

Teoría de Lenguajes

Guía 8 Resuelta

Sebastián Taboh

21 de junio de 2018



Este apunte contiene soluciones a ejercicios de la octava guía de ejercicios del 1er cuatrimestre de 2017 de la materia Teoría de Lenguajes, que está dedicada a Parsers Descendentes, así como también extractos de las clases prácticas dadas en ese cuatrimestre.

Este documento fue creado cuando cursé en el 2017, en caso de encontrar errores, por favor comunicarlo por mail a sebi_282@hotmail.com.

Índice

1. Parsers recursivos descendentes predictivos	3
1.1. Teoría	3
1.1.1. Algunos problemas y cómo resolverlos	3
1.2. Ejercicios	4
Ejercicio 1	4
Ejercicio 2	5
2. Parsers LL(1)	7
2.1. Ejercicios	7
Ejercicio 3	7
Ejercicio 4	8
3. Gramáticas con conflictos	8
3.1. Ejercicios	8
Ejercicio 5	8
Ejercicio 6	9
4. Gramáticas extendidas (ELL) y parsers recursivos iterativos	11
4.1. Teoría	11
4.1.1. Gramáticas extendidas	11
4.1.2. Generación de código y Transformación de gramáticas	11
4.2. Ejercicios	12
Ejercicio 10	12
Ejercicio 12	14
5. Limitaciones de los parsers descendentes	15
6. Material	16

1. Parsers recursivos descendentes predictivos

1.1. Teoría

Vamos a hacer un repaso sobre este tipo de parsers.

- Parsers descendentes (parten del símbolo distinguido)
 - Utilizan la información del próximo token (i.e., token corriente, o τc) y los SD's de cada regla
 - Con eso,
 - se crea un procedimiento por cada no terminal, siendo el procedimiento principal (**Main**) el correspondiente al símbolo distinguido
 - dentro de cada procedimiento, se invocan los procedimientos de los no terminales incluidos, se busca los terminales esperados (**match**), o no se hace nada (si es un λ),
 - si hay más de regla para un mismo no terminal, se agregan condicionales para ejecutar sólo la regla que contenga al token actual (τc) en sus SD's.
- Se agrega un **else** final con **error()**, para cubrir el caso en que el τc no está en los SD de ninguna de las reglas posibles.

Vamos a definir dos funciones: *Primeros* y *Siguientes*

$$\begin{aligned}
 \text{Primeros}(\alpha) &: (V_N \cup V_T)^* \rightarrow \mathcal{P}(V_T) \\
 \text{Primeros}(\alpha) &= \{t \in V_T \mid \alpha \xrightarrow{*} t\beta\} \\
 \text{Siguientes}(N) &: V_N \rightarrow \mathcal{P}(V_T) \\
 \text{Siguientes}(N) &= \{t \in V_T \mid S\$ \xrightarrow{*} \dots Nt\dots\}
 \end{aligned}$$

Con eso creamos la función de Símbolos Directrices (SD):

$$SD(A \rightarrow \beta) = \begin{cases} \text{Primeros}(\beta) & \text{si } \beta \text{ no anulable } (\beta \not\Rightarrow^* \lambda) \\ \text{Primeros}(\beta) \cup \text{Siguientes}(A) & \text{si } \beta \text{ anulable } (\beta \Rightarrow^* \lambda) \end{cases}$$

1.1.1. Algunos problemas y cómo resolverlos

- Ambigüedad por SDs no disjuntos: factorización a la izquierda (*left-factorization*)

Si tenemos producciones de la forma

$A \rightarrow \alpha\beta_1 \mid \alpha\beta_2 \mid \dots \mid \alpha\beta_n \mid \delta_1 \mid \delta_2 \mid \dots \mid \delta_k$ donde $\alpha \neq \lambda$, las reemplazamos por

$$\begin{aligned}
 A &\rightarrow \alpha A' \mid \delta_1 \mid \delta_2 \mid \dots \mid \delta_k \\
 A' &\rightarrow \beta_1 \mid \beta_2 \mid \dots \mid \beta_n
 \end{aligned}$$

