

Introducción a las Gramáticas de Cláusulas Definidas (DCGs)

Procesamiento del Lenguaje Natural

José María Gómez Hidalgo

<http://www.esp.uem.es/jmgomez/>

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Índice

- Representación del conocimiento sintáctico
- Gramáticas de estructura de frase independientes del contexto
- Gramáticas de cláusulas definidas

Representación del conocimiento sintáctico

- **Naturaleza del conocimiento sintáctico**
 - Estructura de las expresiones de un lenguaje
 - Si es un lenguaje natural, estructura de las oraciones, es decir, cómo se relacionan las palabras de una oración entre sí
 - Cómo se organizan las palabras en grupos o sintagmas
 - Qué palabras o grupos modifican a otras palabras o grupos
 - Qué palabras o grupos son los más importantes de la oración

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Representación del conocimiento sintáctico

- **Ejemplo**
 - Oración: "Juan compra una novela en la librería"
 - "En la librería" es un sintagma preposicional que modifica al núcleo de la oración, que es el verbo "compra"
 - Agrupamiento (sintagma preposicional)
 - Modificación (modifica al verbo)
 - Importancia (el verbo es el núcleo)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Representación del conocimiento sintáctico

- Tipos de conocimiento sintáctico
 - Formalismo gramatical
 - Lenguaje de representación de información sintáctica
 - Gramática - G
 - Representación de la estructura de un lenguaje, en un formalismo gramatical
 - Árbol de análisis
 - Representación de la estructura de una oración concreta de un lenguaje concreto (caracterizado por una gramática expresada en un formalismo)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

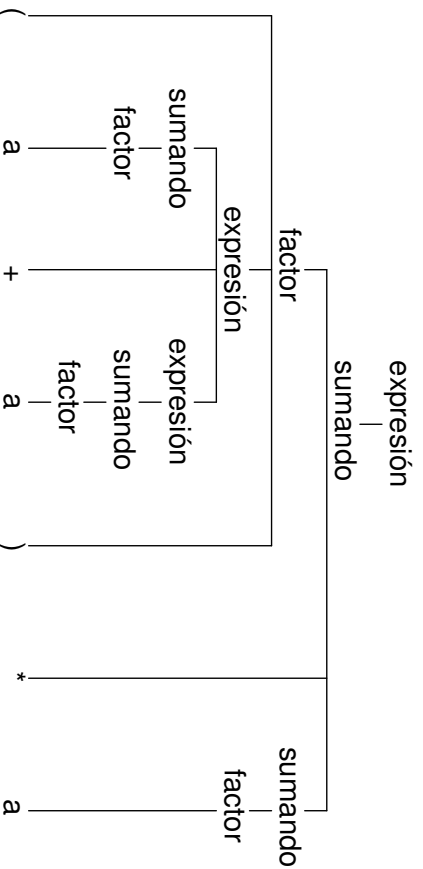
Representación del conocimiento sintáctico

- Ejemplo
 - Para caracterizar el lenguaje de las expresiones matemáticas sencillas, como " $(a + a) * a$ "
 - *Formalismo gramatical*: Gramáticas independientes del contexto
 - *Gramática*:
 - expresión \rightarrow sumando + expresión
 - expresión \rightarrow sumando
 - sumando \rightarrow factor * sumando
 - sumando \rightarrow factor
 - factor $\rightarrow a$
 - factor \rightarrow (expresión)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Representación del conocimiento sintáctico

- Ejemplo
 - *Árbol de análisis de la expresión "(a + a) * a"*



Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Representación del conocimiento sintáctico

- Análisis sintáctico y traducción
 - Dada la gramática G que caracteriza al lenguaje L , algunas expresiones pertenecen a L y otras no
 - **Reconocimiento** = decidir si una expresión pertenece al lenguaje caracterizado por G
 - **Análisis (*parsing*)** = decidir si una expresión pertenece al lenguaje representado por G , y en caso afirmativo, asignar a la expresión una estructura acorde a G
 - Análisis = traducción de expresiones en lenguaje natural a estructuras sintácticas

