

Ciencia de Datos y Lenguaje Natural

Clase 2.2- Análisis Sintáctico

Grupo PLN - INCO

Universidad de la República

16 de agosto de 2022

Qué es el análisis sintáctico

El análisis sintáctico construye otro tipo de estructura al interior de la oración.

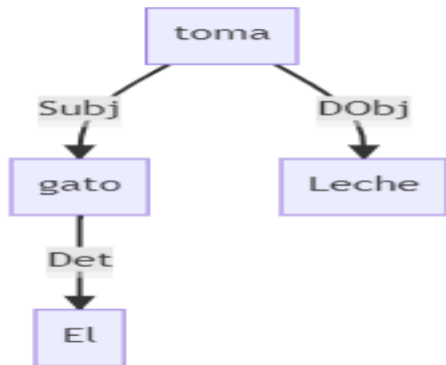
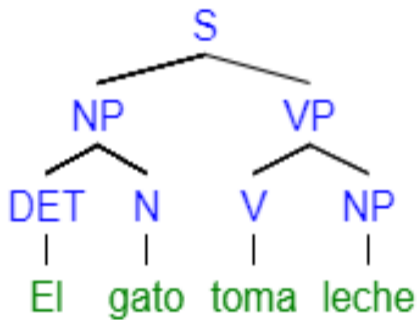
- ▶ O bien agrupa palabras formando *constituyentes*

Qué es el análisis sintáctico

El análisis sintáctico construye otro tipo de estructura al interior de la oración.

- ▶ O bien agrupa palabras formando *constituyentes*
- ▶ O bien vincula palabras entre sí formando un *árbol de dependencias*

Árbol de constituyentes y de dependencias



Comparación dependencias y constituyentes

- ▶ El árbol de constituyentes es mucho más grande

Comparación dependencias y constituyentes

- ▶ El árbol de constituyentes es mucho más grande
- ▶ Quedan menos en evidencia algunas relaciones centrales

Comparación dependencias y constituyentes

- ▶ El árbol de constituyentes es mucho más grande
- ▶ Quedan menos en evidencia algunas relaciones centrales
- ▶ El árbol de constituyentes es más difícil de construir

Comparación dependencias y constituyentes

- ▶ El árbol de constituyentes es mucho más grande
- ▶ Quedan menos en evidencia algunas relaciones centrales
- ▶ El árbol de constituyentes es más difícil de construir
- ▶ Otras aplicaciones utilizan el análisis de dependencias

Aplicaciones del análisis sintáctico

- ▶ Extracción de información

Aplicaciones del análisis sintáctico

- ▶ Extracción de información
- ▶ Corrección ortográfica

Aplicaciones del análisis sintáctico

- ▶ Extracción de información
- ▶ Corrección ortográfica
- ▶ Complejidad de textos, simplificación

Definición de árbol de dependencias

Es un grafo dirigido con palabras en los nodos y relaciones en las aristas, que satisface:

- ▶ Todo nodo, excepto la raíz, tiene una única arista entrante.

Definición de árbol de dependencias

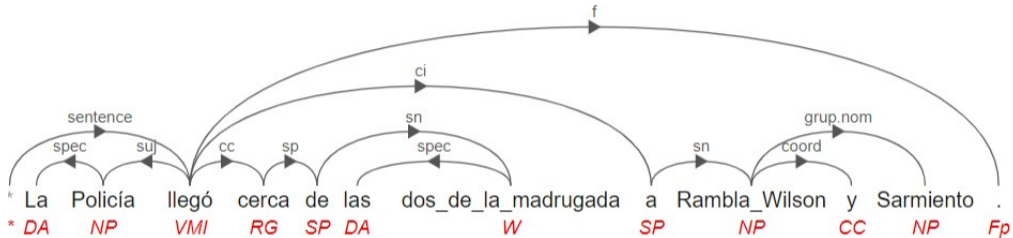
Es un grafo dirigido con palabras en los nodos y relaciones en las aristas, que satisface:

- ▶ Todo nodo, excepto la raíz, tiene una única arista entrante.
- ▶ Existe un único nodo sin aristas entrantes (la raíz).

