Compiladores e Intérpretes

3^{er} curso 2^o cuatrimestre

Tema 8. Análisis sintáctico descendente

- 8.1 Introducción
- 8.2 Análisis descendente no determinista: backtracking
- 8.3 Análisis descendente determinista: LL(1)
 - 8.3.1 Gramáticas LL(1)
 - 8.3.2 Uso de la tabla de análisis

8.1

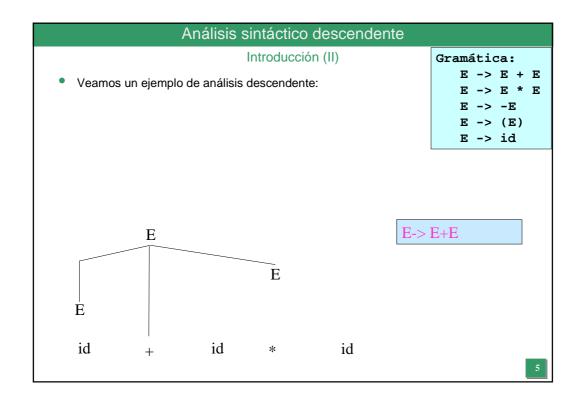
Introducción

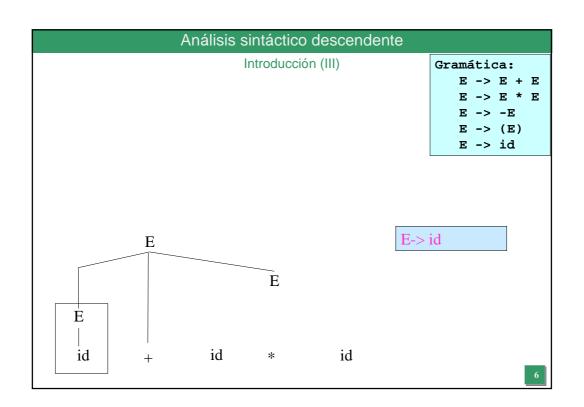
ŧ

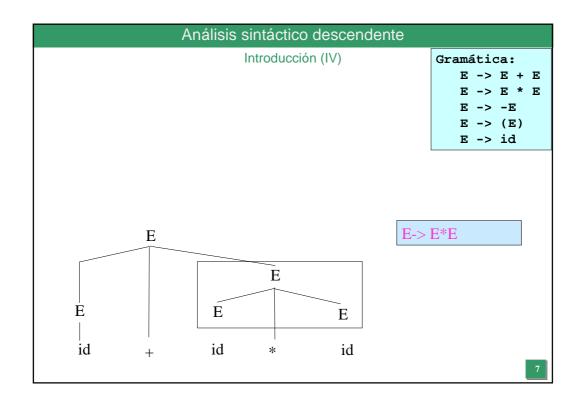
Análisis sintáctico descendente

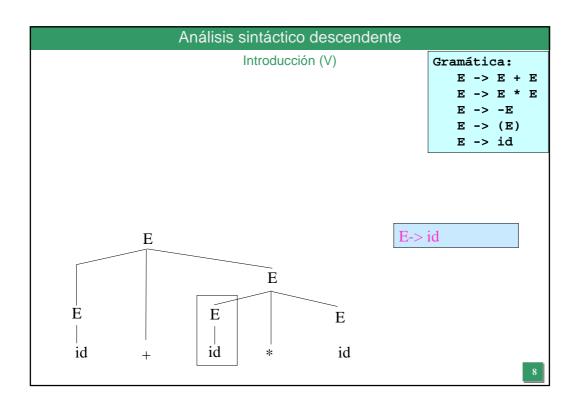
Introducción (I)

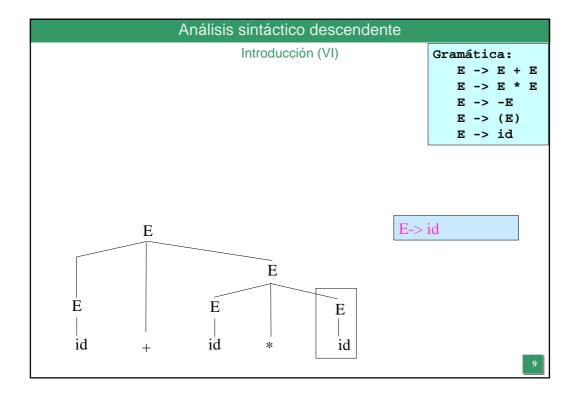
- Hasta ahora hemos estudiado cómo hacer análisis ascendente con las técnicas LR, SLR y LALR. Ahora vamos a ver varios enfoques descendentes.
- La principal diferencia es que ahora el análisis parte del axioma, y hay que encontrar una derivación que lleve hasta el programa proporcionado como entrada.
- El problema es que el número de posibles derivaciones es muy grande.
- Ya se ha estudiado en asignaturas como IA la existencia de estrategias generales (para cualquier problema) de búsqueda ciega y sus propiedades de las que se mencionan algunas:
 - Las dos estrategias básicas son en anchura y en profundidad.
 - La búsqueda en anchura expande todos los nodos del mismo nivel antes de descender a los hijos. De esta forma, siempre se puede conseguir encontrar la solución óptima (aquella para la que se requiere un menor número de pasos).
 - La búsqueda en profundidad expande todos los nodos hijos de un nodo antes que los nodos hermanos. En profundidad, excepto árboles infinitos, se suele tener oportunidad de encontrar alguna solución con menos esfuerzo.