Representación del conocimiento sintáctico

- **Características deseables en los formalismos gramaticales**
 - **Naturalidad lingüística**
 - Conceptos lingüísticos fácilmente expresables
 - **Poder expresivo**
 - Capacidad de representar lenguajes (matemáticamente complejos)
 - **Efectividad computacional**
 - Gramática leída por hombre y máquina
 - Diseño de lenguajes formales

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Representación del conocimiento sintáctico

- **Características deseables en las gramáticas**
 - **General**
 - Debe cubrir el mayor subconjunto posible del idioma
 - **Selectiva**
 - Debe minimizar el número de problemas que encuentra en la oraciones no válidas
 - **Comprensible**
 - Debe ser lo más sencilla posible

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de estructura de frase independientes del contexto

- Gramática de estructura de frase independiente del contexto, CF-PSG (Context-free phrase structure grammar)
- Una tupla $G = (T, NT, S, P)$
 - T es un conjunto finito de símbolos terminales o léxicos
 - NT es un conjunto finito disjunto de T de elementos no terminales o categorías
 - P es un conjunto finito de producciones o reglas de reescritura de la forma $A \rightarrow a$ donde A es una cat. y a es una secuencia de categorías y símbolos léxicos
 - S es un símbolo de NT llamado axioma o símbolo inicial, y no aparece en lado derecho de ninguna regla

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de estructura de frase independientes del contexto

- Ejemplo (G1)
 - $T = \{el, la, perro, salchicha, come\}$
 - $NT = \{O, SN, SV, V, Art, N\}$
 - $S = O$
 - $P = \{O \rightarrow SN\ SV, SN \rightarrow Art\ N, SV \rightarrow V\ SN, Art \rightarrow el, Art \rightarrow la, V \rightarrow come, N \rightarrow perro, N \rightarrow salchicha\}$

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de estructura de frase independientes del contexto

- Una CF-PSG G sirve para
 - Definir un conjunto de frases (subconjuntos de T^*) aceptables en un lenguaje o gramaticales = el lenguaje generado (representado) por la gramática G , $L(G)$
 - Asignar una $o +$ estructuras sintácticas a frases gramaticales de la gramática, frases de $L(G)$
 - Las estructuras asignadas son árboles o "grafos acíclicos dirigidos" (Directed acyclic graphs, DAGs)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

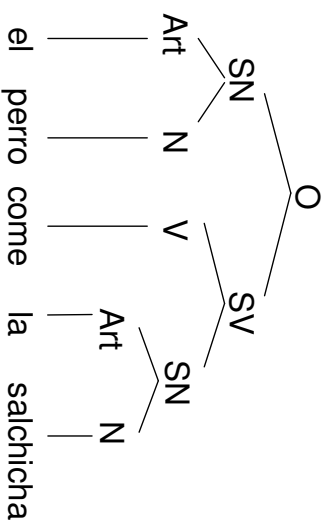
Gramáticas de estructura de frase independientes del contexto

- Ejemplo
 - La frase $F =$ "el perro come la salchicha" es aceptada por G_1 , es decir, $F \in L(G_1)$
 - A la frase F se le puede asignar el árbol siguiente
 - Representado como lista:
 - (O (SN (Art el) (N perro))) (SV (V come) (SN (Art la) (N salchicha))))

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de estructura de frase independientes del contexto

- Ejemplo
 - Árbol representado gráficamente



Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de estructura de frase independientes del contexto

- Un árbol de análisis representa dos tipos de relaciones entre símbolos
 - Dominio
 - Una categoría domina a otra o a un símbolo léxico cuando es su nodo padre en el árbol de análisis
 - Ej. Art domina a "el", y SV domina a V
 - Precedencia
 - Un símbolo precede a otro cuando es un hermano suyo a la izquierda en el árbol de análisis
 - Ej. "el" precede a "perro", y V precede a SN

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de estructura de frase independientes del contexto

