

# Gramáticas libres de contexto (GLC)

# GLC

- Capturan la noción de constituyente sintáctico y la noción de orden.
- Herramienta formal que puede ser vista tanto desde un punto de vista generador como estructurador.
- Propiedades computacionales interesantes: se puede reconocer en tiempo polinómico.

# GLC

Una Gramática Libre de Contexto es una tupla con 4 parámetros :

$$\mathbf{G = (V,T,P,S)}$$

- V – conjunto de símbolos variables
- T – conjunto de símbolos terminales
- $S \in V$ , símbolo inicial
- P – conjunto de reglas de producción :  
 $A \rightarrow \alpha$  , con  $\alpha$  sucesión de símbolos de  $V \cup T$ , eventualmente vacía ( $\alpha = \epsilon$ )

# GLC

- Una GLC es un dispositivo **generador**.
- Definimos el lenguaje  $L_G$  generado por una gramática  $G$  del siguiente modo:

$$L_G = \{ w / S \Rightarrow^* w \}$$

siendo  $\Rightarrow^*$  una “especie” de clausura transitiva de  $\rightarrow$   
y  $w$  una tira de terminales

# GLC para GN sencillos

*el avión*

*un perro andaluz*

*París*

# GLC para GN sencillos

GN → Det Nom

GN → Det Nom Adj

GN → NomProp

**GRAMÁTICA**

Det → el | un | la | ...

Nom → avión | perro | ...

Adj → andaluz | viejo | ...

NomProp → París | Jorge | ...

**LÉXICO**

# GLC para GN sencillos

GN → Det Nom

GN → Det Nom Adj

GN → NomProp

## GRAMÁTICA

**símbolos terminales y no  
terminales**

Det → el | un | la | ...

Nom → avión | perro | ...

Adj → andaluz | viejo | ...

NomProp → París | Jorge | ...

## LÉXICO

**es parte de la gramática**

# GLC en Computación

## Teoría de Lenguajes Formales

- $L1 = \{ w / w = a^{nb^n}, n > 0 \}$
- $L2 = \{ w / w = anbncn, n > 0 \}$
- $L3 = \{ w / w = (ab)^n, n > 0 \}$



# GLC en Computación

## Jerarquía de Chomsky

### Teoría de Lenguajes Formales

- $L1 = \{ w / w = a^n b^n, n > 0 \}$  Libres de contexto (tipo 2)
- $L2 = \{ w / w = a^n b^n c^n, n > 0 \}$  Sensibles al contexto (tipo 1)
- $L3 = \{ w / w = (ab)^n, n > 0 \}$  Regulares (tipo 3)
- *La jeraquía se completa con los lenguajes tipo 0 (gramáticas irrestrictas)*

# GLC en Computación

## Lenguajes de programación

Backus Naur, especificación de sintaxis

- $\langle \text{sentence} \rangle ::= \langle \text{if} \rangle \mid \langle \text{while} \rangle \mid \langle \text{cond} \rangle \mid \langle \text{sequence} \rangle$
- $\langle \text{while} \rangle ::= \text{while } \langle \text{cond} \rangle \text{ do } \langle \text{sequence} \rangle$

# GLC en Computación

## Lenguajes de programación

Backus Naur, especificación de sintaxis, es GLC

–  $\langle \text{sentence} \rangle ::= \langle \text{if} \rangle \mid \langle \text{while} \rangle \mid \langle \text{cond} \rangle \mid \langle \text{sequence} \rangle$

SENT  $\rightarrow$  IF

SENT  $\rightarrow$  WHILE

....

–  $\langle \text{while} \rangle ::= \text{while } \langle \text{cond} \rangle \text{ do } \langle \text{sequence} \rangle$

WHILE  $\rightarrow$  while COND do SEQ

# Ejercicio

*Cuando salí de Monte con un currículum abusivamente sobresaliente y bajo el brazo un recién nacido título de ingeniero, el profesor Paley estaba a mi lado y no me abandonó hasta que pisamos Santamaría.*

Identificar los grupos nominales (GN) contenidos en el texto anterior y escribir una GLC que los reconozca.