- Recursión a izquierda

- eliminación de la recursión inmediata

Si tenemos producciones de la forma

$A \rightarrow A\alpha_1 \mid A\alpha_2 \mid \dots \mid A\alpha_m \mid \beta_1 \mid \beta_2 \mid \dots \mid \beta_k$ las reemplazamos por

$$\begin{aligned}
 A &\rightarrow \beta_1 A' \mid \beta_2 A' \mid \dots \mid \beta_k A' \\
 A' &\rightarrow \alpha_1 A' \mid \alpha_2 A' \mid \dots \mid \alpha_m A' \mid \lambda
 \end{aligned}$$

- eliminación de la recursión no inmediata

Tenemos el siguiente algoritmo.

- Numerar los no terminales A_1, \dots, A_n
- Para $i \leftarrow 1 \dots n$
 - Para $j \leftarrow 1 \dots i - 1$
 - Reemplazar $A_i \rightarrow A_j \gamma$ por $A_i \rightarrow \delta_1 \gamma \mid \delta_2 \gamma \mid \dots \mid \delta_k \gamma$ (donde $A_j \rightarrow \delta_1 \mid \dots \mid \delta_k \in P$)
 - Eliminar la recursión inmediata de A_i

1.2. Ejercicios

Ejercicio 1

- a) $P = \{A \rightarrow A\{A\}A \mid \lambda\}$, $V_T = \{ \{, \} \}$, $V_N = \{A\}$

Primero, notemos que no es posible hacer un parser recursivo predictivo porque hay recursión a izquierda en la producción $A \rightarrow A\{A\}A$. Vamos a hacer eliminación de la recursión inmediata (no hay no inmediata dado que sólo hay una variable, A).

$$A \rightarrow \lambda A' \qquad A' \rightarrow \{A\}AA' \mid \lambda$$

Ahora ya no hay conflictos. Vamos a proceder a calcular los SDs para las distintas reglas y luego a armar los procedimientos para cada variable.

Tomemos la regla $A \rightarrow A'$ (es igual a $A \rightarrow \lambda A'$, antes se escribió el λ para explicitar todo el procedimiento). Como A' es anulable ($A' \xRightarrow{*} \lambda$), obtenemos

$$SD(A \rightarrow A') = \text{Primeros}(A') \cup \text{Siguietes}(A) = \{\{ \} \cup \{ \}, \{, \$\}$$

Además,

$$SD(A' \rightarrow \{A\}AA') = \{\{ \}$$

$$SD(A' \rightarrow \lambda) = \text{Primeros}(\lambda) \cup \text{Siguietes}(A') = \emptyset \cup \text{Siguietes}(A) = \{ \}, \{, \$\}$$

Con la regla $A' \rightarrow \{A\}AA'$ ya se va a matchear $\}$, así que si $ct == '}'$, usamos la regla $A' \rightarrow \lambda$. También usamos $A' \rightarrow \lambda$ si $ct == '\$'$, dado que Main ya matchea $\$$.

```

Proc Main()
  if(ct != $) A();
  match('$');
  accept();
End

Proc A()
  A'();
End

Proc A'()
  if(ct == '{') {
    match('{');
    A();
    match('}');
    A();
    A'();
  }
  elseif(ct in {'}', '$') {}
  else{ error(); }
End

```

Otra opción es dar el parser para la gramática $\langle \{A\}, \{\{, \}\}, A, \{A \rightarrow \{A\}A \mid \lambda\} \rangle$.

b) $P = \{A \rightarrow +AA \mid -AA \mid a\}$, $V_T = \{+, -, a\}$, $V_N = \{A\}$

PRODUCCIÓN	SD
$A \rightarrow +AA$	$\{+\}$
$A \rightarrow -AA$	$\{-\}$
$A \rightarrow a$	$\{a\}$

```

Proc Main()
  A();
  match('$');
  accept();
End

Proc A()
  if(ct == '-') {
    match('-');
    A();
    A();
  }
  elseif(ct == '+') {
    match('+');
    A();
    A();
  }
  elseif(ct == 'a') {
    match('a');
  }
  else {
    error();
  }
End
    
```

c) $P = \{A \rightarrow 0A1 \mid 01\}$, $V_T = \{0, 1\}$, $V_N = \{A\}$

Hay que hacer *left-factorization*.

$$A \rightarrow 0A'$$

$$A' \rightarrow A1 \mid 1$$

Ahora ya está, no hay recursión no inmediata.

