

FACULTAD DE INGENIERIA MECANICA Y ELECTRICA
PROCESAMIENTO DE LENGUAJE NATURAL

NIVELES DEL PROCESAMIENTO DE LENGUAJE NATURAL

PROFESOR: OSWALDO MORAN CARRILLO

ALUMNO:

ULISES YAREL GARCIA DE LOS SANTOS

6-D

26/05/2022

El PLN se concibe como el reconocimiento y utilización de la información expresada en lenguaje humano a través del uso de sistemas informáticos. En su estudio intervienen diferentes disciplinas tales como lingüística, ingeniería informática, filosofía, matemáticas y psicología. Debido a las diferentes áreas del conocimiento que participan, la aproximación al lenguaje en esta perspectiva es también estudiada desde la llamada ciencia cognitiva. Tanto desde un enfoque computacional como lingüístico se utilizan técnicas de inteligencia artificial:

- modelos de representación del conocimiento y de razonamiento,
- lenguajes de programación declarativos,
- algoritmos de búsqueda, y
- estructuras de datos.

Se investiga cómo el lenguaje puede ser utilizado para cumplir diferentes tareas y la manera de modelar el conocimiento.

En los siguientes párrafos se presenta una introducción a las técnicas que se aplican para el tratamiento del lenguaje natural. Generalmente la bibliografía sobre el tema se caracteriza por su estilo técnico y, dada su componente interdisciplinar, se presenta como una materia de difícil comprensión para los legos en el tema.

En vista a conocer estas técnicas de representación y procesamiento, es necesario tener en cuenta una doble dimensión: se trata por una parte de un problema de representación lingüística, y por otra de un problema de tratamiento mediante recursos informáticos.

El uso de técnicas computacionales procedentes especialmente de la inteligencia artificial no aportaría soluciones adecuadas sin una concepción profunda del fenómeno lingüístico. Por otra parte, las gramáticas utilizadas para el tratamiento del lenguaje han evolucionado hacia modelos más adecuados para un tratamiento computacional.

Cuatro niveles de análisis

El estudio del lenguaje natural se estructura normalmente en 4 niveles de análisis:

- morfológico,
- sintáctico,
- semántico y
- pragmático.

Además se pueden incluir otros niveles de conocimiento como es la información fonológica, referente a la relación de las palabras con el sonido asociado a su pronunciación; el análisis del discurso, que estudia cómo la información precedente puede ser relevante para la comprensión de otra información; y, finalmente, lo que se denomina conocimiento del mundo, referente al conocimiento general que los hablantes han de tener sobre la estructura del mundo para mantener una conversación.

Análisis morfológico

Su función consiste en detectar la relación que se establece entre las unidades mínimas que forman una palabra, como puede ser el reconocimiento de sufijos o prefijos. Este nivel de análisis mantiene una estrecha relación con el léxico.

El léxico es el conjunto de información sobre cada palabra que el sistema utiliza para el procesamiento. Las palabras que forman parte del diccionario están representadas por una entrada léxica, y en caso de que ésta tenga más de un significado o diferentes categorías gramaticales, tendrá asignada diferentes entradas.

En el léxico se incluye la información morfológica, la categoría gramatical, irregularidades sintácticas y representación del significado.

Normalmente el léxico sólo contiene la raíz de las palabras con formas regulares, siendo el analizador morfológico el que se encarga de determinar si el género, número o flexión que componen el resto de la palabra son adecuados.

Análisis sintáctico

Tiene como función etiquetar cada uno de los componentes sintácticos que aparecen en la oración y analizar cómo las palabras se combinan para formar construcciones gramaticalmente correctas. El resultado de este proceso consiste en generar la estructura correspondiente a las categorías sintácticas formadas por cada una de las unidades léxicas que aparecen en la oración. Las gramáticas, tal como se muestra en la siguiente figura, están formadas por un conjunto de reglas:

O --> SN, SV

SN --> Det, N

SN --> Nombre Propio

SV --> V, SN

SV --> V

SP --> Preposición, SN

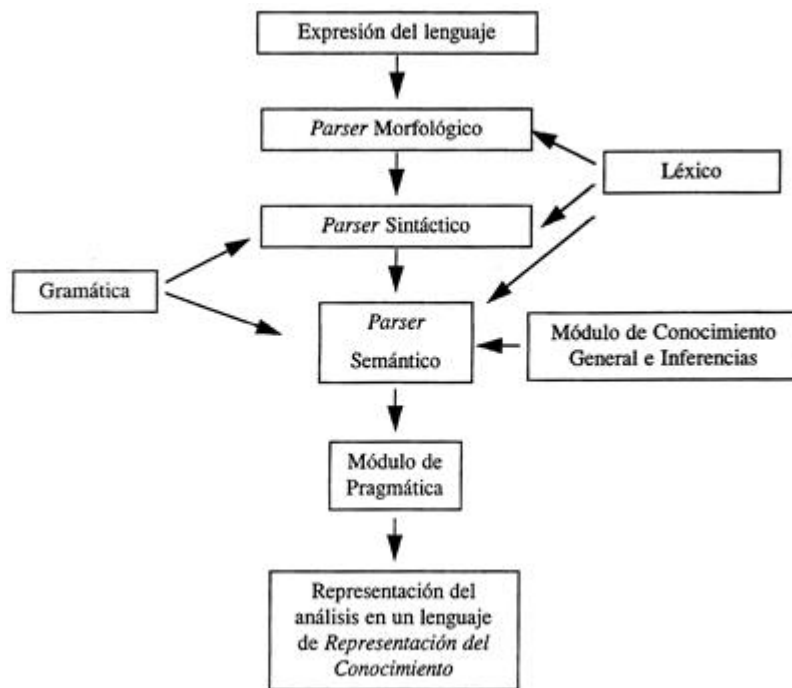
SN = sintagma nominal

SV = sintagma verbal

Det = determinante

Ejemplo de una gramática simple: las reglas tiene como función la composición de estructuras

El resultado del análisis se puede expresar en forma arbórea. Los árboles son formas gráficas utilizadas para expresar la estructura de la oración, consistentes en nodos etiquetados (O, SN, SV..) conectados por ramas:



Análisis semántico

En muchas aplicaciones del PLN los objetivos del análisis apuntan hacia el procesamiento del significado. En los últimos años las técnicas de procesamiento sintáctico han experimentado avances significativos, resolviendo los problemas fundamentales.

Sin embargo, las técnicas de representación del significado no han obtenido los resultados deseados, y numerosas cuestiones continúan sin encontrar soluciones satisfactorias.

Definir qué es el significado no es una tarea sencilla, y puede dar lugar a diversas interpretaciones. A efectos funcionales, para facilitar el procesamiento, la modularidad es una de las propiedades más deseables. Haciendo uso de esta concepción modular es posible distinguir entre significado independiente y significado dependiente del contexto.

El primero, tratado por la semántica, hace referencia al significado que las palabras tienen por sí mismas sin considerar el significado adquirido según el uso en una determinada circunstancia. La semántica, por tanto, hace referencia a las condiciones de verdad de la frase, ignorando la influencia del contexto o las intenciones del hablante. Por otra parte, el componente significativo de una frase asociado a las circunstancias en que ésta se da, es estudiado por la pragmática y conocido como significado dependiente del contexto.

Atendiendo al desarrollo en el proceso de interpretación semántica, es posible optar entre múltiples pautas para su organización, tal como se determinan en los siguientes párrafos.

En referencia a la estructura semántica que se va a generar, puede interesarnos que exista una simetría respecto a la estructura sintáctica, o por el contrario que no se dé tal correspondencia entre ellas. En el primer caso, a partir del árbol generado por el análisis sintáctico se genera una estructura arbórea con las mismas características, sobre la cual se realizará el análisis

semántico. En el segundo caso, en la estructura generada por la sintaxis se produce un curso de transformaciones sobre las cuales se genera la representación semántica.

Cada una de las dos opciones anteriores puede implementarse de forma secuencial o paralela. En la interpretación secuencial, después de haber finalizado la fase de análisis sintáctico, se genera el análisis semántico. En cambio, desde un procedimiento en paralelo, el proceso de análisis semántico no necesita esperar a que el analizador sintáctico haya acabado toda su tarea, sino que puede ir realizando el análisis de cada constituyente cuando éste ha sido tratado en el proceso sintáctico.

Finalmente en combinación con cada una de las opciones anteriores, podemos escoger un modelo en el que exista una correspondencia entre reglas sintácticas y semánticas o, contrariamente, podemos optar por un modelo que no cumpla tal requisito. En caso afirmativo, para cada regla sintáctica existirá una regla semántica correspondiente.

El significado es representado por formalismos conocidos por el nombre de *knowledge representation*. El léxico proporciona el componente semántico de cada palabra en un formalismo concreto, y el analizador semántico lo procesa para obtener una representación del significado de la frase.

Análisis pragmático

Añade información adicional al análisis del significado de la frase en función del contexto donde aparece. Se trata de uno de los niveles de análisis más complejos, la finalidad del cual es incorporar al análisis semántico la aportación significativa que pueden hacer los participantes, la evolución del discurso o información presupuesta.

Incorpora así mismo información sobre las relaciones que se dan entre los hechos que forman el contexto y entre diferentes entidades.