

Inteligencia Artificial

Gramáticas Independientes del Contexto

Edgar Andrade, Ph.D.

Matemáticas Aplicadas y Ciencias de la computación

Última revisión: Enero de 2023



Contenido

Motivación

Gramáticas independientes del contexto

Definición formal

Características gramaticales

Una CFG para el español



Contenido

Motivación

Gramáticas independientes del contexto

Definición formal

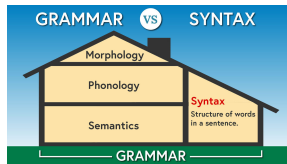
Características gramaticales

Una CFG para el español



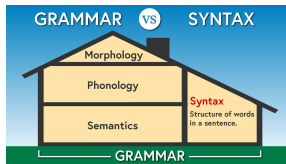
Sintaxis

Estudia la manera en la cual las palabras se pueden agrupar para formar frases, y las frases oraciones.



Sintaxis

Estudia la manera en la cual las palabras se pueden agrupar para formar frases, y las frases oraciones.



Existen principios implícitos que regularmente siguen los hablantes de una comunidad al comunicarse (verbalmente o por escrito).



Ejemplo 1: Oraciones “no gramaticales”

Oraciones que “se ven mal”:



Ejemplo 1: Oraciones “no gramaticales”

Oraciones que “se ven mal”:

- (1) Nosotros nos queremos.
- (2) *Nosotros queremos a nosotros.



Ejemplo 1: Oraciones “no gramaticales”

Oraciones que “se ven mal”:

- (1) Nosotros nos queremos.
- (2) *Nosotros queremos a nosotros.
- (3) Pedro tiene que trabajar.
- (4) *Pedro quiere que trabajar.



Ejemplo 2: Ambigüedades

Ambigüedades relacionadas con la estructura.



Ejemplo 2: Ambigüedades

Ambigüedades relacionadas con la estructura.

(5) Los profesores y los estudiantes antiguos serán reconocidos.



Ejemplo 2: Ambigüedades

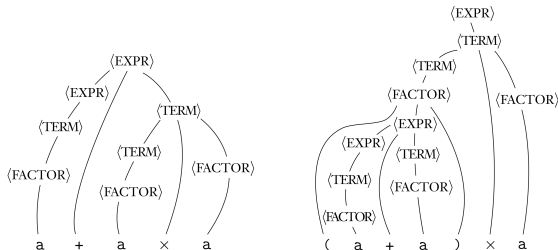
Ambigüedades relacionadas con la estructura.

(5) Los profesores y los estudiantes antiguos serán reconocidos.

(6) Todo hombre ama a una mujer.



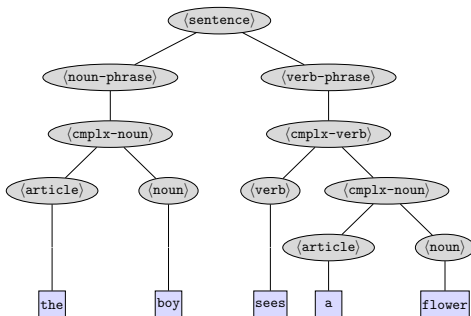
Analogía con el álgebra



- 👉 La representación lineal puede esconder la estructura.
- 👉 Las expresiones algebraicas pueden generarse mediante la aplicación de reglas.



Gramáticas para el lenguaje



Contenido

Motivación

Gramáticas independientes del contexto

Definición formal

Características gramaticales

Una CFG para el español



Ejemplo: gramática

La siguiente es una gramática independiente del contexto, G_1 :

$$A \rightarrow 0A1$$

$$A \rightarrow B$$

$$B \rightarrow 2$$



Ejemplo: gramática

La siguiente es una gramática independiente del contexto, G_1 :

$$A \rightarrow 0A1$$

$$A \rightarrow B$$

$$B \rightarrow 2$$

Una gramática es una colección de **reglas de sustitución** (o reglas de reescritura o producciones) de la forma

`variable` \rightarrow `cadena de variables y terminales`

Por ejemplo, en G_1 las variables son A y B ; los terminales son 0, 1, y 2.