Ejercicio 2

Hecho en clase el 17/05/2017.

Dada la siguiente gramática para la declaración de tipos en un lenguaje de programación:

$G_1 = \langle \{S_0, T, S\}, \{\text{array}, [,], \text{of}, \dots, \text{int}, \text{char}, \text{num}, \uparrow\}, S_0, P \rangle$, con P :

$$S_0 \rightarrow T$$

$$T \rightarrow S \mid \uparrow S \mid \text{array of } T \mid \text{array } [S] \text{ of } T$$

$$S \rightarrow \text{int} \mid \text{char} \mid \text{num} \dots \text{num}$$

(a) ¿Es posible crear un parser recursivo descendente predictivo para G_1 ? Si no, crear una gramática G'_1 que genere el mismo lenguaje, y construir el parser para ella. Justificar.

No es posible hacer un parser descendente recursivo predictivo para la gramática, porque hay solapamiento entre los SDs de dos reglas del no terminal T .

Se lo resuelve aplicando factorización a la izquierda, para crear G'_1 .

$$G'_1 = \langle \{S_0, T, T', S\}, \{\text{array}, [,], \text{of}, \dots, \text{int}, \text{char}, \text{num}, \uparrow\}, S_0, P \rangle, \text{ con } P :$$

	PRODUCCIÓN	SD
$S_0 \rightarrow T$	$S_0 \rightarrow T$	{int, char, num, ↑, array}
$T \rightarrow S \mid \uparrow S \mid \text{array } T'$	$T \rightarrow S$	{int, char, num}
$T' \rightarrow \text{of } T \mid [S] \text{ of } T$	$T \rightarrow \uparrow S$	{↑}
$S \rightarrow \text{int} \mid \text{char} \mid \text{num} \dots \text{num}$	$T \rightarrow \text{array } T'$	{array}
	$T' \rightarrow \text{of } T$	{of}
	$T' \rightarrow [S] \text{ of } T$	{[]}
	$S \rightarrow \text{int}$	{int}
	$S \rightarrow \text{char}$	{char}
	$S \rightarrow \text{num} \dots \text{num}$	{num}

```

Proc T()
    if(ct in {int, char, num}){
        S();
    }
    elseif(ct == '^'){
        match('^');
        S();
    }
    elseif(ct == 'array'){
        match('array');
        T'();
    }
    else{
        error();
    }
}
End

Proc Main()
    T();
    match('$');
    accept();
End
    
```

```

Proc T'()
  if(ct == of){
    match('of');
    T();
  }
  elseif(ct == '['){
    match('[');
    S();
    match(']');
    match('of');
    T();
  }
  else{
    error();
  }
End

Proc S()
  if(ct == 'int'){
    match('int');
  }
  elseif(ct == 'char'){
    match('char');
  }
  elseif(ct == 'num'){
    match('num');
    match('..');
    match('num');
  }
  else{
    error();
  }
End

```

- (b) Mostrar la derivación y el árbol correspondiente, paso a paso, para la siguiente cadena:
array [num..num] of ↑ char

$$\begin{aligned}
 S_0 &\Rightarrow T \Rightarrow \text{array } T' \Rightarrow \text{array } [S] \text{ of } T \Rightarrow \text{array } [\text{num} \dots \text{num}] \text{ of } T \\
 &\Rightarrow \text{array } [\text{num} \dots \text{num}] \text{ of } \uparrow S \Rightarrow \text{array } [\text{num} \dots \text{num}] \text{ of } \uparrow \text{char}
 \end{aligned}$$

- (c) ¿Qué ocurriría con la cadena anterior si se la procesara con un parser descendente que no fuera predictivo?

