13) evaluará los bloques requeridos para el texto narrativo  $CB_{Narrative}$  como se mencionó anteriormente, al estar las condiciones de regla conectadas por el operador *and* y cumplirse los tres bloques retornará verdadero en  $CB_{Narrative}$ , por tanto la regla es verdadera.

En el caso de los  $CB$  para textos expositivos y argumentativos que se conectan con un *or* exclusivo son falsos, porque no hay elementos de sus bloques, por tanto al ser sólo válida la información del bloque narrativo, los demás  $CB$  son falsos. Así, la regla arroja que es un texto narrativo y que por tanto es aceptada.

## Questions

1. It can be inferred from the passage that Louis Armstrong went to Chicago for which of the following reasons?

- (A) To form his own band (B) To learn to play Chicago - style jazz  
(C) To play in Joseph Oliver's band (D) To make recordings with the Hot Five

2. According to the passage, which of the following Black bands was the first to make a significant set of jazz recordings?

- (A) The Hot Seven band (B) Fletcher Henderson's band  
(C) The Red Hot Peppers band (D) King Oliver's Creole jazz Band

3. As used in line 12, the word "note" could best be replaced by which of the following?

- (A) distinction (B) memorandum (C) mood (D) song

4. The nickname "Duke" belonged to which of the following bandleaders?

- (A) Louis Armstrong (B) Joseph Oliver  
(C) Edward Ellington (D) Fletcher Henderson

5. The passage supports which of the following conclusions?

- (A) By the 1930's jazz was appreciated by a wide audience  
(B) Classical music had a great impact on jazz  
(C) jazz originated in New Orleans in the early nineteenth century  
(D) jazz band were better known in, Europe than in the United States

- The first jazz musicians played in New Orleans during the early 1900's. After 1917, many of the New Orleans musicians moved to the south side of Chicago, where they continued to play their style of jazz. Soon Chicago was the new-center for jazz.

- 

- Several outstanding musicians emerged as leading jazz artists in Chicago. Daniel Lotus "Satchmo" Armstrong, born in New Orleans in 1900, was one. Another leading musician was Joseph king Oliver, who is also credited with having discovered Armstrong, when they were both in New Orleans. While in Chicago, Oliver asked Armstrong, who was in New Orleans, to join his band. In 1923 King Oliver's Creole Jazz Band made the first important set of recordings by a Hot Five and Hot Seven bands under Louis Armstrong also made recordings of special note.

- 

- Although Chicago's South Side was the main jazz center, some musicians in New York were also demanding attention in jazz circles. In 1923 Fletcher Henderson already had a ten - piece band that played jazz. During the early 1930's, the number of players grew to sixteen. Henderson's band was considered a leader in what some people have called the Big Band Era. By the 1930's, big dance bands were the rage. Large numbers of people went to ballrooms to dance to jazz music played by big bands.

- 

- One of the most popular and also a very famous jazz band was the Duke Ellington band. Edward "Duke" Ellington was born in Washington, D.C., in 1899 and died in New York City in 1974. He studied the piano as a young boy and later began writing original musical compositions. The first of Ellington's European tours came in 1933. He soon received international fame for his talent as a band leader, composer, and arranger. Ten years later, Ellington began giving annual concerts at Carnegie Hall in New York City. People began to listen to jazz in the same way, that they had always listened to classical music.

Figura 5.12: Ejemplo texto narrativo, fuente Universidad de Chuvanan Chuvanan-University (2018)