Ejemplo: gramática

La siguiente es una gramática independiente del contexto, G_1 :

$$A \rightarrow 0A1$$

$$A \rightarrow B$$

$$B \rightarrow 2$$

Una gramática es una colección de **reglas de sustitución** (o reglas de reescritura o producciones) de la forma

`variable` \rightarrow `cadena de variables y terminales`

Por ejemplo, en G_1 las variables son A y B ; los terminales son 0, 1, y 2. La variable en la primera regla de sustitución es llamada **variable inicial**.



Ejemplo: derivación

Una gramática puede producir frases de la siguiente manera:

1. Escribimos la variable inicial;



Ejemplo: derivación

Una gramática puede producir frases de la siguiente manera:

1. Escribimos la variable inicial;
2. Elegimos una regla de sustitución para reemplazar la variable inicial con una cadena de variables y terminales;



Ejemplo: derivación

Una gramática puede producir frases de la siguiente manera:

1. Escribimos la variable inicial;
2. Elegimos una regla de sustitución para remplazar la variable inicial con una cadena de variables y terminales;
3. Seguimos con el paso 2 hasta que no queden variables.



Ejemplo: derivación

Una gramática puede producir frases de la siguiente manera:

1. Escribimos la variable inicial;
2. Elegimos una regla de sustitución para remplazar la variable inicial con una cadena de variables y terminales;
3. Seguimos con el paso 2 hasta que no queden variables.

La secuencia de sustituciones se llama una **derivación**.



Ejemplo: derivación

Una gramática puede producir frases de la siguiente manera:

1. Escribimos la variable inicial;
2. Elegimos una regla de sustitución para remplazar la variable inicial con una cadena de variables y terminales;
3. Seguimos con el paso 2 hasta que no queden variables.

La secuencia de sustituciones se llama una **derivación**.

Por ejemplo, $0002111 \in G_1$:

$A \Rightarrow 0A1$

$$\begin{cases} A \rightarrow 0A1 \\ A \rightarrow B \\ B \rightarrow 2 \end{cases}$$



Ejemplo: derivación

Una gramática puede producir frases de la siguiente manera:

1. Escribimos la variable inicial;
2. Elegimos una regla de sustitución para remplazar la variable inicial con una cadena de variables y terminales;
3. Seguimos con el paso 2 hasta que no queden variables.

La secuencia de sustituciones se llama una **derivación**.

Por ejemplo, $0002111 \in G_1$:

$A \Rightarrow 0A1 \Rightarrow 00A11$

$$\begin{cases} A \rightarrow 0A1 \\ A \rightarrow B \\ B \rightarrow 2 \end{cases}$$



Ejemplo: derivación

Una gramática puede producir frases de la siguiente manera:

1. Escribimos la variable inicial;
2. Elegimos una regla de sustitución para remplazar la variable inicial con una cadena de variables y terminales;
3. Seguimos con el paso 2 hasta que no queden variables.

La secuencia de sustituciones se llama una **derivación**.

Por ejemplo, $0002111 \in G_1$:

$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111$

$$\begin{cases} A \rightarrow 0A1 \\ A \rightarrow B \\ B \rightarrow 2 \end{cases}$$



Ejemplo: derivación

Una gramática puede producir frases de la siguiente manera:

1. Escribimos la variable inicial;
2. Elegimos una regla de sustitución para remplazar la variable inicial con una cadena de variables y terminales;
3. Seguimos con el paso 2 hasta que no queden variables.

La secuencia de sustituciones se llama una **derivación**.

Por ejemplo, $0002111 \in G_1$:

$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111 \Rightarrow 000B111 \quad \begin{cases} A \rightarrow 0A1 \\ A \rightarrow B \\ B \rightarrow 2 \end{cases}$$



Ejemplo: derivación

Una gramática puede producir frases de la siguiente manera:

1. Escribimos la variable inicial;
2. Elegimos una regla de sustitución para remplazar la variable inicial con una cadena de variables y terminales;
3. Seguimos con el paso 2 hasta que no queden variables.

La secuencia de sustituciones se llama una **derivación**.