2. Parsers LL(1)

Una gramática $G = \langle V, T, P, S \rangle$ es LL(1) sii

$$\forall (A \rightarrow \beta, A \rightarrow \gamma), \text{ con } \beta \neq \gamma, SD(A \rightarrow \beta) \cap SD(A \rightarrow \gamma) = \emptyset$$

2.1. Ejercicios

Ejercicio 3

- a) $G_1 = \langle \{S, A\}, \{a, b\}, P_1, S \rangle$ con $P_1 = \{S \rightarrow aAS \mid b, A \rightarrow a \mid bSA\}$

Tengo que ver si se da que

$$SD(S \rightarrow aAS) \cap SD(S \rightarrow b) = \emptyset = SD(A \rightarrow a) \cap SD(A \rightarrow bSA)$$

$SD(S \rightarrow aAS) = \text{Primeros}(aAS) = \{a\}$ porque aAS no es anulable.

$SD(S \rightarrow b) = \text{Primeros}(b) = \{b\}$ porque b no es anulable.

La intersección de la izquierda es vacía.

$SD(A \rightarrow a) = \{a\}$ y $SD(A \rightarrow bSA) = \{b\}$ dan como resultado que la intersección de la derecha también sea vacía.

Concluimos que la gramática G_1 es LL(1).

b) $G_2 = \langle \{S\}, \{a, b\}, P_2, S \rangle$ con $P_2 = \{S \rightarrow aaSbb \mid a \mid \lambda\}$

Afirmo que no es LL(1). Tomo el par de reglas $S \rightarrow aaSbb$ y $S \rightarrow a$.

$SD(S \rightarrow aaSbb) = Primeros(aaSbb) = \{a\}$ y $SD(S \rightarrow a) = Primeros(a) = \{a\}$ dan una intersección no vacía.

Hago *left-factoring* y obtengo

$$G'_2 = \langle \{S, S'\}, \{a, b\}, P'_2, S \rangle \text{ con } P'_2 = \{S \rightarrow aS' \mid \lambda, S' \rightarrow aSbb \mid \lambda\}$$

Tengo que verificar que

$$SD(S \rightarrow aS') \cap SD(S \rightarrow \lambda) = \emptyset = SD(S' \rightarrow aSbb) \cap SD(S' \rightarrow \lambda)$$

$$\begin{aligned} SD(S \rightarrow aS') \cap SD(S \rightarrow \lambda) &= Primeros(aS') \cap (Primeros(\lambda) \cup Siguients(S)) \\ &= \{a\} \cap (\emptyset \cup \{\$\}) \\ &= \emptyset \end{aligned}$$

$$\begin{aligned} SD(S' \rightarrow aSbb) \cap SD(S' \rightarrow \lambda) &= Primeros(aSbb) \cap (Primeros(\lambda) \cup Siguients(S')) \\ &= \{a\} \cap (\emptyset \cup Siguients(S)) \\ &= \{a\} \cap (\emptyset \cup \{\$\}) \\ &= \emptyset \end{aligned}$$

Ejercicio 4

Crear el parser LL(1) para la gramática del Ejercicio 2.

Al ver los SD para las producciones nos damos cuenta de que la gramática del enunciado no es LL(1) (como vimos en el Ejercicio 2). La gramática corregida que propusimos, G'_1 , sí es LL(1) (como podemos ver en la tabla del Ejercicio 2).

Así, hacemos la tabla del parser LL(1).

NT	SÍMBOLO DE ENTRADA								
	array	[]	of	..	int	char	num	↑
S_0	$S_0 \rightarrow T$					$S_0 \rightarrow T$	$S_0 \rightarrow T$	$S_0 \rightarrow T$	$S_0 \rightarrow T$
T	array T'					$T \rightarrow S$	$T \rightarrow S$	$T \rightarrow S$	$T \rightarrow \uparrow S$
T'		$T' \rightarrow$ [S] of T		$T' \rightarrow$ of T					
S						$S \rightarrow$ int	$S \rightarrow$ char	$S \rightarrow$ num .. num	

3. Gramáticas con conflictos

3.1. Ejercicios

Ejercicio 5

Hecho en clase 18/05/2017.

