

Análisis de Textos

Grupo PLN – InCo
2024

Introducción a la gramática del español

Contenido

- Conceptos básicos
- Categorías gramaticales
 - léxicas
 - sintagmáticas
- Funciones sintácticas
- Funciones semánticas
- Herramientas de análisis lingüístico

Gramática

En un sentido restringido, la gramática estudia las unidades significativas del lenguaje y su combinatoria.

(Definiciones más amplias incluyen otros aspectos del lenguaje, como el estudio del significado.)

Gramática

La gramática comprende dos partes:

- MORFOLOGÍA

Estudia la estructura interna de las palabras.

- SINTAXIS

Estudia la estructura de la oración, es decir, la combinatoria de las palabras.

Gramática

MORFOLOGÍA

unidad mínima: morfema (*pre-histór-ico-s*)

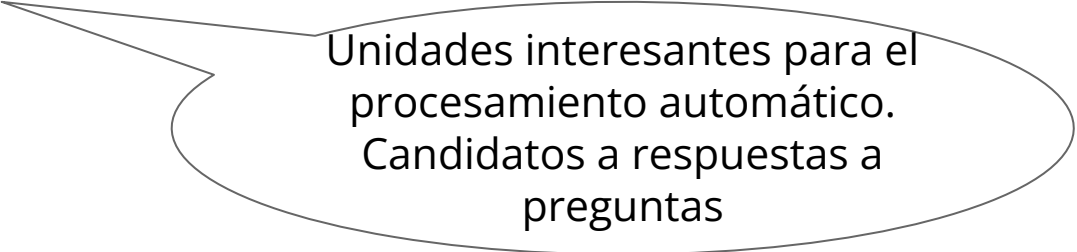
unidad máxima: palabra (*prehistóricos*)

SINTAXIS:

unidad mínima: palabra (*estudiantes*)

unidad máxima: oración (*Los estudiantes de computación trabajan*)

unidad intermedia: sintagma (*Los estudiantes de computación*)



Unidades interesantes para el
procesamiento automático.
Candidatos a respuestas a
preguntas

Oración

Definición de oración

Existen diferentes enfoques.

Tomamos la siguiente definición:

La oración es una unidad de predicación en donde se establece una relación entre dos constituyentes, el sujeto y el predicado.

(Di Tullio & Malcuori, 2013: 21.3 La oración y el enunciado)

Oración

Definición de oración

La oración es una unidad de predicación en donde se establece una relación entre dos constituyentes, el sujeto y el predicado.

*El árbol del jardín perdió sus hojas.
Llegó el tren.*

Oración

Definición de oración

La oración es una unidad de predicación en donde se establece una relación entre dos constituyentes, el sujeto y el predicado.

El árbol del jardín perdió sus hojas.

S

P

Llegó el tren.

P

S

Question & Answering: Detectar que el sujeto no está en la posición prototípica para responder correctamente.
¿Quién llegó?

Oración

Definición de oración

El árbol del jardín perdió sus hojas.

S

P

Llegó el tren.

P

S

El sujeto suele ser un sintagma nominal y el predicado un sintagma verbal (pero no siempre es así).

Oración / Enunciado

En un texto, la unidad es el **enunciado**, que va desde la mayúscula hasta el punto.

Algunos enunciados, los que tienen un verbo conjugado como núcleo, son también oraciones.

La **oración** es un objeto teórico, gramatical.

El **enunciado** es un objeto concreto, existe dentro de un texto.

Oración / Enunciado

Ejercicio

Marque los enunciados y las oraciones en el siguiente texto.
Indique sujeto y predicado de cada oración.

*Los artistas se ofrecieron a colaborar desde el principio.
Muchos, sólo para hacer pinta. Otros, sinceramente, se
quedaron a trabajar siempre que se necesitó.*

Oración / Enunciado

[*Los artistas se ofrecieron a colaborar desde el principio.*]
[*Muchos, sólo para hacer pinta.*] [*Otros, sinceramente, se quedaron a trabajar siempre que se necesitó.*]

Notación:

- Los corchetes segmentan enunciados.
- Los enunciados subrayados son oraciones.