- Sobregeneración
 - La gramática puede generar frases u oraciones no deseables
 - G1 genera "la salchicha come el perro"
 - Puede no ser problema si se pretende reconocer frases, pero lo es si se pretende producir frases

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de estructura de frase independientes del contexto

- Infrageneración
 - La gramática puede no ser capaz de generar oraciones deseables
 - G1 no genera "el perro comió la salchicha"
 - Puede no ser problema para producir frases, pero lo es para reconocer frases

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de estructura de frase independientes del contexto

- Asignación de estructuras correctas
 - La gramática puede no asignar las estructuras deseadas por el desarrollador de la misma
 - Si se pretende asignar la estructura de función (p. ej., SN Objeto o SN Sujeto), G1 no es capaz de hacerlo
 - Las estructuras dependen de la lingüística, donde no hay acuerdos definitivos

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Descripción general
- Notación
- Uso
- Análisis gramatical

Gramáticas de cláusulas definidas

- Definite Clause Grammars (DCGs)
- Extensión natural de las CF-PSGs
- Mejoran a las CF-PSGs
 - Proporcionan dependencia del contexto
 - Construcción de estructuras
 - Sintácticas
 - Semánticas
 - Condiciones adicionales => cálculos auxiliares

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Heredan las propiedades de las CF-PSGs
 - Claridad y modularidad
 - Recursividad
 - Sobre todo relativos a algoritmos de análisis
 - Resultados teóricos
- Básicamente CF-PSGs con extensiones
- Azúcar sintáctico sobre Prolog

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Notación básica
 - Símbolos con minúsculas
 - Símbolos léxicos entre corchetes (listas)
 - Símbolos léxicos con mayúscula entre comillas simples
 - Reglas incluyen "-->", " y "."
- Ejemplo: DCG31 (G31)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Extensiones de la notación
 - Inclusión de variables en las categorías
 - Primera letra con mayúsculas
 - Llamadas a Prolog
 - Entre llaves
- Ejemplo
 - nombre(N) --> [Palabra], {raiz(Palabra,N),es_nombre(N)}.
 - Significa que la palabra Palabra es un nombre N si la raíz de Palabra es N y N es un nombre.

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Usos de las DCGs
 - Dependencia del contexto
 - Concordancia
 - Restricciones selectivas
 - Construcción de estructuras
 - Sintácticas
 - Semánticas
 - Realización de cómputos auxiliares
 - Búsqueda en el diccionario y análisis morfológico

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Concordancia
 - Es una dependencia contextual en **sentido lingüístico**
 - No es una dependencia contextual en sentido de **poder expresivo**
 - Ej. "los hombres"
 - "hombres" tiene que concordar con el número del contexto marcado por "los"
- Se puede tratar con CF-PSGs

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- La gramática
- se convierte en

sintagma_nominal →
artículo nombre
artículo → el
artículo → los
nombre → hombre
nombre → hombres

sintagma_nominal →
artículo_singular
nombre_singular
sintagma_nominal → artículo_plural
nombre_plural
artículo_singular → el
artículo_plural → los
nombre_singular → hombre
nombre_plural → hombres

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- **Concordancia**
 - Se simplifica el tratamiento con DCGs
 - Ejemplo
 - sintagma_nominal --> artículo(Numero), nombre(Numero).
 - artículo(singular) --> [e].
 - artículo(plural) --> [los].
 - nombre(singular) --> [hombre].
 - nombre(plural) --> [hombres].

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- **Restricciones selectivas**
 - Se refieren a dependencias semánticas (actor, instrumento = casos)
 - Imponen chequeos semánticos para disminuir
 - el número de árboles sintácticos generados
 - el grado de regeneración de una gramática
 - Por ejemplo, G31 reconoce "la salchicha come el perro"
=> se convierte en DCG32 que no reconozca la expresión
- Ejemplos: DCG32 y DCG33

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- **Construcción de estructuras**
 - Sintácticas
 - árbol de análisis
 - Semánticas
 - representación del significado
 - forma lógica