Ejercicio 6

Dada la siguiente gramática que genera todas las expresiones regulares sobre los terminales $\{a, b, c\}$, indicar si es LL(1). Si no lo es, indicar qué cambios habría que realizarle para que lo fuera, sin modificar el lenguaje generado, o si se podría resolver de otra manera (e.g., ajustando la tabla del parser). $G_2 = \langle \{E\}, \{a, b, c, (,), |, *, \}, E, P_2 \rangle$, con P_2 :

$$E \rightarrow E|E \mid EE \mid E^* \mid (E) \mid a \mid b \mid c$$

Calculamos los SD para cada regla.

$$SD(E \rightarrow E|E) = \text{Primeros}(E|E) = \{(, a, b, c\}$$

$$SD(E \rightarrow EE) = \text{Primeros}(EE) = \{(, a, b, c\}$$

$$SD(E \rightarrow E^*) = \text{Primeros}(E^*) = \{(, a, b, c\}$$

$$SD(E \rightarrow (E)) = \text{Primeros}((E)) = \{(}$$

$$SD(E \rightarrow a) = \text{Primeros}(a) = \{a\}$$

$$SD(E \rightarrow b) = \text{Primeros}(b) = \{b\}$$

$$SD(E \rightarrow c) = \text{Primeros}(c) = \{c\}$$

Claramente no es LL(1).

NT	SÍMBOLO DE ENTRADA			
	(a	b	c
E	$E \rightarrow E E$	$E \rightarrow E E$	$E \rightarrow E E$	$E \rightarrow E E$
	$E \rightarrow EE$	$E \rightarrow EE$	$E \rightarrow EE$	$E \rightarrow EE$
	$E \rightarrow E^*$	$E \rightarrow E^*$	$E \rightarrow E^*$	$E \rightarrow E^*$
	$E \rightarrow (E)$	$E \rightarrow a$	$E \rightarrow b$	$E \rightarrow c$

No es posible resolver los conflictos ajustando la tabla del parser. Hay que modificar la gramática para marcar la precedencia de los operadores con distintos símbolos no terminales o eliminar la recursión inmediata.

Probamos eliminar la recursión inmediata.

$$E \rightarrow (E)E' \mid aE' \mid bE' \mid cE'$$

$$E' \rightarrow |EE' \mid EE' \mid *E' \mid \lambda$$

PRODUCCIÓN	SD
$E \rightarrow (E)E'$	$\{(}$
$E \rightarrow aE'$	$\{a\}$
$E \rightarrow bE'S$	$\{b\}$
$E \rightarrow cE'$	$\{c\}$
$E' \rightarrow EE'$	$\{ $
$E' \rightarrow EE'$	$\{(, a, b, c\}$
$E' \rightarrow *E'$	$\{*\}$
$E' \rightarrow \lambda$	$\{ , (, a, b, c, *\}$

No funciona, vamos a probar agregando no terminales para marcar la precedencia.

	PRODUCCIÓN	SD
	$E \rightarrow E E_1$	$\{ (, a, b, c \}$
	$E \rightarrow E_1$	$\{ (, a, b, c \}$
	$E_1 \rightarrow E_1 E_2$	$\{ (, a, b, c \}$
	$E_1 \rightarrow E_2$	$\{ (, a, b, c \}$
	$E_2 \rightarrow E_3^*$	$\{ (, a, b, c \}$
	$E_2 \rightarrow E_3$	$\{ (, a, b, c \}$
	$E_3 \rightarrow (E)$	$\{ (\}$
	$E_3 \rightarrow a$	$\{ a \}$
	$E_3 \rightarrow b$	$\{ b \}$
	$E_3 \rightarrow c$	$\{ c \}$

Eliminemos la recursión y hagamos factorización.

$$\begin{aligned}
 E &\rightarrow E_1 E' \\
 E' &\rightarrow | E_1 E' | \lambda \\
 E_1 &\rightarrow E_2 E_1' \\
 E_1' &\rightarrow E_2 E_1' | \lambda \\
 E_2 &\rightarrow E_3 E_3' \\
 E_3' &\rightarrow * | \lambda \\
 E_3 &\rightarrow (E) | a | b | c
 \end{aligned}$$