Oración / Enunciado

[*Los artistas* se **ofrecieron** a colaborar desde el principio.]
[Muchos, sólo *para hacer pinta.*] [*Otros*, sinceramente, se **quedaron a trabajar** siempre que *se necesitó.*]

Notación:

- Los corchetes segmentan enunciados.
- Los enunciados subrayados son oraciones (independientes).
- Oraciones de infinitivo en azul
- Oraciones subordinadas en violeta

Categorías gramaticales

- Categorías léxicas (o clases de palabras)
- Categorías sintagmáticas

Categorías gramaticales

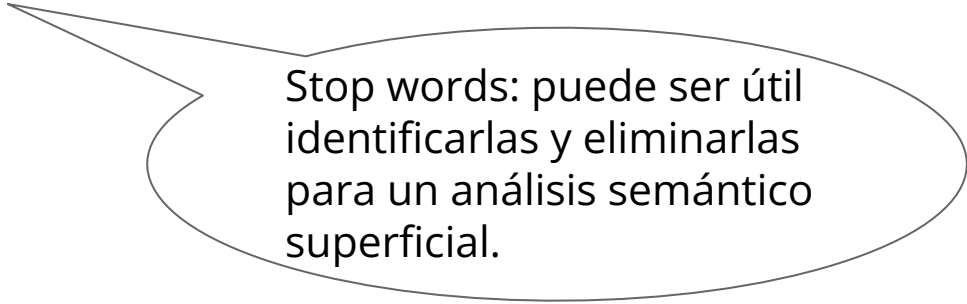
Categorías léxicas

- nombre/sustantivo (*casa, casas, felicidad*)
- verbo (*cantar, cantando, cantáramos, sabe, dé*)
- adjetivo (*alto, alta, lindos, inteligente, solo, media*)
- adverbio (*medio, simplemente, solo, ayer*)
- determinante (incluye artículo) (*el, una, unos, esos, nuestro, algún*)
- preposición (*a, de, por, contra, ...*)
- pronombre (*yo, él, mí, se, aquello, que, ese ...*)
- conjunción (*y, o, si, que, ...*)

Categorías gramaticales

Categorías léxicas

- Categorías de significado léxico (clases abiertas):
nombres, verbos, adjetivos, algunos adverbios
- Categorías de significado funcional o gramatical (clases cerradas):
determinantes, pronombres, preposiciones,
conjunciones, algunos adverbios.



Stop words: puede ser útil identificarlas y eliminarlas para un análisis semántico superficial.

Categorías gramaticales

Categorías léxicas

Ejercicio

¿Cuál es la categoría de las palabras resaltadas?

*De **su** padre **aprendió desde** muy niño el dominio de las armas de fuego, el amor por los caballos y la **maestranza** de las aves de presas **altas**, **pero** de **él** aprendió **también** las buenas artes del valor **y** la prudencia.*

“Crónica de una muerte anunciada”

Gabriel García Márquez

Categorías gramaticales

Categorías léxicas

<i>su</i>	→ determinante (posesivo)
<i>aprendió</i>	→ verbo
<i>desde</i>	→ preposición
<i>maestranza</i>	→ nombre (sustantivo)
<i>altas</i>	→ adjetivo
<i>pero</i>	→ conjunción
<i>él</i>	→ pronombre
<i>también</i>	→ adverbio

Categorías gramaticales

Categorías léxicas

¿A qué categoría pertenecen las palabras:

sobre

joven

como

?

Categorías gramaticales

Categorías léxicas

Ambigüedad: algunas palabras pertenecen a más de una categoría.

Categorías gramaticales

Categorías léxicas

medio: adjetivo

Quiero *medio* tomate.

Quiero *media* naranja.

medio: adverbio

Están *medio* cansadas

*Están *medias* cansadas

Están *un poco* cansadas

Categorías gramaticales

Categorías léxicas

*A Ana le gusta jugar con **la** pelota. **La** lleva al parque cada vez que va. Su hermano **la** acompaña y juega con ella.*

