LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ALFABETOS**
  - Definición: conjunto finito y no vacío de símbolos.
  - Ejemplos: binario, vocales, letras, ajedrez, otros.
  - Pertenencia de símbolos; cardinalidad de alfabetos.
  - Relaciones: igualdad, inclusión e inclusión estricta.
  - Operaciones: unión, intersección, complementos.
  - Propiedades de las operaciones (son conjuntos !!!).
  - Concatenación o yuxtaposición de alfabetos y de símbolos de un alfabeto.

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **PALABRAS, cadenas, tiras o strings**
  - Definición: concatenación de símbolos de un alfabeto.
  - Longitud de una palabra.
  - Palabra vacía. Largo cero !!!
  - Concatenación de palabras sobre un alfabeto. Potenciación.
  - Propiedades: no conmutativa, asociativa, elemento neutro.
  - Subpalabra ($\omega=\alpha\beta\gamma$), sufijo y prefijo ($\omega=\alpha\beta$) propios e impropios.
  - Palabra inversa o refleja. Palíndromos.
  - Nuevas operaciones con alfabetos:
    - Concatenación: $\Sigma_1 . \Sigma_2 = \{\alpha=xy \,/\, x\in\Sigma_1 \wedge y\in\Sigma_2\}$
    - Potenciación: $\Sigma^n = \Sigma^{n-1} . \Sigma$ si $n > 0$, $\Sigma^n = \{\lambda\}$ si $n = 0$
    - Clausura positiva: $\Sigma^+ = \bigcup_{i=1}^{\infty} \Sigma^i$
    - Clausura o Cierre: $\Sigma^* = \bigcup_{i=0}^{\infty} \Sigma^i = \Sigma^+ \cup \{\lambda\}$
    - Estas operaciones generan conjuntos de “palabras”.
  - Universo de Discurso de un alfabeto: $W(\Sigma) = \Sigma^*$

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **LENGUAJES**
  - Lenguaje sobre un alfabeto Σ: $L \subseteq \Sigma^*$
  - Lenguaje vacío y lenguaje con cadena vacía. Cardinalidad.
  - Igualdad, inclusión e inclusión estricta.
  - Operaciones con lenguajes:
    - Unión / Intersección / Complementos / Concatenación
    - Potenciación, clausura positiva, clausura o cierre.
    - Inversión o reflexión.
  - Nueva operación con cadenas:
    - Regla de reescritura o producción: $\alpha := \beta$
    - Derivación directa por aplicación de producción: $\alpha \rightarrow \beta$
    - Derivación (en cero o más pasos): $\alpha \rightarrow^* \beta$
    - Derivación por derecha, por izquierda y mixta.
    - Reducción (en cero o más pasos): $\alpha \; {}^*\!\leftarrow \beta$

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **LENGUAJES**
  - Dado un alfabeto Σ, un lenguaje es un conjunto de palabras definidas sobre él: $\mathbf{L} \subseteq \Sigma^*$
  - Descripción de lenguajes (son conjuntos !!!):
    - Por enumeración o extensión
    - Por comprensión
      - Conjunto con una propiedad (propiedad)
      - Conjunto con fórmula-patrón (algebraicamente)
      - **GRAMÁTICA FORMAL** (estableciendo cómo se derivan sus elementos desde un símbolo inicial)
  - **Recordar: LENGUAJES NATURALES vs LENGUAJES FORMALES**

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **GRAMÁTICA FORMAL**
  - Definición de la gramática **G**: $(\Sigma_T, \Sigma_N, S, P)$; $\Sigma_T \cap \Sigma_N = \varnothing$; $S \in \Sigma_N$
  - Lenguaje generado por **G**: $L(G) = \{ \alpha \in \Sigma_T^* \;/\; S \to^* \alpha \}$
  - Noam Chomsky, Backus, Naur y **BNF** (ver formato).
  - Forma sentencial: $S \to^* \alpha$; $\alpha \in (\Sigma_T \cup \Sigma_N)^*$
  - Sentencia: $S \to^* \alpha$; $\alpha \in \Sigma_T^*$
  - Equivalencia: $G1 \equiv G2 \leftrightarrow L(G1) = L(G2)$
  - Regla no compresora: $\alpha := \beta$; $|\alpha| \leq |\beta|$
  - Regla compresora: $\alpha := \beta$; $|\alpha| > |\beta|$
  - Regla lambda: $S := \lambda$; **S = axioma**
  - Regla innecesaria: **A := A ; A = un símbolo no terminal**
  - Regla no generativa: $A := \lambda$; $A \neq$ **axioma**
  - Regla de redenominación: **A := B ; A y B símbolos no terminales**

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **Tipos de Lenguajes**: **Jerarquía de Chomsky (1956)**
  - **Tipo 0**: Estructurados por frases, sin restricciones o recursivamente enumerables.