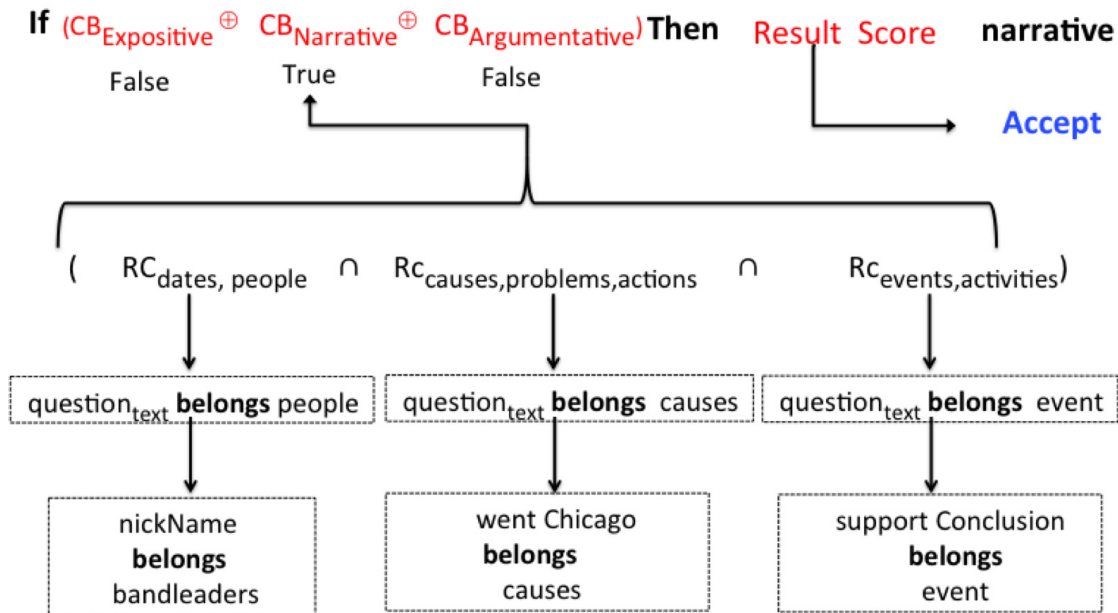


Figura 5.13: Aplicación de la regla de bloques cognitivos para textos narrativos

Posteriormente se ejecuta el paso 11 del Algoritmo 2, cuyo propósito es generar las respuestas correctas para cada pregunta. En este caso se seleccionó la pregunta 4 del texto de la Figura 5.12, para ejemplificar el funcionamiento.

En el paso 12 de Algoritmo 2, se validan las respuestas para la pregunta 4: *The Nickname Duke belonged to which of the following bandleader?*, corresponde al patrón de personas. Por lo que se ejecuta el paso 13 de este algoritmo, que aplica la regla de respuesta que aparece en la Figura 4.13 .

En el caso de la regla de respuesta mencionada en la Figura 4.13, se tienen descritos los patrones narrativos, de  $CG_{pa,i}$  al  $CG_{pz,i}$  que son todos los posibles patrones encontrados de acuerdo al bloque pregunta respuesta que se realizó en el análisis del corpus, por lo que se ejecuta del paso 14 al 19 del Algoritmo 2, donde se utilizará una bandera con valor falso para que se encienda cuando se encuentre el patrón correcto de acuerdo a uno de los casos de la sentencia *switch*.

Como se observa en las Figuras 5.14 y 5.15, se realiza la aplicación de la regla de respuesta antes mencionada, se evalúan los patrones de personas, eventos, objetos entre otros. En este caso, se muestran la aplicación de estos patrones para ilustrar el funcionamiento de la obtención de la respuesta correcta y una respuesta incorrecta.

En la Figura 5.14 se inicia el análisis con la respuesta correcta, en este caso la respuesta  $A_3 = Edward Ellington$ , se aplica la regla de condición, como se observa en el

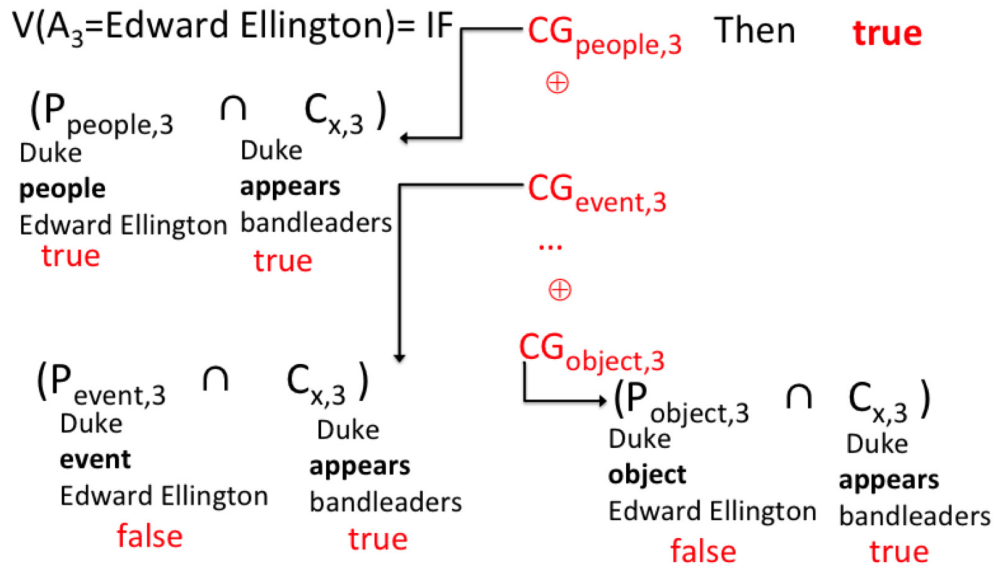


Figura 5.14: Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta correcta

antecedente de la regla, los  $CG_{\text{TipoPatron}}$  se aplican al mismo tiempo, por lo que se tendrá que probar si esa respuesta corresponde a uno de los patrones y al contexto para cada  $CG_{\text{TipoPatron}}$ . Por lo que se verifica, si en el predicado el argumento izquierdo *Duke* corresponde con el operador *people* respecto al argumento derecho que es *Edward Ellington*, en este caso es verdadero, porque hay evidencia de esa respuesta en el pasaje. Además de que también es verdadero el contexto puesto que si aparece la palabra *Duke* en el texto del pasaje respecto a la palabra *bandleaders*. Así la regla de condición es verdadera pues ambos elementos están conectados mediante el operador *and*.