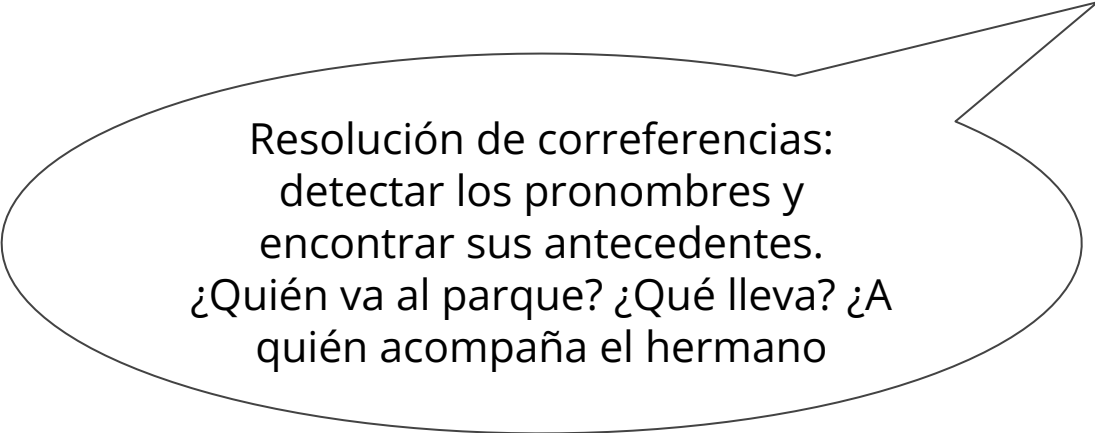
Ejercicio: ¿Cuál es la categoría de cada ocurrencia de **la**?

Categorías gramaticales

Categorías léxicas

*A Ana le gusta jugar con **la** [det] pelota. **La** [pron] lleva al parque cada vez que va. Su hermano **la** [pron] acompaña y juega con ella.*

¿Cuál es la categoría de cada ocurrencia de **la**?



Resolución de correferencias:
detectar los pronombres y
encontrar sus antecedentes.
¿Quién va al parque? ¿Qué lleva? ¿A
quién acompaña el hermano

Freeling (<https://nlp.lsi.upc.edu/freeling/demo/demo.php>)

▼ Coreferences

A Ana le gusta jugar con la pelota .

La lleva a el parque cada vez que va .

Su hermano la acompaña y juega con ella .



You

Dado el siguiente texto:

"A Ana le gusta jugar con la pelota. La lleva al parque cada vez que va. Su hermano la acompaña y juega con ella."

Responder:

1. ¿A quién le gusta jugar con la pelota?
2. ¿Quién va al parque?
3. ¿Qué lleva?
4. ¿A quién acompaña el hermano?



ChatGPT

1. A Ana le gusta jugar con la pelota.
2. Ana va al parque.
3. Ana lleva la pelota.
4. El hermano acompaña a Ana.

Categorías gramaticales

Categorías léxicas

Palabras desconocidas:

Ejercicio: ¿Cuál es la categoría léxica de las palabras resaltadas?

*Apenas se entreplumaban, algo como un ulucordio los **encrestoriaba**, los extrayuxtaba y paramovía, de pronto era el clinón, la **esterfurosa** convulcante de las mátricas, la jadehollante embocapluvia del orgumio, los **esproemios** del merpasmo en una **sobrehumítica** agopausa.*

Julio Cortázar

Rayuela

*Las mediciones de **transaminasas** o **aminotransferasas** son exámenes que indican inflamación hepática. Este examen está alterado casi por definición en todas las hepatitis.*

Categorías gramaticales

Categorías léxicas

encrestoriaba → verbo

esterfurosa → ¿adjetivo? ¿nombre?

esproemios → nombre

sobrehumítica → adjetivo

transaminasa, aminotransferasa → nombre

evidencia:

- morfología
- contexto

Categorías gramaticales

Categorías léxicas

Palabras compuestas

locución prepositiva

*Lo invitaron **a partir de** que demostró estar interesado en el tema.*

preposición + verbo + preposición

*Lo obligaron **a partir de** aquel lugar inmediatamente.*

Categorías gramaticales

Categorías léxicas

Expresiones multi-palabra

ojo de buey, lobo de mar

Categorías gramaticales

Categorías léxicas

- La ambigüedad en la asignación de la categoría gramatical,
- la determinación de la categoría de palabras desconocidas y
- la identificación de palabras compuestas

son algunos de los problemas que enfrenta la etapa de análisis léxico en el procesamiento automático del lenguaje natural.

Categorías gramaticales

Categorías sintagmáticas

- Los sintagmas (grupos / frases) son grupos de palabras en donde podemos identificar un núcleo.
- Tienen estructura interna.
- Cumplen diversas funciones dentro de la oración.