PRODUCCIÓN	SD
$E \rightarrow E_1 E'$	$Primeros(E_1) = Primeros(E_2) = Primeros(E_3) = \{ (, a, b, c \}$
$E' \rightarrow E_1 E'$	$\{ \}$
$E' \rightarrow \lambda$	$Sigüientes(E') = Sigüientes(E) = \{ \$,) \}$
$E_1 \rightarrow E_2 E_1'$	$Primeros(E_2) = \{ (, a, b, c \}$
$E_1' \rightarrow E_2 E_1'$	$Primeros(E_2) = \{ (, a, b, c \}$
$E_1' \rightarrow \lambda$	$Sigüientes(E_1') = Sigüientes(E_1)$ $= Primeros(E') \cup Sigüientes(E')$ $= \{ \} \cup Sigüientes(E) = \{ ,), \$ \}$
$E_2 \rightarrow E_3 E_3'$	$Primeros(E_3) = \{ (, a, b, c \}$
$E_3' \rightarrow *$	$\{ * \}$
$E_3' \rightarrow \lambda$	$Sigüientes(E_3') = Sigüientes(E_2) = Primeros(E_1') \cup Sigüientes(E_1)$ $= Primeros(E_2) \cup (Primeros(E') \cup Sigüientes(E'))$ $= Primeros(E_3) \cup (\{ \} \cup Sigüientes(E)) = \{ (, a, b, c \} \cup (\{ \} \cup \{ \$,) \})$ $= \{ (, a, b, c, , \$,) \}$
$E_3 \rightarrow (E)$	$\{ (\}$
$E_3 \rightarrow a$	$\{ a \}$
$E_3 \rightarrow b$	$\{ b \}$
$E_3 \rightarrow c$	$\{ c \}$

Esta gramática sí es LL(1), habría que ver que el lenguaje generado es el mismo que el de la gramática original.

4. Gramáticas extendidas (ELL) y parsers recursivos iterativos

4.1. Teoría

Supongamos que tenemos la siguiente gramática:

$$\begin{aligned} L &\rightarrow E \mid L, E \\ E &\rightarrow n \mid (L) \end{aligned}$$

Escribamos

$$\begin{aligned} L &\rightarrow E(, E)^* \\ E &\rightarrow n \mid (L) \end{aligned}$$

y obtenemos los siguientes códigos:

$L \rightarrow E(, E)^*$ <pre> Proc L(): E(); while(tc == ','){ match(','); E(); } End </pre>	$E \rightarrow n \mid (L)$ <pre> Proc E(): if(tc == 'n'){match('n');} elseif(tc == '('){ match('('); L(); match(')'); } else error(); End </pre>
---	--

4.1.1. Gramáticas extendidas

Las gramáticas extendidas y los métodos ELL permiten escribir usando ERs y generar parsers descendentes iterativos-recursivos.

Definición: Una gramática extendida es una tupla $\langle V_N, V_T, p, S \rangle$ con $p : V_N \rightarrow ER(V)$ ($V = V_N \cup V_T$).

p puede ser una función porque se pueden combinar las producciones para un mismo no terminal con la unión.

Operadores: *, +, ?. * se puede implementar con while, el + con do while y el ? con un if.

4.1.2. Generación de código y Transformación de gramáticas

Definimos por inducción la función $Cod(E)$ que recibe una expresión regular sobre V y devuelve el código correspondiente:

E	$\text{Cod}(E)$	E no extendida
λ	<code>skip;</code>	
a	<code>match('a')</code>	
$R?$	<code>if (tc in Primeros(R)) { Cod(R) }</code>	$A \rightarrow R \mid \lambda$
R^*	<code>while (tc in Primeros(R)) { Cod(R) }</code>	$A \rightarrow RA \mid \lambda$
R^+	<code>do { Cod(R) } while (tc in Primeros(R))</code>	$A \rightarrow RR^*$
R_1R_2	<code>Cod(R₁); Cod(R₂);</code>	
$R_1 \mid R_2 \mid \dots \mid R_n$	<code>if (tc in SD(A → R₁)) { Cod(R₁) } eif (tc in SD(A → R₂)) { Cod(R₂) } ... else error</code>	$A \rightarrow R_1 \mid R_2 \mid \dots$

Notar que al agregar producciones de esta manera evitamos generar conflictos LL(1).

¿Cómo se usa?