Con el resto de las patrones: eventos y objetos, *Edward Ellington* no corresponde a ellos, porque *Edward Ellington* es una persona, aunque el contexto es verdadero, por la ocurrencia del nombre en el pasaje. Eso implica que en cada regla de condición todos los resultados serán falsos y aplicando el *or* exclusivo. la evaluación resulta verdadera porque sólo una regla fue verdadera, por lo que  $A_3 = \text{Edward Ellington}$  es la respuesta correcta, como se muestra en la Figura 5.14 .

En la Figura 5.15 se verifica la respuesta incorrecta  $A_2 = \text{Joseph Oliver}$ , se aplica la regla de condición, como se observa en el antecedente de la regla, los  $CG_{\text{TipoPatron}}$  se aplican al mismo tiempo, por lo que se tendrá que probar si esa respuesta corresponde a uno de los patrones y al contexto para cada  $CG_{\text{TipoPatron}}$ . Por lo que se verifica si el argumento izquierdo *Duke* corresponde con el operador *people* respecto al argumento

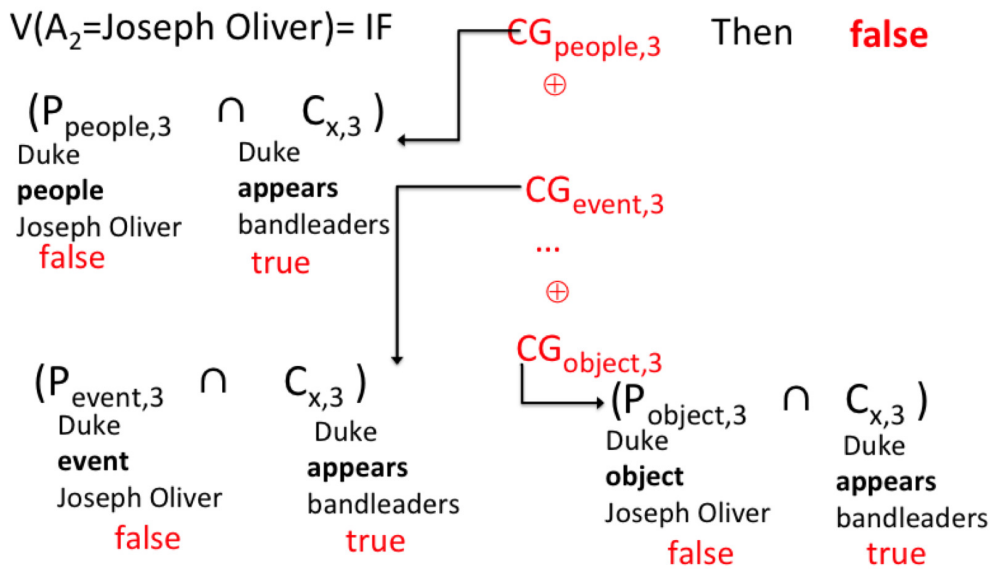


Figura 5.15: Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta incorrecta

derecho que es *Joseph Oliver*, en este caso es falso, puesto que no hay evidencia dentro del pasaje que sea su nombre, aunque el contexto es verdadero, porque *Duke* aparece en el texto del pasaje.

Por otro lado, en cuestión de los otros patrones, aunque el contexto es verdadero para los dos casos, la regla se hace falsa pues en el caso de los eventos y objetos devuelve falso, porque *Duke* no es un objeto ni un evento, por lo que todas las reglas serán falsas y la evaluación será falsa.

Este proceso se realiza de la misma manera con las otras respuestas donde sucederá el mismo proceso (repetición del paso 12 del Algoritmo 2).

De la misma forma se abordan las otras preguntas del pasaje, verificando los patrones para encontrar la respuesta correcta (repetición del paso 11 del Algoritmo 2).

### 5.2.3. Caso: textos argumentativos

En este caso, en la Figura 5.16 se tiene un ejemplo de un texto argumentativo, cuyo tema principal es el arte. Aplicando el Algoritmo 2 se procesa el pasaje en los pasos 1 y 2 del mismo.

Después se aplica el paso 3 del Algoritmo 2, que funciona internamente como se muestra en la Figura 5.17 que corresponde a la evaluación de los bloques cognitivos.

A partir del texto de la Figura 5.16, se toman tres preguntas para la validación de los

bloques (paso 4 del Algoritmo 2), la primera pregunta es la 4: *Compared to those of the ballet, Isadora Duncan's costumes were less*, es decir se enfoca a que los atuendos se comparan con alguna propiedad lo que corresponde a una hipótesis, porque se requiere demostrar su existencia, resultando el predicado como se ha denotado, argumento izquierdo operador argumento derecho de la forma: *compareDuncan'sCostumes belongs hypothesis*. Por tanto corresponde a la verificación de la condición de reglas de hipótesis (paso 9 del Algoritmo 2).