Categorías gramaticales

Categorías sintagmáticas

el estudiante de ingeniería de Montevideo

muy propenso a las infecciones respiratorias en primavera

casi compra otro libro por equivocación

bastante cerca de la escuela

más hacia la derecha

Categorías gramaticales

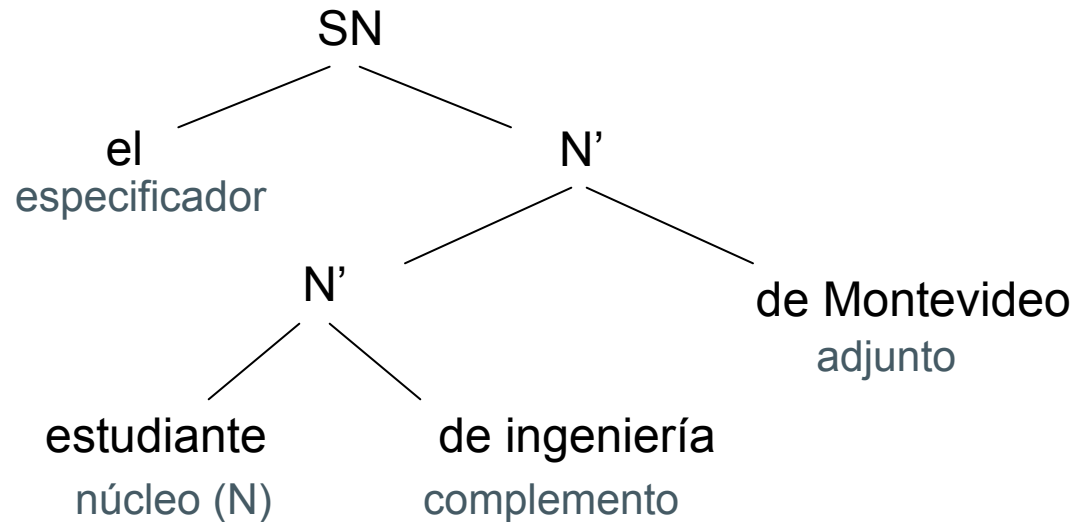
Categorías sintagmáticas

<i>el estudiante de ingeniería de Montevideo</i>	SN
<i>muy propenso a las infecciones respiratorias en primavera</i>	SAdj
<i>casi compra otro libro por equivocación</i>	SV
<i>bastante cerca de la escuela</i>	SAdv
<i>más hacia la derecha</i>	SP

Categorías gramaticales

Categorías sintagmáticas

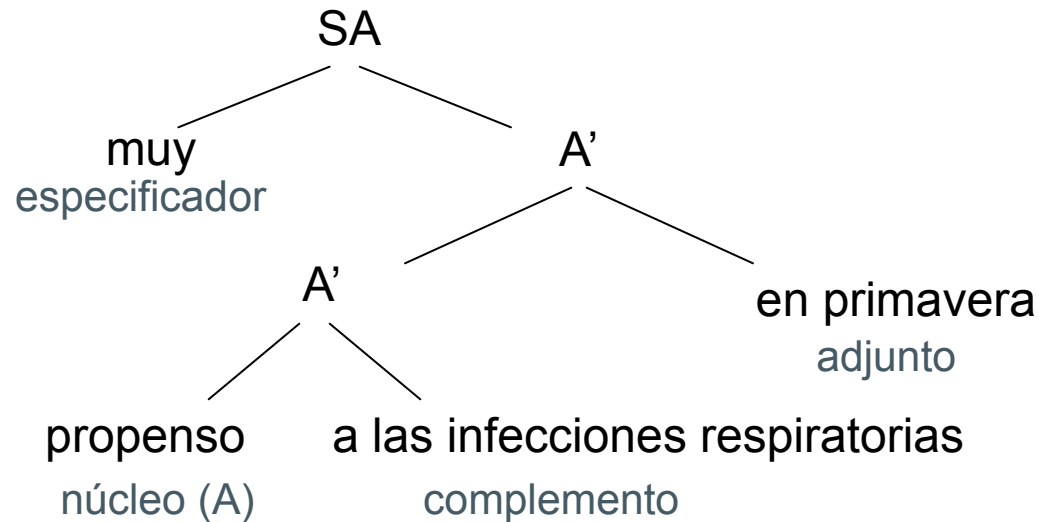
el estudiante de ingeniería de Montevideo



Categorías gramaticales

Categorías sintagmáticas

muy propenso a las infecciones respiratorias en primavera



Constituyentes

Constituyentes: sintagmas principales de la oración

- No existen procedimientos mecánicos para segmentar los constituyentes.
- Pueden realizarse algunas pruebas para identificarlos.