# Ejercicio

*Monte*

*un currículum abusivamente sobresaliente  
el brazo*

*un recién nacido título de ingeniero  
ingeniero*

*el profesor Paley  
mi lado*

*me*

*Santamaría*

# Ejercicio: una solución

GN → Nom  
GN → NomProp  
GN → Det Nom  
GN → Det GAdj Nom  
GN → Det Nom GAdj  
GN → GN GP  
GN → GN NomProp  
GN → Pron

GAdj → Adj  
GAdj → Adv Adj

GP → Prep GN

**GRAMÁTICA**

Det → el | un | mi

Nom → currículum | brazo | título |  
ingeniero | profesor | lado

Adj → sobresaliente | nacido

NomProp → Monte | Paley |  
Santamaría

Adv → abusivamente | recién

Prep → de

Pron → me

**LÉXICO**

# La oración

Varias formas posibles:

- Aseverativa: *Julio vende la casa.*
- Imperativa: *Traeme ya el informe.*
- Interrogativa: *¿Podés pasarme la sal?*
- Exclamativa: *¡Cuánto ha crecido!*
- Desiderativa: *Ojalá salga el sol.*
- Dubitativa: *Quizá venga María.*

# Oración aseverativa

*Los niños juegan.*

*Julio vende la casa.*

*Compré un libro de sintaxis.*

*Compré un libro en la librería.*

*El resultado de la interpelación permanece incierto.*

*El hombre llegó muy tranquilamente a su casa.*



# Oración aseverativa

[Los niños]<sub>GN</sub> [juegan]<sub>GV</sub>.

[Julio]<sub>GN</sub> [vende la casa]<sub>GV</sub>.

[Compré un libro de sintaxis]<sub>GV</sub>.

[Compré un libro en la librería]<sub>GV</sub>.

[El resultado de la interpelación]<sub>GN</sub> [permanece incierto]<sub>GV</sub>.

[El hombre]<sub>GN</sub> [llegó muy tranquilamente a su casa]<sub>GV</sub>.

# Oración aseverativa

[Los niños]<sub>GN</sub> [juegan]<sub>GV</sub>.

[Julio]<sub>GN</sub> [vende la casa]<sub>GV</sub>.

[Compré un libro de sintaxis]<sub>GV</sub>.

[Compré un libro en la librería]<sub>GV</sub>.

[El resultado de la interpelación]<sub>GN</sub> [permanece incierto]<sub>GV</sub>.

[El hombre]<sub>GN</sub> [llegó muy tranquilamente a su casa]<sub>GV</sub>.

# Reglas para oraciones aseverativas (esbozo)

$O \rightarrow GN\ GV$

$O \rightarrow GV$

$GV \rightarrow V$

$GV \rightarrow V\ GN$

$GV \rightarrow V\ GAdj$

$GV \rightarrow V\ GAdv$

$GV \rightarrow GV\ GP$

# Dos posibles notaciones para la estructura

Para representar el análisis que resulta de la aplicación de las reglas utilizamos:

- Notación con paréntesis rectos (bracketed notation)
- Árboles de análisis sintáctico

# Dos posibles notaciones para la estructura

Ejemplo:

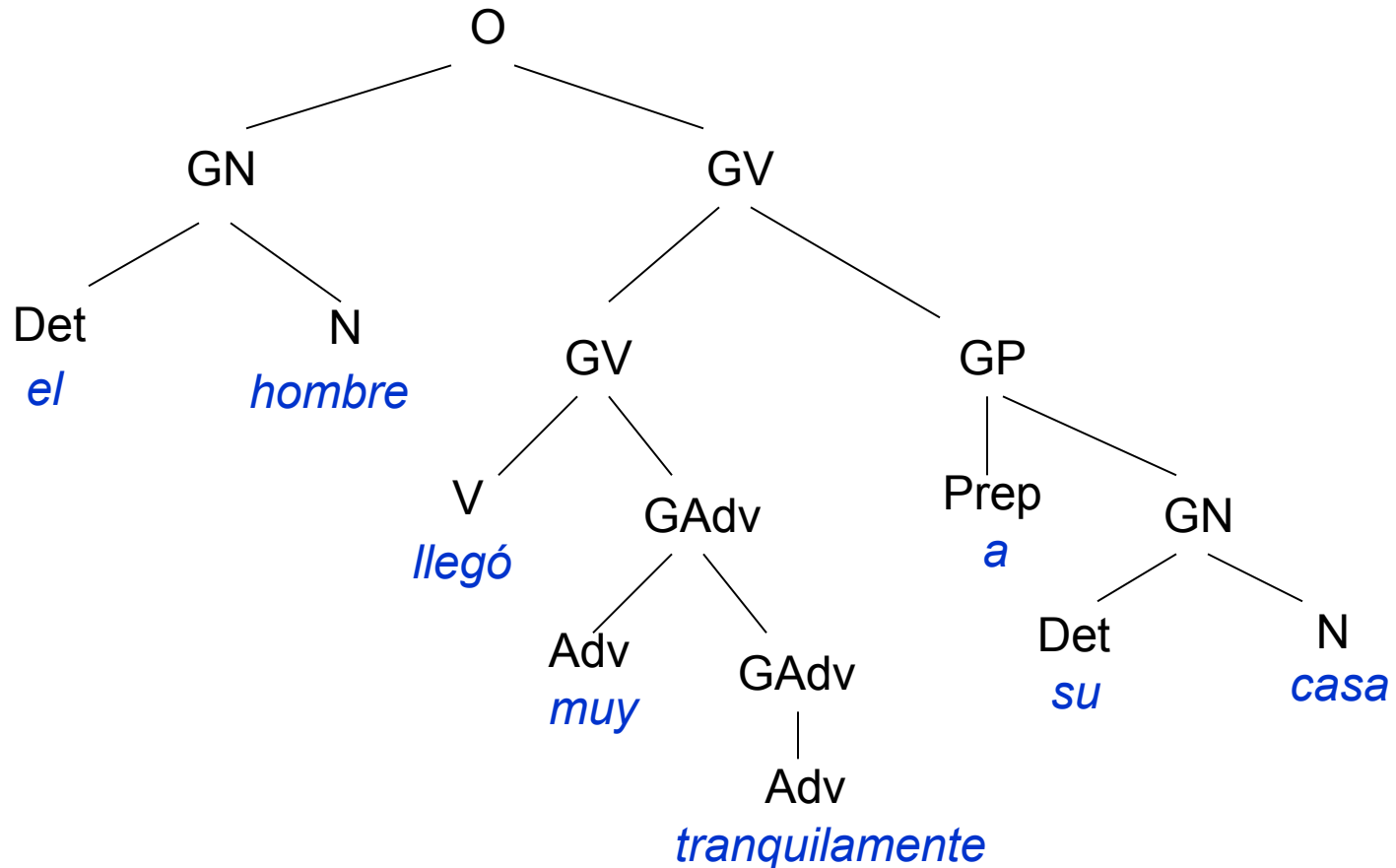
*El hombre llegó muy tranquilamente a su casa.*

