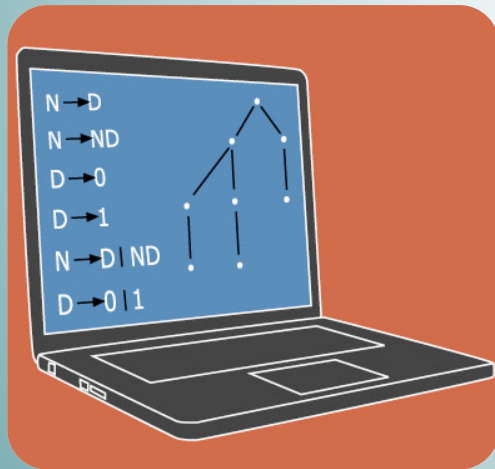


# Procesadores de Lenguaje

## Procesamiento de Lenguaje Natural. Algoritmo de Earley



**Cristina Tirnauca**

DPTO. DE MATEMÁTICAS,  
ESTADÍSTICA Y COMPUTACIÓN

Este tema se publica bajo Licencia:

[Creative Commons BY-NC-SA 3.0](https://creativecommons.org/licenses/by-nc-sa/3.0/)

# El algoritmo de Earley

- ▶ Es un algoritmo descendiente
- ▶ Funciona para cualquier gramática independiente de contexto.
- ▶ Complejidad:  $\mathcal{O}(n^3)$  para el caso general,  $\mathcal{O}(n^2)$  para gramáticas no ambiguas,  $\mathcal{O}(n)$  para casi todas las gramáticas LR(k).
- ▶ Funciona muy bien con la recursividad izquierda.
- ▶ El algoritmo original de Earley incluía un “look-ahead”.
- ▶ Hay que manejar con cuidado reglas tipo  $A \rightarrow \epsilon$

## Tipo de reglas

Se construye una secuencia de conjuntos llamados **Earley sets**

Input:  $x_1 x_2 \dots x_n$

Secuencia de conjuntos:  $S_0, S_1, \dots, S_n$

Un conjunto  $S_i$  contiene varios "tokens":  $[A \rightarrow \alpha \bullet \beta, j]$

- ▶ **Scan**:  $[A \rightarrow \dots \bullet a \dots, j] \in S_i$ ,  $a = x_{i+1}$ ,  
añadimos  $[A \rightarrow \dots a \bullet \dots, j]$  a  $S_{i+1}$
- ▶ **Predict**:  $[A \rightarrow \dots \bullet B \dots, j] \in S_i$   
añadimos  $[B \rightarrow \bullet \alpha, i]$  a  $S_i$  para todas las reglas  $B \rightarrow \alpha$
- ▶ **Complete**:  $[A \rightarrow \dots \bullet, j] \in S_i$   
añadimos  $[B \rightarrow \dots A \bullet \dots, k]$  a  $S_i$  para todos los tokens  
 $[B \rightarrow \dots \bullet A \dots, k]$  en  $S_j$

Aceptar  $x_1 x_2 \dots x_n$  si en  $S_n$  hay una regla de tipo  $[S \rightarrow \alpha \bullet, 0]$

# El algoritmo de Earley en un ejemplo

$G = (\{S, M, T\}, \{a, +, *\}, S, \{S \rightarrow S + M | M, M \rightarrow M * T | T, T \rightarrow a\})$

Input:  $a + a * a$

Scan

Predict

Complete

$S_0 : \bullet a + a * a$	$S_1 : a \bullet + a * a$	$S_2 : a + \bullet a * a$
$[S \rightarrow \bullet S + M, 0]$	$[T \rightarrow a \bullet, 0]$	$[S \rightarrow S + \bullet M, 0]$
$[S \rightarrow \bullet M, 0]$	$[M \rightarrow T \bullet, 0]$	$[M \rightarrow \bullet M * T, 2]$
$[M \rightarrow \bullet M * T, 0]$	$[S \rightarrow M \bullet, 0]$	$[M \rightarrow \bullet T, 2]$
$[M \rightarrow \bullet T, 0]$	$[M \rightarrow M \bullet * T, 0]$	$[T \rightarrow \bullet a, 2]$
$[T \rightarrow \bullet a, 0]$	$[S \rightarrow S \bullet + M, 0]$	

$S_3 : a + a \bullet * a$	$S_4 : a + a * \bullet a$	$S_5 : a + a * a \bullet$
$[T \rightarrow a \bullet, 2]$	$[M \rightarrow M * \bullet T, 2]$	$[T \rightarrow a \bullet, 4]$
$[M \rightarrow T \bullet, 2]$	$[T \rightarrow \bullet a, 4]$	$[M \rightarrow M * T \bullet, 2]$
$[S \rightarrow S + M \bullet, 0]$		$[S \rightarrow S + M \bullet, 0]$
$[M \rightarrow M \bullet * T, 2]$		$[M \rightarrow M \bullet * T, 2]$
$[S \rightarrow S \bullet + M, 0]$		$[S \rightarrow S \bullet + M, 0]$