El procedimiento consiste en reemplazar cada subexpresión (empezando por las más externas) por un no terminal nuevo y agregando las producciones que hagan falta.

Ejemplo: Si tenemos la producción $A \rightarrow a(bA?c)^*$ entonces se transforma en

$$A \rightarrow aA_1$$

$$A_1 \rightarrow bA?cA_1 \mid \lambda$$

y eso a su vez en

$$A \rightarrow aA_1$$

$$A_1 \rightarrow bA_2cA_1 \mid \lambda$$

$$A_2 \rightarrow A \mid \lambda$$

¿Es ELL(1)?

Si tenemos una Gramática Extendida EG y su GLC derivada es LL(1), decimos que EG es ELL(1) y el parser generado reconocerá exactamente las cadenas de $L(EG)$.

4.2. Ejercicios

Ejercicio 10

Dada la gramática $G_1 = \langle \{L, R, A, D\}, \{f, d, h, +, t\}, P, L \rangle$, con P :

$$L \rightarrow R \mid R + L$$

$$R \rightarrow A \mid RA$$

$$A \rightarrow dLhD \mid t \mid dLh$$

$$D \rightarrow \lambda \mid f$$

- (a) ¿Es G_1 ELL(1)? Justificar. Si no lo es, dar una gramática G_1' que sea ELL(1) y tal que $L(G_1') = L(G_1)$ usando al menos una vez cada uno de los siguientes operadores: *, + y ?.

Describir cómo se llega a G_1' a partir de G_1 y justificar por qué G_1' es ELL(1).

La gramática G_1 es equivalente a

$$\begin{aligned} L &\rightarrow R(+R)^* \\ R &\rightarrow A^+ \\ A &\rightarrow \text{dLh}(D?) \mid \mathfrak{t} \\ D &\rightarrow \mathfrak{f}? \end{aligned}$$

que es equivalente a

$$L \rightarrow (\text{dLh}(f?) \mid \mathfrak{t})^+ \left(+ (\text{dLh}(f?) \mid \mathfrak{t})^+ \right)^*$$

$$\begin{aligned} L &\rightarrow A_1A_2 \\ A_1 &\rightarrow A_3A_4 \\ A_2 &\rightarrow +A_1A_2 \mid \lambda \\ A_3 &\rightarrow \text{dLh}A_5 \mid \mathfrak{t} \\ A_4 &\rightarrow A_3A_4 \mid \lambda \\ A_5 &\rightarrow \mathfrak{f} \mid \lambda \end{aligned}$$

PRODUCCIÓN	SD
$L \rightarrow A_1A_2$	$SD(A_1) = SD(A_3) = \{\mathfrak{d}, \mathfrak{t}\}$
$A_1 \rightarrow A_3A_4$	$SD(A_3) = \{\mathfrak{d}, \mathfrak{t}\}$
$A_2 \rightarrow +A_1A_2$	$\{+\}$
$A_2 \rightarrow \lambda$	$Siguientes(A_2) = Siguietes(L) = \{\$\}$
$A_3 \rightarrow \text{dLh}A_5$	$\{\mathfrak{d}\}$
$A_3 \rightarrow \mathfrak{t}$	$\{\mathfrak{t}\}$
$A_4 \rightarrow A_3A_4$	$SD(A_3) = \{\mathfrak{d}, \mathfrak{t}\}$
$A_4 \rightarrow \lambda$	$Siguientes(A_4) = Siguietes(A_1)$ $= Primeros(A_2) \cup Siguietes(L)$ $= \{+\} \cup \{\$\} = \{+, \$\}$
$A_5 \rightarrow \mathfrak{f}$	$\{\mathfrak{f}\}$
$A_5 \rightarrow \lambda$	$Siguientes(A_5) = Siguietes(A_3)$ $= Primeros(A_4) \cup Siguietes(A_4)$ $= Primeros(A_3) \cup \{+, \$\}$ $= \{\mathfrak{d}, \mathfrak{t}\} \cup \{+, \$\} = \{\mathfrak{d}, \mathfrak{t}, +, \$\}$