La segunda pregunta es la 2: *According to the passage, what did nature represent to Isadora Duncan?*, corresponde a una pregunta de una opinión sobre lo que representa la naturaleza para Isadora Duncan, por lo que corresponde a la regla de posiciones, cuyo predicado generado es: *representNatureToIsadora belongs position*.

Finalmente la tercera pregunta es la 1: *Which of the following would be the best title for the passage?* que evalúa el posible título del pasaje, que implica la regla de frases y estructura, que se modela en el predicado *wouldBeBestTitle belongs structure*, en este caso se cumplen los tres bloques cognitivos de los textos argumentativos.

Es decir en el Algoritmo 2, se han ejecutado los pasos: 4 y 9 para seleccionar el bloque y el 10 que retorna al cumplirse la regla que el texto es argumentativos. Los pasos 5,6, 7 y 8 del Algoritmo 2, no se ejecutan por ser un *switch* la estructura condicional empleada. Internamente la regla de los bloques cognitivos (ver Figura 5.17) evaluará los bloques requeridos para el texto argumentativo  $CB_{Argumentative}$  como se mencionó anteriormente, al estar las condiciones de regla conectadas por el operador *and* y cumplirse los tres bloques retornará verdadero en  $CB_{Argumentative}$ , por tanto la regla es verdadera.

En el caso de los  $CB$  para textos expositivos y narrativos que se conectan con un *or* exclusivo son falsos, porque no hay elementos de sus bloques, por tanto al ser sólo válida la información del bloque argumentativo, los demás  $CB$  son falsos. Así, la regla arroja que es un texto argumentativo y que por tanto es aceptada.

Posteriormente se ejecuta el paso 11 del Algoritmo 2, cuyo propósito es generar las respuestas correctas para cada pregunta. En este caso se seleccionó la pregunta 1 del texto de la Figura 5.16, para ejemplificar el funcionamiento.

En el paso 12 de Algoritmo 2, se validan las respuestas para la pregunta 1: *Compared to those of the ballet, Isadora Duncan's costumes were less?*, que corresponde al patrón de hipótesis. Por lo que se ejecuta el paso 13 de este algoritmo, que aplica la regla de respuesta que aparece en la Figura 4.13 .

En el caso de la regla de respuesta mencionada en la Figura 4.13, se tienen descritos los patrones argumentativos, de  $CG_{pa,i}$  al  $CG_{pz,i}$  que son todos los posibles patrones encontrados de acuerdo al bloque pregunta respuesta que se realizó en el análisis del

**Topic: Art**

Many artists late in the last century were in search of a means to express their individuality. Modern dance was one of the ways some of these people sought to free their creative spirit. At the beginning there was no exacting technique, no foundation from which to build. In later years trial, error, and genius founded the techniques and the principles of the movement. Eventually, innovators even drew from what they considered the dread ballet, but first they had to discard all that was academic so that the new could be discovered. The beginnings of modern dance were happening before Isadora Duncan, but she was the first person to bring the new dance to general audiences and see it accepted and acclaimed.

Her search for a natural movement form sent her to nature. She believed movement should be as natural as the swaying of the trees and the rolling waves of the sea, and should be in harmony with the movements of the Earth. Her great contributions are in three areas.

First, she began the expansion of the kinds of movements that could be used in dance. Before Duncan danced, ballet was the only type of dance performed in concert. In the ballet the feet and legs were emphasized, with virtuosity shown by complicated, codified positions and movements. Duncan performed dance by using all her body in the freest possible way. Her dance stemmed from her soul and spirit. She was one of the pioneers who broke tradition so others might be able to develop the art.

Her second contribution lies in dance costume. She discarded corset, ballet shoes, and stiff costumes. These were replaced with flowing Grecian tunics, bare feet, and unbound hair. She believed in the natural body being allowed to move freely, and her dress displayed this ideal.

Her third contribution was in the use of music. In her performances she used the symphonies of great masters, including Beethoven and Wagner, which was not the usual custom. She was as exciting and eccentric in her personal life as in her dance.

1. Which of the following would be the best title for the passage?

- (A) The Evolution of Dance in the Twentieth Century  
 (B) Artists of the Last Century  
 (C) Natural Movement in Dance  
 (D) A Pioneer in Modern Dance

2. According to the passage, what did nature represent to Isadora Duncan?

- (A) Something to conquer (B) A model for movement  
 (C) A place to find peace (D) A symbol of disorder

3. Which of the following is NOT mentioned in the passage as an area of dance that

Isadora Duncan worked to change?