Definición de árbol de dependencias

Es un grafo dirigido con palabras en los nodos y relaciones en las aristas, que satisface:

- ▶ Todo nodo, excepto la raíz, tiene una única arista entrante.
- ▶ Existe un único nodo sin aristas entrantes (la raíz).
- ▶ Hay un camino único desde la raíz a cada nodo.



▼ CONLL format

1	La	el	DA0FS0	DA	-	-	-	-	2	spec	-	-
2	Policía	policía	NP00000	NP	-	B-ORG	-	-	3	subj	-	-
3	llegó	llegar	VMIS3S0	VMI	-	-	-	-	0	sentence	-	-
4	cerca	cerca	RG	RG	-	-	-	-	3	cc	-	-
5	de	de	SP	SP	-	-	-	-	4	sp	-	-
6	las	el	DA0FP0	DA	-	-	-	-	7	spec	-	-
7	dos_de_la_madrugada	[?:?:/?:?:?:2.00:am]	W	W	-	-	-	-	5	sn	-	-
8	a	a	SP	SP	-	-	-	-	3	ci	-	-
9	Rambla_Wilson	rambla_wilson	NP00G00	NP	-	B-LOC	-	-	8	sn	-	-
10	y	y	CC	CC	-	-	-	-	9	coord	-	-
11	Sarmiento	sarmiento	NP00G00	NP	-	B-LOC	-	-	9	grup.nom	-	-
12	.	.	Fp	Fp	-	-	-	-	3	f	-	-

Relaciones binarias asimétricas

- ▶ Es uno de los postulados de la gramática de dependencias

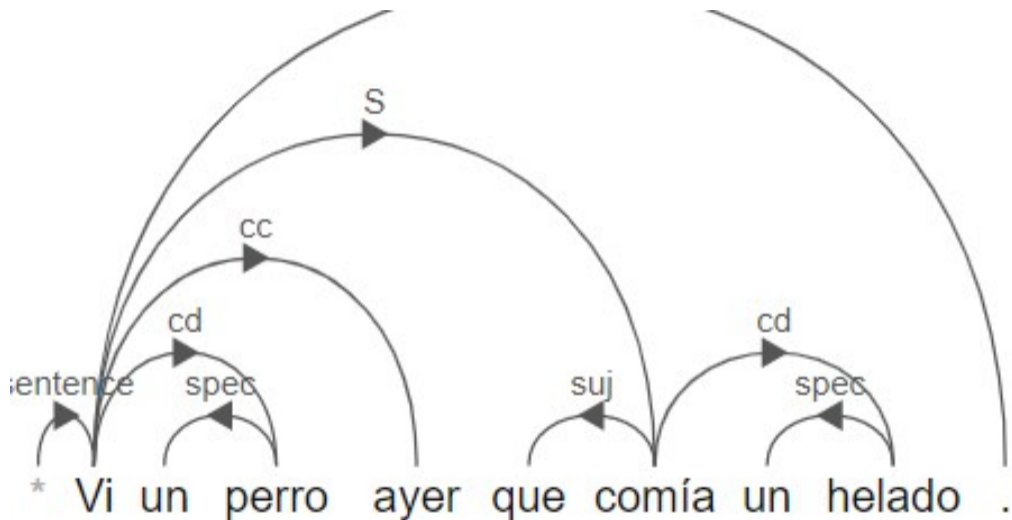
Relaciones binarias asimétricas

- ▶ Es uno de los postulados de la gramática de dependencias
- ▶ La coordinación es un caso problemático

Relaciones binarias asimétricas

- ▶ Es uno de los postulados de la gramática de dependencias
- ▶ La coordinación es un caso problemático
- ▶ Se resuelve linealizando, dando calidad de gobernante al 1er elemento

Proyectividad



Proyectividad

- ▶ El árbol anterior no está bien.

