

1. Revisión bibliográfica, estado del arte

- 1.1. Máquinas de traducción: Sistemas capaces de realizar traducciones simples de palabras para un texto preferiblemente si está delimitado a un tema específico, usualmente se dedican a realizar traducciones de palabras simples y no del corpus de un texto. Normalmente son implementadas basadas en un sistema de reglas [1].
- 1.2. Lenguaje Natural Controlado: Se limita la gramática que el usuario pueda usar para hacer más sencilla a la hora de analizar. Se imponen reglas como no usar voz pasiva o evitar el uso de pronombres [2].
- 1.3. Memoria de traducción: Una base de datos que se encarga de traducir frases o segmentos de un texto que previamente fueron traducidos y almacenados en la base [3].

2. Definición de problema, objetivos y productos.

2.1. Definición del Problema

Uno de los grandes problemas con los que se enfrentan los traductores que implementan “traducción automática” es que nos afirman los argumentos debatidos entre Alan Turing e investigadores “¿pueden pensar las máquinas?”, llevando a la conclusión de que este tipo de máquinas no son la excepción y no han logrado por el momento, reemplazar al ser humano. Si utilizamos cualquiera de los famosos traductores para trasladar textos de un lenguaje a otro, y nos encontramos con que el texto original tiene errores gramaticales, lamentablemente las máquinas no pueden dar sentido a la frase como lo haría una persona que se dedica a la traducción, esto porque esta técnica requiere de aprendizaje para lograr habilidad y experiencia, así mismo, es indispensable emplear trabajo duro y conocimiento del tema a tratar. Es por esta razón que el presente trabajo pretende dar una respuesta a esta problemática, estudiando el funcionamiento de estas máquinas y aportando sugerencias de árboles de parsing para frases en español, identificando la transitividad de las oraciones e implementando reglas en el lenguaje prolog que generen árboles que respondan a interrogantes de usuarios ante ciertas frases ya limitadas.

2.2. Objetivo General

- Desarrollar un producto de software en Prolog que mejore la definición de cláusulas gramaticales (DCG).

2.3. Objetivos Específicos

- Estudiar cómo emplea Prolog el procesamiento de lenguaje natural.
- Representar árboles que analicen las categorías gramaticales aplicadas a la descomposición de sus partes.

- Implementar árboles sintácticos o árboles de parsing que representen la estructura sintagmática de oraciones de la lengua española.
- Identificar la transitividad de frases para responder a diferentes preguntas.

3. Producto

3.1. Definición de producto

Mediante el paradigma lógico con técnicas de producción interpretada “Prolog”, se pretende diseñar e implementar un prototipo que permita realizar procesamiento de lenguaje natural, siendo en este caso el vocabulario español. La idea es que mediante el uso de gramáticas libres de contexto y árboles de análisis sintáctico, se pueda definir la estructura correcta del lenguaje ya mencionado, caracterizando principalmente las partes que se consideren más importantes de una oración como por ejemplo: artículo, sujeto, verbo y predicado. A su vez es importante considerar el uso de gramáticas con cláusulas definidas que permitan verificar el sentido de la frase, es decir, que pueda determinar, por ejemplo, de manera correcta la oración “Los niños comen helado” y de forma incorrecta “Los comen helado”. Es importante además que las gramáticas con cláusulas definidas tomen en cuenta la concordancia de género y número, restringiendo de este modo oraciones del tipo “Los niño come helado”. Una vez teniendo definidos en el prototipo dichos términos o gramáticas es de suma importancia incorporar razonamiento de nuestro lenguaje natural, esto con el fin de lograr la meta propuesta del prototipo, que es emplear transitividad entre frases definidas, logrando que el programa sea capaz de responder o procesar interrogantes del usuario sobre una o varias oraciones ya definidas.

3.2. Planificación del producto

4. Conceptualización

- 4.1. Lenguaje: Es un conjunto infinito de símbolos de algún alfabeto que en su combinación correcta tiene el fin de formar frases. La correctitud de una frase radica en el formalismo de utilizar gramáticas.
- 4.2. Procesamiento de lenguaje natural: El lenguaje empleado por los seres humanos se basa en el contexto y la experiencia, lo cual resulta que en general sea muy complicado de procesar por una máquina. Procesar lenguaje natural por parte de cualquier ordenador implica de tres fases: análisis léxico, análisis sintáctico y análisis semántico. En Prolog, por ejemplo, se da de manera restringida, tiene su origen en la lógica de reglas, sin embargo, es el más apropiado por ser diseñado desde sus inicios como traductor del lenguaje natural.

- 4.3. Análisis léxico: Lectura de un conjunto de caracteres, agrupándolos de manera significativa, mediante etiquetas o tokens para facilitar el proceso de traducción en las siguientes fases.
- 4.4. Análisis sintáctico: Define una especificación de las reglas gramaticales del lenguaje empleado, es decir, describe la forma correcta del lenguaje.
- 4.5. Análisis semántico: Verifica la consistencia semántica en base a la definición gramatical del lenguaje, mediante la definición del significado de la frase, es decir, corrobora que la frase tenga un significado o sentido correcto.
- 4.6. Gramáticas libres de contexto GLC: Son un conjunto de reglas que dan estructura correcta a un lenguaje específico. Sus constituyentes deben de ser independientes uno sobre el otro. Poseen conjunto de objetos terminales y no terminales, que son unidos mediante producciones para crear objetos más complejos. Por ejemplo, considere $G \rightarrow a$, donde “G” es un símbolo no terminal y “a” la unión de símbolos no terminales y terminales, dando origen a lo que llamamos vocabulario.
- 4.7. Gramáticas de cláusulas definidas: Son gramáticas que permiten extender las GLC, manejando el orden de las sentencias. Permite a los símbolos no terminales poseer argumentos con dependencia recíprocas entre los componentes que forman una frase.
- 4.8. Árboles de parsing: Es la representación de secuencia de derivaciones, por el cual, dada una frase, se debe descomponer en sus constituyentes sintácticos, es decir, los componentes de una frase se estructuran en diferentes categorías sintácticas. Posee nodos internos que representan los símbolos no terminales, hojas que constituyen a los símbolos terminales.
- 4.9. Eliminación de ambigüedades: Después de la fase de análisis semántico, se procede a escoger uno de los posibles significados mediante la eliminación de ambigüedades.
- 4.10. Razonamiento con lenguaje natural:

5. Representación e inferencia

6. Casos de uso

7. Desarrollo

8. Validación y Plan futuro

9. Referencias Bibliográficas

[1] Hutchins, W. & somers, H. (1992). *Introduction to Machine Translation: table of contents*
Recuperado el 30 de Abril del 2015, de <http://www.hutchinsweb.me.uk/IntroMT-TOC.htm>

[2] S. (2003). *Controlling Controlled English An Analysis of Several Controlled Language Rule Sets*
Consultado el 30 de Abril del 2015, de <http://www.mt-archive.info/CLT-2003-Obrien.pdf>

[3] Elina Lagoudaki (2006), "Translation Memory systems: Enlightening users' perspective. Key finding of the TM Survey 2006 carried out during July and August 2006. (Imperial College London, Translation Memories Survey 2006)

[4] Aho, A. (2008). *Compiladores: Principios, técnicas y herramientas*. México : Pearson.

[5] Bratko, I. *Prolog Programming for Artificial Intelligence: Cap 21: "Language Processing with Grammar Rules"* Pearson.