- (A) The music (B) The stage sets  
 (C) Costumes (D) Movements

4. Compared to those of the ballet, Isadora Duncan's costumes were less

- (A) costly (B) colorful (C) graceful (D) restrictive

5. What does the paragraph following the passage most probably discuss?

- (A) Isadora Duncan's further contribution to modern dance  
 (B) The music customarily used in ballet  
 (C) Other aspects of Isadora Duncan's life  
 (D) Audience acceptance of the new form of dance

Figura 5.16: Ejemplo texto argumentativo, fuente Universidad de Chuvanan Chuvanan-University (2018)

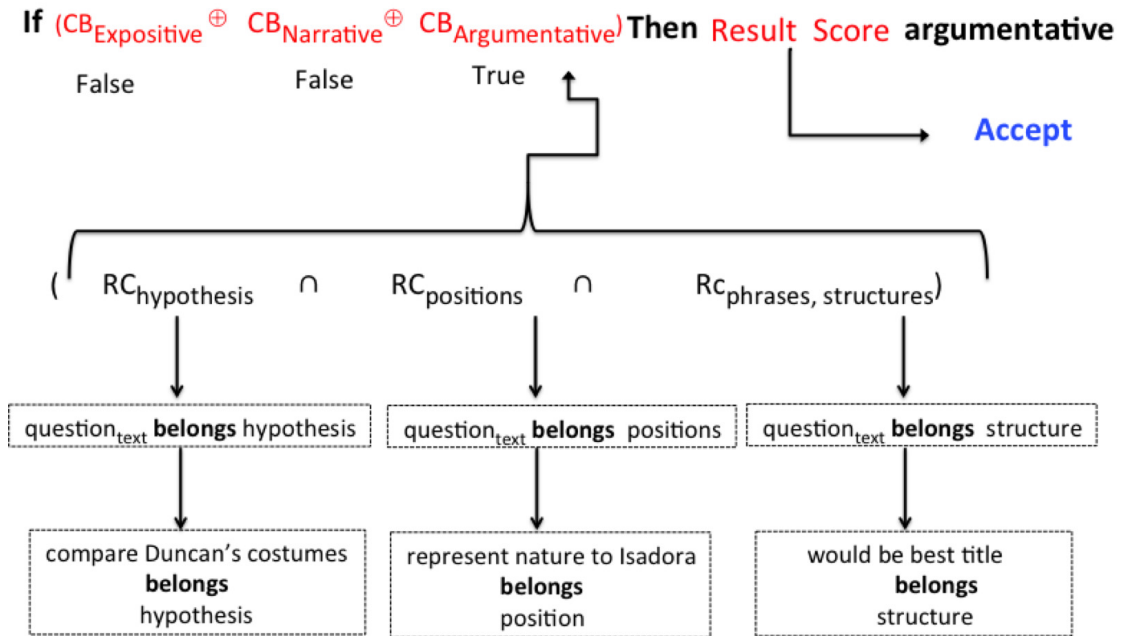


Figura 5.17: Aplicación de la regla de bloques cognitivos para textos argumentativos

corpus, por lo que se ejecuta del paso 14 al 19 del Algoritmo 2, donde se utilizará una bandera con valor falso para que se encienda cuando se encuentre el patrón correcto de acuerdo a uno de los casos de la sentencia *switch*.

Como se observa en las Figuras 5.18 y 5.19, se realiza la aplicación de la regla de respuesta, se evalúan los patrones: hipótesis, opiniones y frases, entre otros. En este caso, se muestran la aplicación de estos patrones para ilustrar el funcionamiento de la obtención de la respuesta correcta y una respuesta incorrecta.

En la Figura 5.18, se hace el análisis con la respuesta correcta  $A_4 = less\ restrictive$ , ahora se verifica si el argumento izquierdo *compare costumes* corresponde con el operador *hypothesis* respecto al argumento derecho que es *lessRestrictive*, en este caso es verdadero, porque hay evidencia que sustenta el tipo de prenda y sus restricciones. Además de que también es verdadero el contexto puesto que si aparece *lessRestrictive* en la forma de describir las prendas. Así la regla de condición es verdadera pues ambos elementos están conectados mediante el operador *and*.

Con el resto de las patrones: opinión y frase, *less restrictive* no corresponde a ellos, aunque el contexto es verdadero. Eso implica que en cada regla de condición todos los resultados serán falsos y aplicando el *or* exclusivo, la evaluación resulta verdadera porque sólo una regla con el patrón de hipótesis fue verdadera, por lo que  $A_4 = less\ restrictive$  es la respuesta correcta .