Proyectividad

- ▶ El árbol anterior no está bien.
- ▶ Debe haber una relación entre 'perro' y 'que comía un helado'.

Proyectividad

- ▶ El árbol anterior no está bien.
- ▶ Debe haber una relación entre 'perro' y 'que comía un helado'.
- ▶ El algoritmo usual para árboles de dependencia (Nivre) solo genera árboles proyectivos.

Gramática universal

- ▶ Único conjunto de relaciones de dependencia, más de 60 lenguas.

Gramática universal

- ▶ Único conjunto de relaciones de dependencia, más de 60 lenguas.
- ▶ REF https://nlp.stanford.edu/pubs/USD_LREC14_paper_camera_ready.pdf

Gramática universal

- ▶ Único conjunto de relaciones de dependencia, más de 60 lenguas.
- ▶ REF https://nlp.stanford.edu/pubs/USD_LREC14_paper_camera_ready.pdf
- ▶ O sea, tenemos un texto segmentado en palabras y le ponemos a cada palabra la categoría indicada por el diccionario.

Evaluación

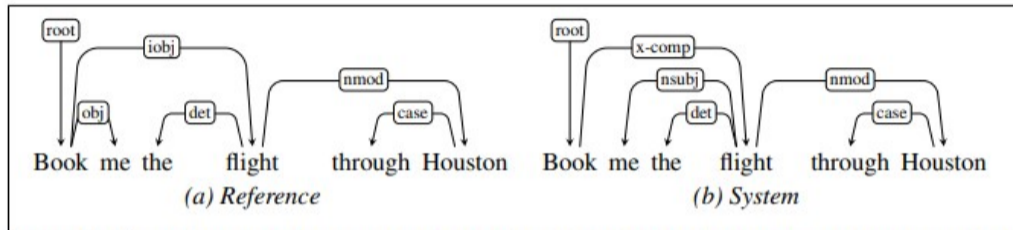


Figure 14.16 Reference and system parses for *Book me the flight through Houston*, resulting in an LAS of $2/3$ and an UAS of $5/6$.

- ▶ Se usan 2 medidas: UAS y LAS

Evaluación

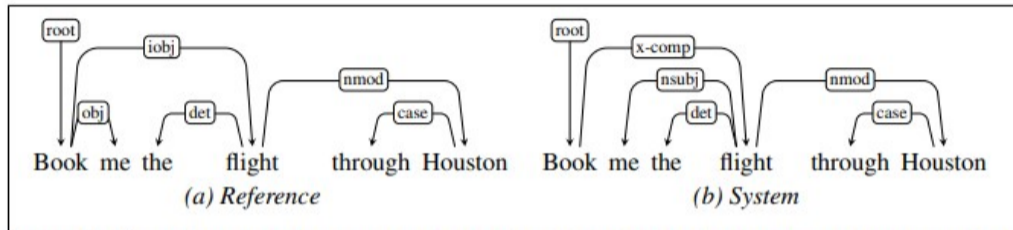


Figure 14.16 Reference and system parses for *Book me the flight through Houston*, resulting in an LAS of $2/3$ and an UAS of $5/6$.

- ▶ Se usan 2 medidas: UAS y LAS
- ▶ UAS Unlabeled attachment score porcentaje de las palabras con nodo padre correcto

Evaluación

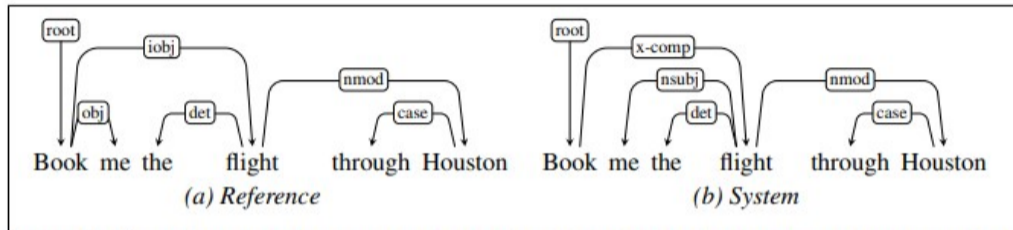


Figure 14.16 Reference and system parses for *Book me the flight through Houston*, resulting in an LAS of 2/3 and an UAS of 5/6.