Identificación de constituyentes

Sustituir los constituyentes complejos por otros más simples:

Los estudiantes de Letras encontraron el libro de Sintaxis en la Biblioteca con mucho esfuerzo.

Ellos lo encontraron allí así

Identificación de constituyentes

Se puede mover (anteponer o posponer) solo un constituyente entero.

Los estudiantes de Letras encontraron el libro de Sintaxis en la Biblioteca con mucho esfuerzo.

El libro de Sintaxis encontraron los estudiantes de Letras con mucho esfuerzo en la Biblioteca.

Identificación de constituyentes

Se interroga por un constituyente entero.

¿*Quiénes* encontraron el libro de Sintaxis en la Biblioteca?
Los estudiantes de Letras.

¿*Qué* encontraron en la Biblioteca los estudiantes de
Letras? *El libro de Sintaxis.*

¿*Dónde* encontraron el libro de Sintaxis los estudiantes de
Letras? *En la Biblioteca*

Funciones sintácticas

Las funciones sintácticas expresan la relación que el sintagma mantiene con los otros componentes de la oración:

- Sujeto: Un árbol frondoso nos protege del sol.
- Objeto directo: Vimos un árbol frondoso.
- Objeto Indirecto: Le cortaron una rama a un árbol frondoso.
- Complemento de Régimen: Soñé con un árbol frondoso.
- Adjunto: Nos sentamos bajo un árbol frondoso.

Funciones semánticas

Las funciones semánticas, o roles semánticos, establecen una relación entre sintaxis y semántica

El niño come la manzana.

La manzana es comida por el niño.

Roles semánticos

Establecen una relación entre sintaxis y semántica

El niño come la manzana.

sujeto <-> agente (el que come)

objeto directo <-> tema (paciente) (lo comido)

La manzana es comida por el niño.

sujeto <-> tema (paciente)

complemento agente <-> agente

Para un sistema de Q&A en general nos interesa más el rol semántico.
¿Quién come una manzana?
¿Qué come el niño?

Roles semánticos

El ruido desagrada *al maestro*.

sujeto <-> tema

objeto indirecto <-> experimentante

El maestro detesta *el ruido*.

sujeto <-> experimentante

objeto directo <-> tema

Roles semánticos

Juan corta el pan con el cuchillo.

sujeto <-> agente

objeto directo <-> tema

complemento preposicional <-> instrumento

Roles semánticos

Agente (Arg0): *Juan corre.*

Paciente (Arg1): *Juan come manzanas. La mujer fue asesinada.*

Tema (Arg1): *Juan lee un libro.*

Experimentante (Arg0): *Juan le teme a la tormenta.*

Locativo (LOC): *Juan vive en París. Conocí a Juan en París.*

Temporal (TEMP): *Esta ópera dura tres horas.*

Instrumento (Arg2): *Esta llave abre la puerta. Abro con la llave.*

etc.

Herramientas para análisis lingüístico

- POS-taggers (POS: Part Of Speech):
 - Tokenizan el texto y realizan un análisis morfosintáctico de cada token (lema y pos-tag).
- Parsers:
 - realizan un análisis sintáctico de cada oración, basados en algún formalismo gramatical:
 - GLC (constituyentes): muy utilizadas durante muchos años, GLC probabilísticas.
 - Dependencias: muy utilizadas actualmente.
 - HPSG (gramáticas de rasgos) y Catoriales: pocas herramientas.

Herramientas para análisis lingüístico

Para el español:

Freeling (<http://nlp.lsi.upc.edu/freeling/demo/demo.php>)

Stanza (<https://stanfordnlp.github.io/stanza/>)

Spacy (<https://spacy.io/api/dependencyparser>)

...

Bibliografía

- Di Tullio, Ángela, Malcuori Marisa *Gramática del español Para maestros y profesores del Uruguay*. Montevideo, Codicen. 2013.