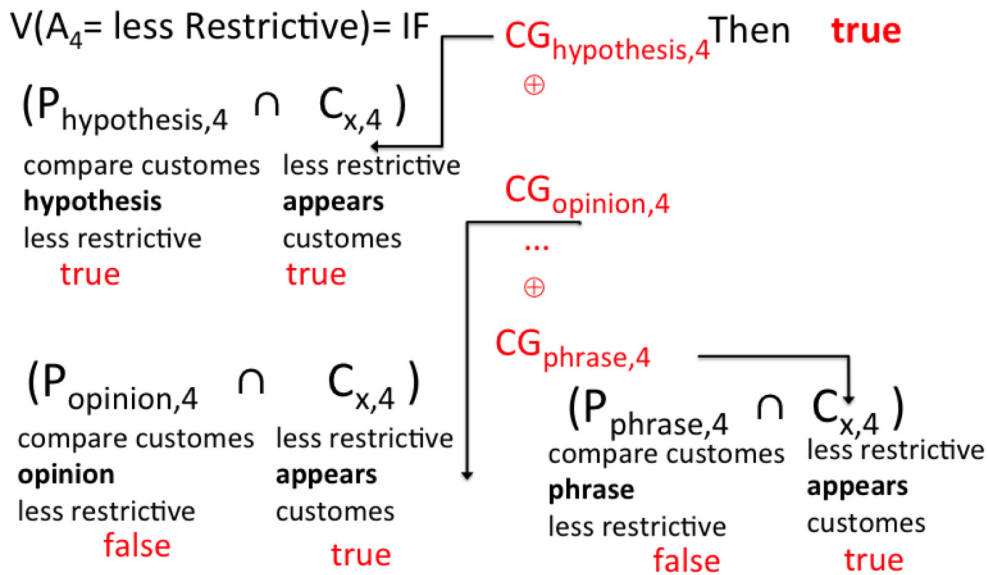


Figura 5.18: Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta correcta

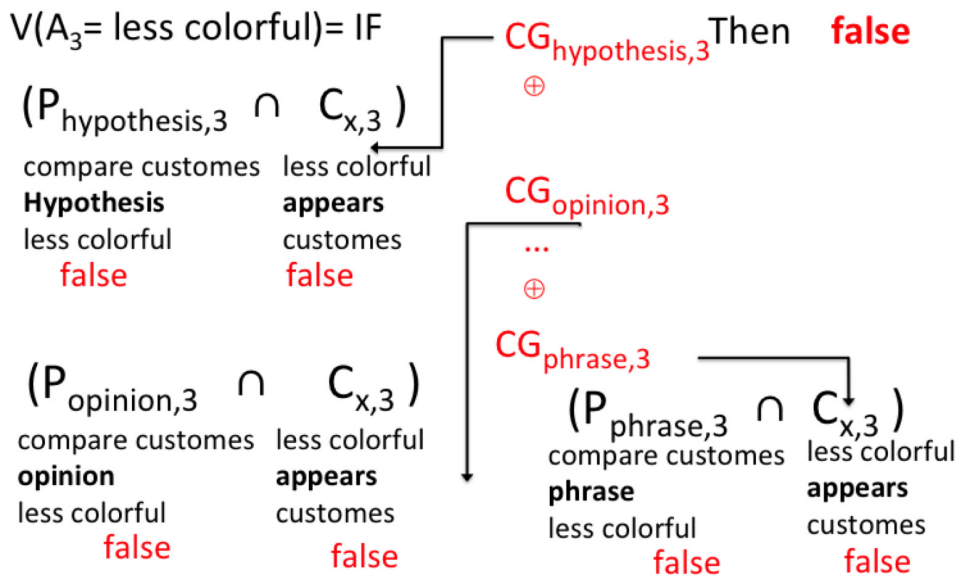


Figura 5.19: Ejemplo de la aplicación de la regla de la respuesta que proporciona la respuesta incorrecta

En la Figura 5.19, se muestra el caso de la respuesta incorrecta  $A_3 = less\ color\ ful$ , se aplica la regla de condición, como se observa en el antecedente de la regla, los  $CG_{TipoPatron}$  se aplican al mismo tiempo, por lo que se tendrá que probar si esa respuesta corresponde a uno de los patrones y al contexto para cada  $CG_{TipoPatron}$ . Por lo que se verifica si el argumento izquierdo  $compareCustomes$  corresponde con el operador  $hypothesis$  respecto al argumento derecho que es  $lessColorful$ , en este caso es falso puesto que no hay evidencia dentro del pasaje que se hable del color en comparación de las prendas y el contexto es falso porque no hay ocurrencia del color en el pasaje. Por otro lado, en cuestión de los otros patrones, aunque el contexto es verdadero para los dos casos, la regla se hace falsa pues en el caso de las opiniones y frases devuelve falso, porque no tienen relación con la comparación en sus estructuras, por lo que todas las reglas serán falsas y la evaluación será falsa.

Este proceso se realiza de la misma manera con las otras respuestas donde sucederá el mismo proceso (repetición del paso 12 del Algoritmo 2).

De la misma forma se abordan las otras preguntas del pasaje, verificando los patrones para encontrar la respuesta correcta (repetición del paso 11 del Algoritmo 2).



# Capítulo 6

## Conclusiones y Trabajo Futuro

Considerando los objetivos propuestos y la hipótesis planteada fue posible realizar la construcción de un modelo semántico, a partir de la formalización teórica de los modelos cognitivos, la lógica de predicados y el cálculo de situaciones.