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Construcción del árbol de análisis
 - La estructura representada con términos
 - Ejemplo: "El perro come la salchicha" en G31

o(sn(art(el),
n(perro)),
sv(v(come),
sn(art(la),
n(salchicha))))

- Ejemplo: DCG34

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Construcción de la forma lógica (DCG35)
 - Modificamos G31 => G33

O → SN SV
SN → Det N SV → V SN
Det → un Det → todo
Det → una Det → toda
V → come
N → perro N → salchicha

Gramáticas de cláusulas definidas

– Forma lógica

- Queremos representar la frase "todo perro come una salchicha" como
$$\forall X (\text{perro}(X) \rightarrow \exists Y (\text{salchicha}(Y) \rightarrow \text{come}(X, Y)))$$
- Los **nombres** se representan por medio de predicados, propiedades sobre variables lógicas
- Los **determinantes** se representan por medio de cuantificadores
- Los **verbos** se representan por medio de predicados

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

– Notación de las fórmulas

- \forall y \exists son operadores binarios \Rightarrow todo(X, P(X)) y existe(X, P(X)) para algún predicado P
- \rightarrow es un operador binario \Rightarrow implica(P, Q)
- Ej. La fórmula anterior sería

todo(X,
 implica(perro(X),
 existe(Y,
 implica(salchicha(Y),
 come(X, Y))))))

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Cómputos auxiliares
 - Ejemplo
 - nombre(N) --> [Palabra], {raiz(Palabra,N), es_nombre(N)}.
 - Se obtiene la raíz de la palabra por medio de raiz/2
 - Se busca en el diccionario una raíz por medio de es_nombre/1

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Análisis gramatical en Prolog
 - Prolog convierte automáticamente las reglas a una representación interna operativa
 - DCG = azúcar sintáctico sobre Prolog
 - De manera inmediata, DCG = analizador / generador

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- Representación interna similar a los grafos del análisis basado en diagramas
 - Estado = par de listas diferencia
 - Ejemplos
 - oracion --> sintagma_nominal, sintagma_verbal.
 - oracion(S0, S2) :- sintagma_nominal(S0, S1),
sintagma_verbal(S1, S2).
 - nombre(N) --> [W],{raiz((W, N), es_nombre(N))}.
 - nombre(N, [W|X], X) :- raiz(W, N), es_nombre(N).

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- DCG como programa
 - Invocación = objetivo
 - Se puede usar como analizador / generador
 - El tipo de uso está limitado por el comportamiento de los predicados extragramaticales
 - Ejemplo:
 - oracion(o(SN, SV)) --> sintagma_nominal(SN),
sintagma_verbal(SV).

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- **Uso como analizador**
 - oracion(X, [aquí, va, la, lista, de, entrada],[[]]).
 - X toma como valor con el árbol de análisis si la entrada es gramatical, o se produce fallo
 - Por cada valor de X, un árbol de análisis

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- **Uso como generador**
 - oracion(o(sn(...)), X,[[]]).
 - X toma como valor una oración cuyo árbol de análisis es el primer argumento
 - Esto es posible porque dado un significado, hay varias formas de realizarlo superficialmente (con activa, pasiva, elipsis, etc.)
 - Se puede especificar parcialmente el árbol

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- DCG como analizador
 - Tipo de análisis = descendente, de izquierda a derecha y con vuelta atrás, es decir, similar a ADS => ineficiente
 - Ventaja: depuración de la gramática más rápida
 - Se gana en tiempo de desarrollo y se pierde en ejecución (preferible, hardware más barato que software)

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid

Gramáticas de cláusulas definidas

- DCG como analizador
 - Problema en reglas con recursión por la izquierda
 - ¡¡¡Evitarlas!!!
 - Pasar a recursión por la derecha
 - Gestionar con argumentos y otros predicados la construcción de las estructuras deseables
 - Predicado auxiliar de ayuda
 - Para gramáticas con símbolos sin argumentos
phrase(P, L) :- Goal =.. [P,L,[]], call(Goal).

Procesamiento del Lenguaje Natural – José María Gómez Hidalgo – U. Europea de Madrid