    $\boldsymbol{\alpha A \beta := \gamma}$ $\quad \alpha, \beta, \gamma \in (\Sigma_T \cup \Sigma_N)^* \wedge A \in \Sigma_N$
  - **Tipo 1**: Dependiente del contexto o sensibles al contexto. (admiten reglas contextuales)

    $\boldsymbol{\alpha A \beta := \alpha\gamma\beta}$ $\quad \alpha, \beta \in (\Sigma_T \cup \Sigma_N)^* \wedge \gamma \in (\Sigma_T \cup \Sigma_N)^+ \wedge A \in \Sigma_N$

    $\mathbf{S := \lambda}$ $\quad$ (sin reglas compresoras salvo para el axioma)
  - **Tipo 2**: Independiente del contexto o de contexto libre.

    $\mathbf{A := \alpha}$ $\quad A \in \Sigma_N \wedge \alpha \in (\Sigma_T \cup \Sigma_N)^+$

    $\mathbf{S := \lambda}$ $\quad$ S = axioma
  - **Tipo 3**: Regulares o Lineales.

    A := a|aB ó $\quad$ Lineales por derecha

    A := a|Ba ó $\quad$ Lineales por izquierda

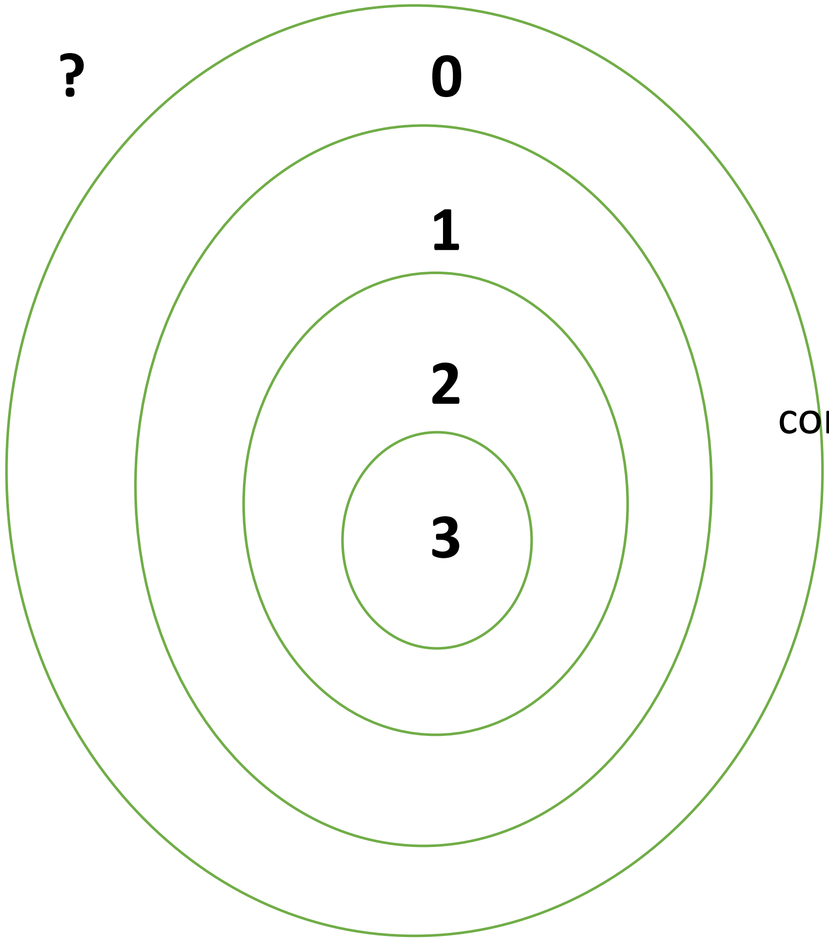
    $\mathbf{S \to \lambda}$ $\quad A, B, S \in \Sigma_N \wedge a \in \Sigma_T \wedge S$ = axioma

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **TIPOS DE LENGUAJES: Jerarquía de Chomsky (1956)**



Estructurados por Frases

Sin reglas compresoras

Dependientes del Contexto

Independientes del Contexto

Regulares o Lineales

$$L3 \subset L2 \subset L1 \subset L0$$

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **LENGUAJES REGULARES**
  - Todo lenguaje finito es regular.
  - Si **L1** y **L2** son lenguajes regulares, también lo son su **unión, concatenación y clausura transitiva y reflexiva**.
  - Solo son regulares los lenguajes construidos con lo anterior.
- **EXPRESIONES REGULARES (ER)**
  - $\varnothing$ es una ER que denota al lenguaje **L($\varnothing$) = {}**
  - λ es una ER que denota al lenguaje **L(λ) = {λ}**
  - $\forall$**a**$\in\Sigma$**, a** es una ER que denota al lenguaje **L(a) = {a}**

