

Informe Técnico: Análisis Sintáctico de Oraciones Simples en Prolog utilizando Gramáticas DCG

1. Introducción.

El presente informe describe el funcionamiento y estructura de un programa escrito en Prolog que implementa una **gramática libre de contexto definida mediante reglas DCG (Definite Clause Grammar)**. El objetivo principal del programa es analizar oraciones simples en español, construyendo su **estructura gramatical en forma de árbol sintáctico**.

Este tipo de herramienta resulta útil para tareas de análisis sintáctico en el ámbito del procesamiento de lenguaje natural (PLN), especialmente en la etapa de reconocimiento y segmentación gramatical.

2. Objetivo.

El programa tiene como finalidad:

- Reconocer oraciones gramaticalmente correctas en español (limitadas a estructuras simples).
- Validar la concordancia de género y número entre determinantes y sustantivos.
- Distinguir entre verbos transitivos e intransitivos.
- Construir la **estructura de árbol sintáctico** de cada oración reconocida.

3. Estructura del Programa

El programa se compone de tres partes fundamentales:

A. Léxico

Se definen listas de palabras agrupadas por categorías léxicas, junto con sus propiedades de **género y número**:

- **Determinantes (det/3):** Incluyen artículos definidos e indefinidos (ej. "el", "una").
- **Sustantivos (noun/3):** Referentes a empleados, empleadas y sueldos.
- **Verbos intransitivos (iv/2):** Como “trabaja”, “trabajan”.

- **Verbos transitivos (tv/2):** Como “cobra”, “cobran”.

Esto permite que la gramática valide la concordancia morfosintáctica.

B. Gramática (DCG)

La gramática está escrita utilizando notación DCG, una extensión de Prolog para definir reglas de producción similares a una gramática libre de contexto.

Reglas principales:

- `o_tree/3`: Produce un árbol sintáctico completo con un **sintagma nominal (SN)** seguido de un **sintagma verbal (SV)**.
- `sn/3`: Reconoce un sintagma nominal compuesto por un determinante y un sustantivo, verificando concordancia de género y número.
- `sv/2`: Reconoce un sintagma verbal. Puede ser:
 - **Intransitivo**: un solo verbo (trabaja, trabajan).
 - **Transitivo**: un verbo más un objeto directo (cobra el sueldo).

Reglas auxiliares:

Estas conectan las palabras en la oración con su representación léxica:

`det(W,Gen,Num) --> [W], { det(W,Gen,Num) }.`

`noun(W,Gen,Num) --> [W], { noun(W,Gen,Num) }.`

`iv(W,Num) --> [W], { iv(W,Num) }.`

`tv(W,Num) --> [W], { tv(W,Num) }.`

C. Entradas y análisis

El programa puede analizar oraciones como:

- "el empleado trabaja"
- "una empleada cobra el sueldo"
- "los empleados cobran sueldos"

Y lo hace utilizando consultas como:

?- phrase(o(Tree), [el, empleado, trabaja]).

Tree = o(sn(det(el), n(empleado)), sv(vt(trabaja))).

También permite validar estructuras gramaticales:

?- phrase(o, [la, empleados, trabaja]).

false. % Falla por concordancia incorrecta

4. Resultados

El sistema reconoce y analiza correctamente oraciones simples que siguen las reglas gramaticales establecidas. Devuelve una estructura tipo árbol con la descomposición gramatical de cada parte de la oración.

Esto permite extender fácilmente el análisis a tareas más complejas como traducción automática, generación de árboles para PLN, y construcción de parsers naturales para el español.

5. Aplicaciones

Este tipo de sistema tiene aplicaciones en:

- **Procesamiento de Lenguaje Natural (PLN).**
- **Desarrollo de analizadores sintácticos automáticos.**
- **Sistemas de corrección gramatical.**
- **Asistentes de escritura o lectura automatizada.**
- **Construcción de chatbots lingüísticamente coherentes.**

6. Conclusión

El programa representa un ejemplo funcional y extensible de un parser sintáctico simple en Prolog. Aprovecha el poder expresivo de las DCG para modelar la estructura del lenguaje humano, y permite construir árboles sintácticos a partir de oraciones correctas, respetando la concordancia de género y número.

Gracias a su modularidad, puede ser fácilmente adaptado para incorporar nuevos sustantivos, tiempos verbales, adjetivos o estructuras gramaticales más complejas, avanzando hacia una comprensión más robusta del lenguaje natural.