Realizando el análisis de los pasajes se tuvieron dos enfoques, en el primero se le dio relevancia al cuerpo del pasaje lo que resultó en la identificación de estructuras semánticas propias de cada tipo de texto. Sin embargo, este enfoque no garantizaba que hubiera confusión en la interpretación del tipo de pasaje considerando por ejemplo, que en los textos argumentativos se incluyen algunos elementos de los textos expositivos y narrativos para sustentar opiniones o refutaciones, por lo que se decidió cambiar de enfoque.

En el segundo enfoque se le dio relevancia al bloque pregunta respuesta, lo que resultó en una caracterización más precisa sobre el tipo de texto, puesto que las preguntas contaban con patrones específicos que determinaban el tipo de texto que se evaluaba, esto permitió definir bloques cognitivos de memorización y retención, inferencia, ordenamiento, comparación y semejanza e interpretación.

Posteriormente se analizó el mecanismo de inferencia a partir de los patrones para determinar la respuesta correcta, por lo que se revisó la teoría de los sistemas basados en reglas y se propuso una adaptación de las reglas para la deducción de la respuesta; esta adaptación incluye dos reglas: la de bloques cognitivos para determinar el tipo de texto y la de inferencia de la respuesta que aplica los patrones encontrados y el contexto para discriminarla.

El factor clave para construir este modelo fue la representación de las estructuras semánticas y los bloques cognitivos, así como la teoría de los sistemas basados en reglas, con estos elementos se realizó la arquitectura del modelo, la determinación de

los patrones y el algoritmo para la deducción de la respuesta a partir de las reglas.

Dentro de las contribuciones de este trabajo se tiene: un corpus etiquetado en formato xsd-xml, los patrones cognitivos extraídos a partir del corpus, la metodología de trabajo realizado mediante un algoritmo, así como un algoritmo determinista para la deducción de las respuestas correctas de los pasajes tipo TOEFL.

Como trabajo futuro se contempla implementar técnicas de web scrapping para recuperar pasajes de otras fuentes de forma automática para así contar con un repositorio actualizado de pasajes, realizar validaciones y adaptaciones a las nuevas versiones del Examen TOEFL, así como generar un sistema de apoyo a los estudiantes en el entrenamiento de comprensión lectora en este tipo de textos.

Por otro lado también se considera implementar las reglas inferenciales en node-rules, clips o prolog para contar con un sistema que apoye el entrenamiento de los usuarios en los ejercicios de comprensión lectora del Examen TOEFL.

# Capítulo 7

## Publicaciones

En este capítulo se mencionan las publicaciones indexadas en Scopus producto de mi trabajo de investigación.

Contreras González, M., Tovar Vidal, M., De Ita Luna, G. (2021). Inferential Rules for Identifying Answers in TOEFL Texts, *Lecture Notes in Computer Science*, 12725. doi: 10.1007/978-3-030-77004-4. [https://link.springer.com/chapter/10.1007/978-3-030-77004-4\\_2/](https://link.springer.com/chapter/10.1007/978-3-030-77004-4_2/)

Contreras González, M., Tovar Vidal, M., De Ita Luna, G., López López, A. (2020). Knowledge Representation in TOEFL Expository Texts. *Computación Y Sistemas*, 24(2), 511-522. <https://www.cys.cic.ipn.mx/ojs/index.php/CyS/article/view/3375/2814/>

Contreras González, M., Tovar Vidal, M., De Ita Luna, G., Mendoza Rojas, A. (2020). Estrategias cognitivas para determinación de inferencias en la comprensión de textos narrativos. En H. Castillo, M. Tovar, C. Zepeda (Eds.), *Los objetos de aprendizaje y sus aplicaciones en la Educación* (pp. 109-116). Dirección General de Publicaciones BUAP.

Contreras González, M., Tovar Vidal, M., De Ita Luna, G. (2019). Cognitive relations for argumentative texts. *CEUR-WS*, 2585.

Contreras González, M., Tovar Vidal, M., De Ita Luna, G. (2019). Repositorio de contextos para el apoyo de la comprensión lectora. En A. Olvera López, I. Olmos Pineda (Eds.), *Avances en Tecnologías de Lenguaje y del Conocimiento* (pp. 105- 112). Direc-

ción General de Publicaciones BUAP.

Contreras González, M., Tovar Vidal, M., De Ita Luna, G., Torres Acuitlapa, O., Mejía Matías, H. (2019). Identificación de estrategias semánticas para comprensión lectora. En H. Castillo, M. Tovar, C. Zepeda (Eds.), *Las entidades digitales educativas y sus aplicaciones* (pp. 55-61). Dirección General de Publicaciones BUAP.