  Si **E1** y **E2** son expresiones regulares que denotan a **L1** y **L2**, entonces:
  - **E1+E2** es una ER que denota al lenguaje **L(E1+E2) = L1 $\cup$ L2**
  - **E1.E2** es una ER que denota al lenguaje **L(E1 . E2) = L1 . L2**
  - **E1*** es una ER que denota al lenguaje **L(E1*) = L*(E1) = L1***
  - **(E1)** es una ER que denota al lenguaje **L((E1)) = L(E1) = L1**
  - Sólo son ER las construidas con las reglas anteriores.

LINGÜÍSTICA MATEMÁTICA
- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **LENGUAJES INDEPENDIENTES DEL CONTEXTO**

  - **Gramática Limpia**: una gramática independiente del contexto SIN:
    - Reglas innecesarias: $\mathbf{A := A}$, $\mathbf{A} \in \Sigma_N$
    - Símbolos inaccesibles: $\nexists\, \boldsymbol{S} \to^{*} \boldsymbol{\alpha X \beta}$, $\mathbf{X} \in (\Sigma_T \cup \Sigma_N)$
    - Símbolos superfluos: $\nexists\, \boldsymbol{A} \to^{*} \boldsymbol{\alpha}\ \boldsymbol{con}\ \boldsymbol{\alpha} \in \boldsymbol{\Sigma_T^{*}}$

  - **Gramática Bien Formada**: una gramática limpia SIN:
    - Reglas no generativas: $\mathbf{A := \lambda}$, $\mathbf{A \neq axioma}$
    - Reglas de redenominación: $\mathbf{A := B}$, $\mathbf{A,B \in \Sigma_N}$

En cada caso, hay que ver cómo encontrar en la gramática y cómo quitar de ella la característica no deseada, obteniendo una gramática equivalente

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO**
  - **¿$\alpha \in L(G)$?**, siendo $\alpha \in \Sigma_T^*$. Los procesos que usamos para responder esta pregunta se denominan Análisis Sintáctico de la cadena α dada una gramática **$G = (\Sigma_T, \Sigma_N, S, P)$**.
  - Hasta ahora, respondemos:
    - **SI**, si podemos encontrar **$S \to^* \alpha$** usando las reglas de **P**, y
    - **NO**, si demostramos que no existe **$S \to^* \alpha$**
  - **Árbol de Derivación o de Análisis Sintáctico**: representación pictórica de la derivación **$S \to^* \alpha$** de una cadena $\alpha \in \Sigma_T^*$.
    - El axioma **S** de la gramática se sitúa en la raíz del árbol.
    - Si para **$A \in \Sigma_N$** existe en **P** la producción **$A := a_1a_2 \ldots a_n$** y ésta es usada en la derivación, entonces se crean como hijos del nodo **A** del árbol, nodos para cada uno de los **a_i** en el orden anterior.
    - Así, los nodos internos son símbolos de Σ_N y las hojas de Σ_T.

LINGÜÍSTICA MATEMÁTICA
- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO** (continuación)
  - Ahora, también podemos responder a **¿$\alpha \in L(G)$?**:
    - **SI**, si podemos construir el árbol de **análisis sintáctico de** α, y
    - **NO**, si demostramos que tal árbol no existe.
  - **Ambigüedad**.
    - Decimos que una cadena $\alpha \in \Sigma_T^*$ **es ambigua** si y solo si existe más de un árbol de análisis sintáctico ella en **G**.
    - Decimos que la gramática **G es ambigua**, si genera al menos una cadena ambigua.
    - Un lenguaje **L(G)** se dice que es **inherentemente ambiguo**, si las únicas gramáticas que lo generan son ambiguas.
    - **La ambigüedad es una característica indecidible de las GIC**.
    - La ambigüedad genera problemas de **significado**.

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO** (continuación)
  - **Recursividad**.
    - Decimos que una producción es recursiva, si el mismo no terminal aparece en el lado derecho e izquierdo: **A := αAβ**.
    - Si la gramática **G** posee una producción recursiva, se dice que tiene **recursividad en un paso**.
    - Si no tiene producciones recursivas, pero puede efectuarse la derivación **A$\rightarrow$*αAβ**, se dice que **G** posee recursión en más de un paso.
    - **Importancia de la recursión para generar lenguajes infinitos.**
    - Recursión por izquierda: **A := Aβ** - **A$\rightarrow$*Aβ**
    - Recursión por derecha: **A := αA** - **A$\rightarrow$*αA**
    - Algunos algoritmos de análisis sintácticos para poder funcionar requieren que **no existe recursión por izquierda** en la gramática.