Por ejemplo, $0002111 \in G_1$:

$$\begin{aligned} A &\Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111 \Rightarrow 000B111 \\ &\Rightarrow 0002111 \end{aligned} \quad \left\{ \begin{array}{l} A \rightarrow 0A1 \\ A \rightarrow B \\ B \rightarrow 2 \end{array} \right.$$

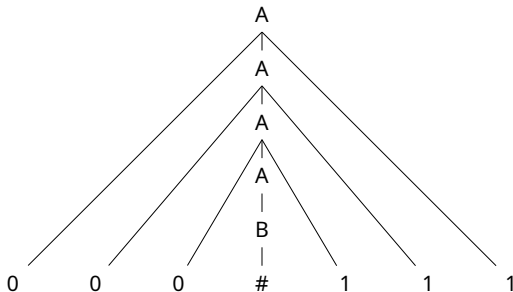


Ejemplo: árbol de parsing

Una manera muy útil de representar la derivación

$$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 000A111 \Rightarrow 000B111 \Rightarrow 0002111.$$

es utilizar un **árbol de parsing**:



Contenido

Motivación

Gramáticas independientes del contexto

Definición formal

Características gramaticales

Una CFG para el español



Definición

Estamos listos para definir formalmente una gramática independiente del contexto (o **CFG**).

Definición

Una **gramática independiente del contexto** es una 4-tupla (V, Σ, R, S) , donde:

- ▶ V es un conjunto finito de **variables**;
- ▶ Σ es un conjunto finito, disyunto de V , de **terminales**;
- ▶ R es el conjunto finito de **reglas**, cada regla asocia una variable a una cadena de variables y terminales;
- ▶ $S \in V$ es la variable inicial.



Definición

Si u y v son dos cadenas de terminales y A es una variable, con una regla $A \rightarrow w$, decimos que uwv se **deriva directamente** de uAv , escrito $uAv \Rightarrow uwv$.



Definición

Si u y v son dos cadenas de terminales y A es una variable, con una regla $A \rightarrow w$, decimos que uwv se **deriva directamente** de uAv , escrito $uAv \Rightarrow uwv$.

En general, decimos que v **deriva** u si existe una secuencia u_1, u_2, \dots, u_k , con $k \geq 0$ tal que

$$u = u_1 \Rightarrow u_2 \Rightarrow \dots \Rightarrow u_k = v,$$

lo que se denota como $u \xRightarrow{*} v$.



Definición

Si u y v son dos cadenas de terminales y A es una variable, con una regla $A \rightarrow w$, decimos que uwv se **deriva directamente** de uAv , escrito $uAv \Rightarrow uwv$.

En general, decimos que v **deriva** u si existe una secuencia u_1, u_2, \dots, u_k , con $k \geq 0$ tal que

$$u = u_1 \Rightarrow u_2 \Rightarrow \dots \Rightarrow u_k = v,$$

lo que se denota como $u \xRightarrow{*} v$.

El **lenguaje de una gramática** es $\{w \in \Sigma^* : S \xRightarrow{*} w\}$.



Lenguajes independientes del contexto

Todas las frases que se pueden generar a partir de una gramática independiente del contexto constituyen su lenguaje.



Lenguajes independientes del contexto

Todas las frases que se pueden generar a partir de una gramática independiente del contexto constituyen su lenguaje.

Por ejemplo $L(G_1) = \{0^n 21^n : n \geq 0\}$.



Lenguajes independientes del contexto

Todas las frases que se pueden generar a partir de una gramática independiente del contexto constituyen su lenguaje.

Por ejemplo $L(G_1) = \{0^n 21^n : n \geq 0\}$.

Todos los lenguajes que pueden ser descritos por una gramática independiente del contexto se llaman **lenguajes independientes del contexto** (o CFL).



Contenido

Motivación

Gramáticas independientes del contexto

Definición formal

Características gramaticales

Una CFG para el español



Motivación

Supongamos que tenemos la siguiente “toy grammar” para el español, en la cual las variables son O para oraciones, SN para sintagmas nominales, D para determinantes, N para sustantivos y V para verbos:

$$\left\{ \begin{array}{ll} O & \rightarrow SN_V \\ SN & \rightarrow D_N \\ D & \rightarrow \text{un} \mid \text{unos} \\ N & \rightarrow \text{hombre} \mid \text{hombres} \\ V & \rightarrow \text{camina} \mid \text{caminan} \end{array} \right.$$



Motivación

Supongamos que tenemos la siguiente “toy grammar” para el español, en la cual las variables son O para oraciones, SN para sintagmas nominales, D para determinantes, N para sustantivos y V para verbos:

- | | | |
|------|---|--------------------------|
| O | $\rightarrow SN_V$ | (7) Un hombre camina. |
| SN | $\rightarrow D_N$ | (8) *Unos hombre camina. |
| D | $\rightarrow \text{un} \mid \text{unos}$ | (9) *Un hombres camina. |
| N | $\rightarrow \text{hombre} \mid \text{hombres}$ | (10) *Un hombre caminan. |
| V | $\rightarrow \text{camina} \mid \text{caminan}$ | |



Características gramaticales

Ejemplos: género (f=femenino; m=masculino), número (sn=singular; pl=plural) y persona (1a, 2a, 3a).

Género y número:

- ▶ Luna (f, sn)
- ▶ Carros (m, pl)
- ▶ Perro (m, sn)
- ▶ Aguas (f, pl)

Persona y número:

- ▶ Duermo (1a, sn)
- ▶ Corren (2a, pl)
- ▶ Camina (3a, sn)
- ▶ Leemos (1a, pl)



Solución inicial

$$\left\{ \begin{array}{ll} O & \rightarrow SN_{sn} V_{sn} \\ O & \rightarrow SN_{pl} V_{pl} \\ SN_{sn} & \rightarrow D_{sn} N_{sn} \\ SN_{pl} & \rightarrow D_{pl} N_{pl} \\ D_{sn} & \rightarrow \text{un} \\ D_{pl} & \rightarrow \text{unos} \\ N_{sn} & \rightarrow \text{hombre} \\ N_{pl} & \rightarrow \text{hombres} \\ V_{sn} & \rightarrow \text{camina} \\ V_{pl} & \rightarrow \text{caminan} \end{array} \right.$$



Solución inicial

$$\left\{ \begin{array}{ll} O & \rightarrow SN_{sn} \sqcup V_{sn} \\ O & \rightarrow SN_{pl} \sqcup V_{pl} \\ SN_{sn} & \rightarrow D_{sn} \sqcup N_{sn} \\ SN_{pl} & \rightarrow D_{pl} \sqcup N_{pl} \\ D_{sn} & \rightarrow \text{un} \\ D_{pl} & \rightarrow \text{unos} \\ N_{sn} & \rightarrow \text{hombre} \\ N_{pl} & \rightarrow \text{hombres} \\ V_{sn} & \rightarrow \text{camina} \\ V_{pl} & \rightarrow \text{caminan} \end{array} \right.$$

👉 **Problema:**

Hay un crecimiento exponencial en el número de reglas respecto al número de características.



Uso de características gramaticales

Las características gramaticales se pueden usar para impedir cierto tipo de combinaciones.

Por ejemplo, consideremos la regla:

$$SN_{[\text{num}: ?n]} \rightarrow D_{[\text{num}: ?n]} N_{[\text{num}: ?n]}$$

Esta regla determina que SN , D y N deben tener el mismo número.



Uso de características gramaticales

Las características gramaticales se pueden usar para impedir cierto tipo de combinaciones.

Por ejemplo, consideremos la regla:

$$SN_{[\text{num}: ?n]} \rightarrow D_{[\text{num}: ?n]} N_{[\text{num}: ?n]}$$

Esta regla determina que SN , D y N deben tener el mismo número.

Si no se menciona una característica, la regla permite cualquier valor.



Usando la característica del número

(11) Un hombre camina.

(12) *Un hombre caminan.

(13) Unos hombres caminan.

(14) *Unos hombre camina.

$$\begin{cases}
 O & \rightarrow SN_{[\text{num}: ?n]} \cup V_{[\text{num}: ?n]} \\
 SN_{[\text{num}: ?n]} & \rightarrow D_{[\text{num}: ?n]} \cup N_{[\text{num}: ?n]} \\
 D_{[\text{num}: \text{sg}]} & \rightarrow \text{un} \\
 D_{[\text{num}: \text{pl}]} & \rightarrow \text{unos} \\
 N_{[\text{num}: \text{sg}]} & \rightarrow \text{hombre} \\
 N_{[\text{num}: \text{pl}]} & \rightarrow \text{hombres} \\
 V_{[\text{num}: \text{sg}]} & \rightarrow \text{camina} \\
 V_{[\text{num}: \text{pl}]} & \rightarrow \text{caminan}
 \end{cases}$$