# Dos posibles notaciones para la estructura

[El hombre]<sub>GN</sub> [ llegó [ muy [ tranquilamente ] ]<sub>GAdv</sub> [ a [ su casa ]<sub>GN</sub> ]<sub>GP</sub> ]<sub>Gv</sub>.

# Dos posibles notaciones para la estructura

*El hombre llegó muy tranquilamente a su casa.*



# Coordinación

*Cleo y Manu leen a Herrera y Reissig.*

*Mandame un capuccino y una medialuna.*

*Valentina come y llora al mismo tiempo.*

*El hombre encendió el auto y arrancó sin mirar si venía alguien.*



# Reglas para coordinación

O → O ConjCoord O

GN → GN ConjCoord GN

GV → GV ConjCoord GV

ConjCoord → y | o | pero | ...

# Algunos problemas

GN  $\rightarrow$  Det Nom

Det  $\rightarrow$  el | la | los | las

Nom  $\rightarrow$  libro | libros | casa | casas

¿Qué GNs se pueden generar?

# Algunos problemas

## Sobregeneración

Una solución posible: partir los símbolos

Det --- DetFS , DetMS, Det FP, DetMP

Nom --- NomFS, .....

GN → DetFD NomFS | ...

# Algunos problemas

*Juan quiere salir.*

*Juan comió pizza.*

Pero no : *Juan comió salir*

**Subcategorización :**

un verbo determina la sintaxis posible de sus complementos

# Problemas con el enfoque GLC

- Concordancia - multiplicidad de reglas
  - GN Masc Sing → Det Masc Sing Nom Masc Sing  
(x 4 + “el agua”)
- Subcategorización
  - sintáctica:

*Juan quiere salir. // \* Juan comió + Vinf*

No todas las combinaciones de categorías gramaticales son posibles, las reglas deben preverlo.
  - semántica

*Juan dijo la verdad. // \* Juan dijo la mesa.*  
*Presenció el incendio. // \* Presenció la silla.*

# Características del enfoque GIC

- Reconocimiento eficiente

pero ...

- Poca adecuación lingüística: su **capacidad generativa débil** (conjunto de tiras generadas / aceptadas) es adecuada pero su **capacidad generativa fuerte** (descripciones estructurales que se generan) no.
- Dificultad para expresar adecuadamente algunas propiedades.

# Características del enfoque GLC

- Es un formalismo general, lingüísticamente neutro.
- No hay por qué respetar criterios lingüísticos sobre constituyentes.
- La construcción manual de grandes gramáticas independientes de contexto no fue una solución tecnológica adecuada.

# Referencias

Jurafsky, D. y J. Martin. 2009. *Speech and Language Processing*. Second edition. Capítulo 12.