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO** (continuación)
  - **Recursividad** (continuación)
    - **Eliminación de recursión por izquierda en un paso**: Dada una gramática independiente del contexto $\mathbf{G = (\Sigma_T, \Sigma_N, S, P)}$ con producciones recursivas por izquierda para el no terminal **A**

$$\mathbf{A := A\alpha_1 \mid A\alpha_2 \mid \ldots \mid A\alpha_n \mid \beta_1 \mid \beta_2 \mid \ldots \mid \beta_m}$$

con $\alpha_i, \beta_j \in (\Sigma_T \cup \Sigma_N)^+$, entonces puede construirse una gramática equivalente a la dada sin recursión por izquierda en **A**, reemplazando esas producciones por:

$$\mathbf{A := \beta_1 X \mid \beta_2 X \mid \ldots \mid \beta_m X \mid \beta_1 \mid \beta_2 \mid \ldots \mid \beta_m}$$

$$\mathbf{X := \alpha_1 X \mid \alpha_2 X \mid \ldots \mid \alpha_n X \mid \alpha_1 \mid \alpha_2 \mid \ldots \mid \alpha_n}$$

donde **X** es un nuevo símbolo no terminal.

LINGÜÍSTICA MATEMÁTICA
- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO** (continuación)
  - **Recursividad** (continuación)
    - **Eliminación de recursión por izquierda en más de un paso**: Se sigue un algoritmo iterativo donde se aplica la eliminación en un paso reiteradamente (ver bibliografía).
  - **Factorización por izquierda**
    - Si en una **GIC** para un mismo no terminal **A**, hay producciones que inician con los mismos símbolos en el lado derecho:

$$\mathbf{A := \alpha\beta_1 \mid \alpha\beta_2}$$

entonces, se obtiene una gramática equivalente al reemplazar esas producciones por:

$$\mathbf{A := \alpha X \quad ; \quad X := \beta_1 \mid \beta_2}$$

donde **X** es un nuevo símbolo no terminal.

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO** (continuación)
  - **Formas Normales**

    Las **GIC** siempre pueden ser convertidas en gramáticas equivalentes, en las cuales los lados derechos de sus producciones tengan un formato uniforme. Estas gramáticas equivalentes reciben el nombre de **Formas Normales**.

    - Estas normalizaciones **son a veces convenientes** para desarrollar ciertos algoritmos de análisis sintáctico.
    - Recordemos que las **GIC** pueden tener como única producción compresora, la regla lambda: **S :=** λ. Todas las otras reglas tendrán la forma **A :=** α donde $\alpha \in (\Sigma_T \cup \Sigma_N)^+$, es una cadena de terminales y no terminales.

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO** (continuación)
  - **Primera Forma Normal** (no en libro)

    Dada una **GIC = (Σ_T, Σ_N, S, P) limpia y bien formada**, todas las producciones en **P** tendrán la forma **A := α** donde $\alpha \in (\Sigma_T \cup \Sigma_N)^+$, salvo la regla lambda **S := λ**. Para cada una de ellas:
    - Si α es un solo terminal, se deja la producción sin cambiar.
    - En caso contrario, sea α = $X_1X_2 \ldots X_n$, donde los $X_i \in (\Sigma_T \cup \Sigma_N)$ son terminales o no terminales; sin alterar el lenguaje generado por la gramática, podemos realizar la siguiente conversión:
      1) Si X_i es un símbolo no terminal, lo dejamos como está en α.
      2) Si X_i es un símbolo terminal, entonces:
         a) Lo reemplazamos por Y_i en α.
         b) Agregamos Y_i al alfabeto de símbolos no terminales.
         c) Agregamos la producción Y_i := X_i al conjunto **P**.
    - Al finalizar todas las producciones tendrán en su lado derecho o un **solo símbolo terminal**, o una **cadena de sólo no terminales**.