- ▶ Se usan 2 medidas: UAS y LAS
- ▶ UAS Unlabeled attachment score porcentaje de las palabras con nodo padre correcto
- ▶ LAS Labeled attachment score porcentaje de las palabras con nodo padre correcto y etiqueta correcta

Evaluación

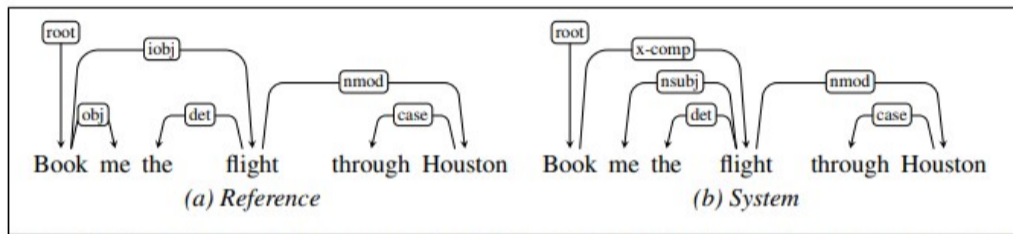


Figure 14.16 Reference and system parses for *Book me the flight through Houston*, resulting in an LAS of 2/3 and an UAS of 5/6.

- ▶ http://nlpprogress.com/english/dependency_parsing.html

Evaluación

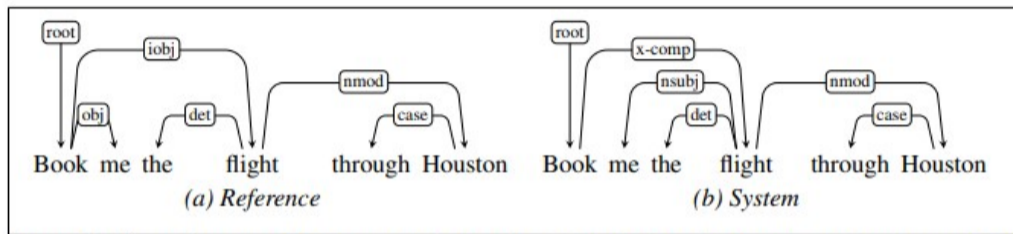


Figure 14.16 Reference and system parses for *Book me the flight through Houston*, resulting in an LAS of 2/3 and an UAS of 5/6.

- ▶ http://nlpprogress.com/english/dependency_parsing.html
- ▶ UAS Se reporta 97 %

Evaluación

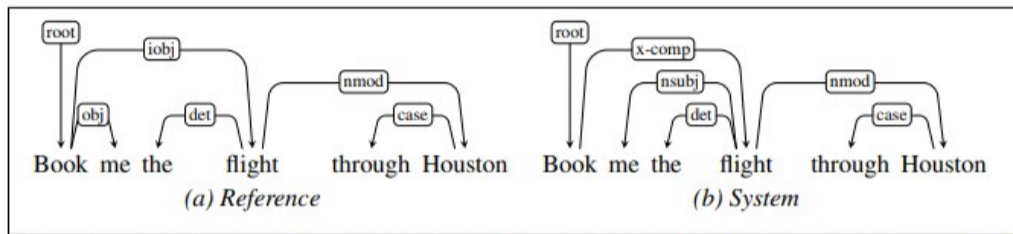


Figure 14.16 Reference and system parses for *Book me the flight through Houston*, resulting in an LAS of 2/3 and an UAS of 5/6.

- ▶ http://nlpprogress.com/english/dependency_parsing.html
- ▶ UAS Se reporta 97 %
- ▶ LAS se reporta 95 %