Contenido

Motivación

Gramáticas independientes del contexto

Definición formal

Características gramaticales

Una CFG para el español



Léxico

Categoría en Español	Símbolo No Terminal	Símbolos Terminales
Sustantivo	N	hombre, mujer, libro, calle, parque
Nombre propio	T	Pedro, María, Juan
Determinante	D	un, una, unos, unas, el, la, los, las
Verbo Intransitivo	VI	camina, bebe
Verbo Transitivo	VT	ama, invita, lee
Preposiciones	P	en, por, con



Componentes (3/3)

Usaremos las siguientes características semánticas:

Característica	Abreviación	Valores
Género	gen	f=femenino, m=masculino
Número	num	sg=singular, pl=plural



Reglas estructurales LITE (1/2)

Regla 1. $O \rightarrow SN \quad V$

Regla 2. $SN \rightarrow T$

Regla 3. $SN \rightarrow D \quad N$

Regla 4. $V \rightarrow VI$



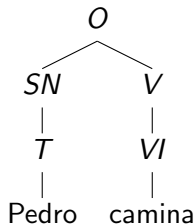
Reglas estructurales LITE (1/2)

Regla 1. $O \rightarrow SN \ V$

Regla 2. $SN \rightarrow T$

Regla 3. $SN \rightarrow D \ N$

Regla 4. $V \rightarrow VI$



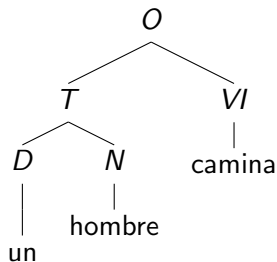
Reglas estructurales LITE (1/2)

Regla 1. $O \rightarrow SN \ V$

Regla 2. $SN \rightarrow T$

Regla 3. $SN \rightarrow D \ N$

Regla 4. $V \rightarrow VI$



Reglas estructurales LITE (2/2)

Regla 5. $V \rightarrow VT \text{ a } SN$

Regla 6. $V \rightarrow V \text{ } SP$

Regla 7. $SP \rightarrow P \text{ } SN$

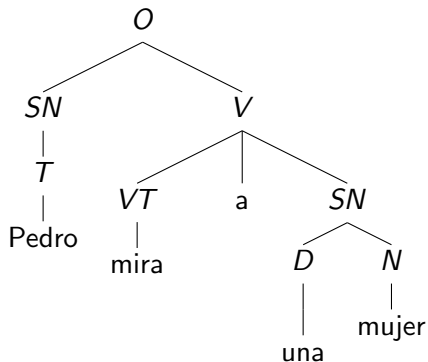


Reglas estructurales LITE (2/2)

Regla 5. $V \rightarrow VT \text{ a } SN$

Regla 6. $V \rightarrow V \text{ } SP$

Regla 7. $SP \rightarrow P \text{ } SN$

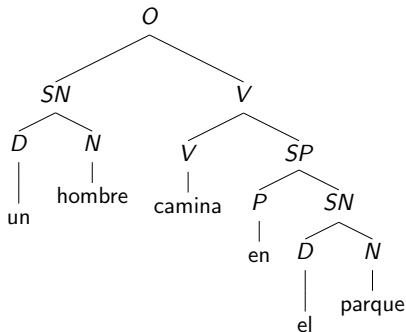


Reglas estructurales LITE (2/2)

Regla 5. $V \rightarrow VT \text{ a } SN$

Regla 6. $V \rightarrow V \text{ } SP$

Regla 7. $SP \rightarrow P \text{ } SN$



Ejemplos

(15) Un hombre ama a la mujer.

(16) Juan lee un libro en el parque.



Ejemplos

(15) Un hombre ama a la mujer.

(16) Juan lee un libro en el parque.

Encuentre dos ejemplos de oraciones que “suenan raro” pero que sean generadas por estas reglas.

(17)

(18)



En esta sesión usted aprendió

- ▶ Qué es un lenguaje independiente del contexto.
- ▶ Qué es un árbol de parsing.
- ▶ Qué es una característica gramatical y su uso.
- ▶ Construir gramáticas independientes del contexto.