LINGÜÍSTICA MATEMÁTICA
- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO** (continuación)
  - **Forma Normal de Chomsky (FNC)**

    Dada una **GIC = (Σ_T, Σ_N, S, P)**, se dice que la misma está en **forma normal de Chomsky** si todas sus producciones tienen la forma:

    $$\mathbf{S := \lambda \quad o \quad A := BC \quad o \quad A := a \qquad con\ S,A,B,C \in \Sigma_N \ \wedge \ a \in \Sigma_T}$$

    Toda **GIC** puede ser convertida en **FNC** haciendo:

    1) Obtener una gramática equivalente **limpia y bien formada**.
    2) Convertir la anterior a la **Primera Forma Normal**.
       Ahora todas las producciones son de la forma **no terminal produce un terminal o una cadena de sólo no terminales.**
    3) Para las producciones **A := Bη** donde $\eta \in \Sigma_N^+$, generar un nuevo símbolo no terminal **X** y reemplazar la producción por **A := BX** y **X := η** hasta que todos los lados derechos queden de largo 2.

    - Al finalizar todas las producciones estarán en **FNC**.

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO** (continuación)
  - **Forma Normal de Chomsky (FNC)** (continuación)
    - Notar que en el libro no se utiliza la **Primera Forma Normal**, sino que se realiza una conversión directa desde la **GIC** a una gramática equivalente en **FNC**.
    - Si una **GIC** está en Forma Normal de Chomsky, entonces sus árboles de derivación serán siempre árboles binarios; tenemos excelentes algoritmos para el manejo de éstos árboles.
    - Algoritmos como los de CYK (Cooke, Young, Kasami) necesitan que la gramática esté en **FNC**.

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO** (continuación)
  - **Forma Normal de Greibach (FNG)**

    Dada una **GIC = (Σ_T, Σ_N, S, P)**, se dice que la misma está en **forma normal de Greibach** si todas sus producciones tienen la forma:

    $$S \to \lambda \;\; \text{o} \;\; A := a\eta \quad \text{donde} \quad A \in \Sigma_N \;\wedge\; a \in \Sigma_T \;\wedge\; \eta \in \Sigma_N^*$$

    Toda **GIC** puede ser convertida en **FNG** haciendo:

    1) Obtener una gramática equivalente **limpia y bien formada**.
    2) Quitar la **recursividad por izquierda** en uno o más pasos.
    3) Asignar un orden cualquiera a los símbolos no terminales, digamos **$A_1, A_2, \ldots, A_n$** (sólo para poder iterar ordenadamente)
    4) Separar las producciones en tres grupos:

       Grupo 1: $S \to \lambda$ **o** $A := a\alpha$; **donde** $A \in \Sigma_N \wedge a \in \Sigma_T \wedge \alpha \in (\Sigma_T \cup \Sigma_N)^*$

       Grupo 2: $A_i := A_j\alpha$; **donde** $A_i, A_j \in \Sigma_N \wedge \alpha \in (\Sigma_T \cup \Sigma_N)^+ \wedge i < j$ **ant**

       Grupo 3: $A_i := A_j\alpha$; **donde** $A_i, A_j \in \Sigma_N \wedge \alpha \in (\Sigma_T \cup \Sigma_N)^+ \wedge i > j$ **pos**

LINGÜÍSTICA MATEMÁTICA

- Alfabetos
- Palabras
- Lenguajes
- Gramática formal
- Tipos de lenguajes
- L. Regulares
  - Definición
  - Expresiones regulares
- L. Independientes del Contexto
  - Limpia
  - Bien Formada
- Análisis Sintáctico
  - ¿$\alpha \in L(\Sigma)$?
  - Árbol
  - Ambigüedad
  - Recursión
  - Factorización
  - Forma Normal

SINTAXIS Y SEMÁNTICA DE LOS LENGUAJES

- **ANÁLISIS SINTÁCTICO** (continuación)
  - **Forma Normal de Greibach (FNG)** (continuación)

5) Para cada producción $\mathbf{A_i := A_j\alpha}$ del **Grupo 3**, iniciando con las de índice **i** menor, reemplazarlas (eliminar y agregar) por las producciones $\mathbf{A_i := \delta_1\alpha \mid \delta_2\alpha \mid \ldots \delta_k\alpha}$ donde los δ_i son los lados derechos de todas las producciones de $\mathbf{A_j}$. Al terminar este proceso todas las producciones serán de los **Grupos 1 o 2**.
6) Proceder igual que en (5) con las producciones del **Grupo 2**. Al terminar este proceso todas las producciones serán del **Grupo 1**.
7) Para todas las reglas del **Grupo 1** de la forma $\mathbf{A := a\alpha}$ donde $\mathbf{A \in \Sigma_N \land a \in \Sigma_T \land \alpha \in (\Sigma_T \cup \Sigma_N)^+}$, reemplazar los terminales **x** en α por nuevos no terminales $\mathbf{Y_x}$ y agregar las producciones $\mathbf{Y_x := x}$. Al terminar este proceso, todas las producciones estarán en **FNG**.