Contreras González, M., Tovar Vidal, M., De Ita Luna, G., Bello López, P. (2019). Categorización de inferencias en la comprensión de textos en inglés. En H. Castillo, M. Tovar, C. Zepeda (Eds.), *Los objetos de aprendizaje como apoyo en el proceso de enseñanza-aprendizaje* (pp. 77-83). Dirección General de Publicaciones BUAP.

# Bibliografía

- Amostoy, M. (2002). La investigación sobre el desarrollo y la enseñanza de las habilidades de pensamiento. *Revista Electronica de Investigacion Educativa*, 4(1), 1–10.
- Burkert, G. (1995). *Computational Lexical Semantics*, chapter Lexical semantics and terminological knowledge. Cambridge.
- Castillo, E., G. J. H. A. (1997). *Expert Systems and Probabilistic Network Models*. Springer-Verlag.
- Chuvanan-University (2018). Collection of TOEFL passages.
- Dennis, S. & Kintsch, W. (2012). *Text Comprehension: Evaluating a Memory-Based Account*, chapter The Text Mapping and Inference Rule Generation Problems.
- Dijk, T. A. V. (2009). *The Construction of Mental Representations During Reading* (1 ed.), volume 1, chapter Context Models in Discourse Processing, (pp. 404). H. van Oostendorp and S. Goldman.
- Frank, S., Koppen, M., Noordman, L., & Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science*, 27, 875–910.
- García, E. G. (1993). La comprensión de textos. Modelo de procesamiento y estrategias de mejora. *Didáctica*, 1(5), 87–113.
- Garrido, N. C. (2010). Relaciones semanticas entre las palabras: hiponimia, sinonimia, polisemia, homonimia y antonimia. los cambios de sentido. *Contribuciones a las Ciencias Sociales*, mayo 2010.
- Gildea, D. & Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational Linguistics*, 28(3), 245–288.

- Goldman, S. & Varma, S. (1995). Discourse comprehension: Essays in honor of Walter Kintsch, chapter CAPping the construction-integration model of discourse comprehension, (pp. 337–358). C. A. Weaver, S. Mannes, C. R. Fletcher.
- Gutiérrez-Calvo, M. (1999). *Psicolingüística del español*, chapter Inferencias en la comprensión del lenguaje. Trotta.
- Hernández-Guanir, P. (2010). *Test de Estrategias Cognitivo-Emocionales* (1 ed.). Fray Bernardino Sahagún, 24; Madrid: Tea Ediciones.
- Johnson-Laird, P. (2006). *Como razonamos*. Oxford University Press.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163–182.
- Kintsch, W. & Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(1), 363–394.
- León, J. & Pérez, O. (2003). *Conocimiento y Discurso: Claves para inferir y comprender*, chapter Taxonomías y tipos de inferencias. Pirámide.
- Lezcano, E. (1977). Una aproximación a la gramática de valencias. *Lenguaje y Textos*, 6(7), 161–173.
- Lin, D. & Pantel, P. (2001). Discovery of inference rules for question answering. *Natural Language Engineering*, 7(4), 343–360.
- Mahnke, M. & C.B. Duffy (1993). *TOEFL Preparation Course*. Heinemann.
- McCarthy, J. & Buvac, S. (1994). *Formalizing context* (expanded notes). Technical report, Stanford University.
- McNamara, D., Levinstein, I., & Bonthum, C. (2004). *istart: Interactive strategy training for active reading and thinking*. *Behavior Research Methods Instruments Computers*, 36(2), 222–233.
- Morales, M. G. (2017). *Uso y producción de verbos en niños con trastornos de aprendizaje*. Master's thesis, Benemérita Universidad Autónoma de Puebla, Avenida San Claudio y 14 sur.
- Peronard, M., L. Gómez, L., S.G. Parodi, S., & Nuñez, P. (1997). *Comprensión de Textos Escritos: de la teoría a la sala de clase*. Valparaíso: Editorial Andrés Bello.

- Pozo, P. M. (2008). Los roles semánticos en la tecnología del lenguaje humano: anotación y aplicación. PhD thesis, Universidad de Alicante, España.
- Reither, R. (2001). Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems. Oxford MIT Press.
- Singer, M. & Kintsch, W. (2001). Text retrieval: A theoretical exploration. *Discourse Processes*, 31, 27–59.
- Snow, C. (2001). Reading for understanding. CA: RAND Education the Science and Technology Police Institute.
- Viramonte, M. (2008). Comprensión Lectora, Dificultades estratégicas en resolución de preguntas inferenciales. Ediciones Colihue.
- Woszezenki, C., Gonzaleves, A., & Souza, J. (2018). A knowledge discovery model based on semantic and temporal associations between textual elements. *IEEE Latin America Transactions*, 16(4), 1243–1249.
- Yeari, M. & van den Broek, P. (2011). A cognitive account of discourse understanding and discourse interpretation: The landscape model of reading. *Discourse Studies*, 13(5), 635–643.
- Zenteno, C. (2000). El entañamiento.¿inferencia proposicional o léxica? 26, 227–244.