(b) Dar el parser recursivo-iterativo de la gramática G_1' .

```
L():
  do{
    if(tc == d){
      match('d');
      L();
      match('h');
      if(tc == f) match('f');
    }
  }
```

```

    elseif(tc == t){
        match('t');
    }
    else error();
} while(tc in {d, t})
while(tc == +){
    match('+');
    do{
        if(tc == d){
            match('d');
            L();
            match('h');
            if(tc == f) match('f');
        }
        elseif(tc == t){
            match('t');
        }
        else error();
    } while(tc in {d, t})
}

```

Ejercicio 12

Dada la gramática extendida G_3 , que genera un directorio telefónico con entradas delimitadas por punto y coma (;), donde cada entrada tiene un nombre y un número de teléfono con, opcionalmente, un carácter y dos puntos (:) al comienzo, que indican bajo qué letra debe agendarse esa entrada (por defecto, cada entrada se agenda bajo la letra de comienzo del nombre, si lo hay). $G_3 = \langle \{A\}, \{a, \text{num}, ,, :, ;\}, A, P_3 \rangle$, con P_3 :

$$A \rightarrow \left(\left((a : a^*, \text{num}) \mid (a^*, \text{num}) \right); \right)^*$$

(a) Mostrar que G_3 no es ELL(1).

Si G_3 fuera ELL(1), la GLC derivada sería LL(1). Veamos que esto no es así.

$$\begin{aligned}
 A &\rightarrow A_1 \\
 A_1 &\rightarrow (a : a^*, \text{num} \mid a^*, \text{num}); A_1 \mid \lambda
 \end{aligned}$$

$$\begin{aligned}
 A &\rightarrow A_1 \\
 A_1 &\rightarrow (a : A_2, \text{num} \mid a^*, \text{num}); A_1 \mid \lambda \\
 A_2 &\rightarrow aA_2 \mid \lambda
 \end{aligned}$$

$$\begin{aligned}
 A &\rightarrow A_1 \\
 A_1 &\rightarrow (a : A_2, \text{num} \mid A_3, \text{num}); A_1 \mid \lambda \\
 A_2 &\rightarrow aA_2 \mid \lambda \\
 A_3 &\rightarrow aA_3 \mid \lambda
 \end{aligned}$$

$$\begin{aligned}
 A &\rightarrow A_1 \\
 A_1 &\rightarrow A_4; A_1 \mid \lambda \\
 A_2 &\rightarrow aA_2 \mid \lambda \\
 A_3 &\rightarrow aA_3 \mid \lambda \\
 A_4 &\rightarrow a : A_2, \text{num} \mid A_3, \text{num}
 \end{aligned}$$

Calculemos ahora los SD para las producciones de A_4 .

$$SD(A_4 \rightarrow a : A_2, \text{num}) = \{a\}$$

$$SD(A_4 \rightarrow A_3, \text{num}) = \text{Primeros}(A_3) \cup \text{Siguientes}(A_3) = \{a\} \cup \{,\}$$

La intersección es $\{a\} \neq \emptyset$, por lo que concluimos que la GLC derivada no es LL(1).

- (b) Crear una gramática de una sola producción para el lenguaje especificado, pero que sí sea ELL(1).

$$A \rightarrow \left((, \text{num}) \mid (a (: a^*, \text{num} \mid a^*, \text{num})); \right)^*$$

5. Limitaciones de los parsers descendentes

6. Material

■ Videos

- Elimination of left recursion and left factoring the grammars:
https://www.youtube.com/watch?v=3_VCoBfirt9c&t=4s
- Introduction to parsers and LL(1) parsing:
<https://www.youtube.com/watch?v=N9UuAPU6Dag>
- Examples on how to find first and follow in LL(1):
https://www.youtube.com/watch?v=_uS1P91jmTM&t=441s
- Construction of LL(1) parsing table:
<https://www.youtube.com/watch?v=R1Z1WEZWMKk>
- Recursive descent parser:
<https://www.youtube.com/watch?v=SH5F-rwWEog>
- Operator grammar and Operator precedence parser:
https://www.youtube.com/watch?v=n5UWAaw_byw&t=1s

■ Libro del Dragón.

- Top-Down Parsing: Capítulo 4.4.