Análisis sintáctico descendente

Introducción (VII)

- En las técnicas vistas hasta ahora (ascendentes) siempre se ha sabido en cada circunstancia "qué regla de la gramática utilizar (reducir)"
- En las técnicas descendentes se tiene el mismo problema: hay que asegurarse de que siempre se conoce qué regla de la gramática aplicar en función del símbolo de la entrada que se está analizando y el estado del análisis en el que nos encontremos
- Una aproximación simple e intuitiva al análisis descendente puede abordar este problema probando reglas y, en caso de no poder aplicar más para analizar la cadena de entrada, volver hacia atrás para probar otras (backtracking).
- Sin embargo, usar backtracking es poco eficiente, por lo que la técnica descendente más importante que veremos es la LL(1):
 - Con procesamiento Left-To-Right (como los analizadores que ya hemos visto).
 - Expandiendo el símbolo no terminal más a la izquierda (Leftmost) primero.
 - Requiere que la gramática sea LL(1).

8.2

Análisis descendente con "backtracking" (o con vuelta a atrás)

П

Análisis sintáctico descendente

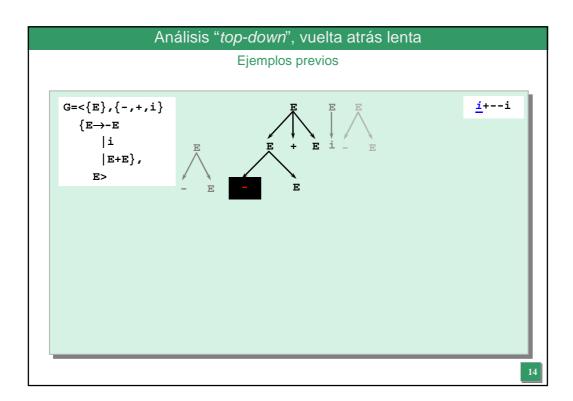
"Top-down" con vuelta atrás lenta: conceptos

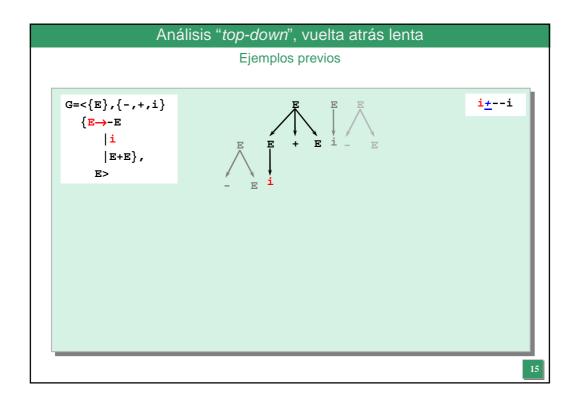
- Podríamos considerar el análisis sintáctico top-down con vuelta atrás lenta como un caso particular de búsqueda ciega en profundidad (se desempata de izquierda a derecha)
- El criterio para decidir si una regla es aplicable en una posición es la coincidencia de esa posición con el no terminal que está en la parte izquierda de la regla.
- No podemos continuar por un camino cuando tengamos terminales que discrepen con el programa (podemos suponer el orden de recorrido natural del programa (de izquierda a derecha)
- Habremos terminado el proceso en dos situaciones:
 - Si el árbol de derivación "se cierra" porque se ha llegado a eliminar todos los no terminales y se obtiene la misma secuencia de terminales del programa. En este caso se ha construido un árbol sintáctico para el programa que resulta ser correcto
 - Si no se ha podido "cerrar" el árbol de derivación pero se han probado todas las reglas en cada posición y no se puede aplicar ninguna otra opción. En ese caso se ha demostrado que el programa es sintácticamente incorrecto.

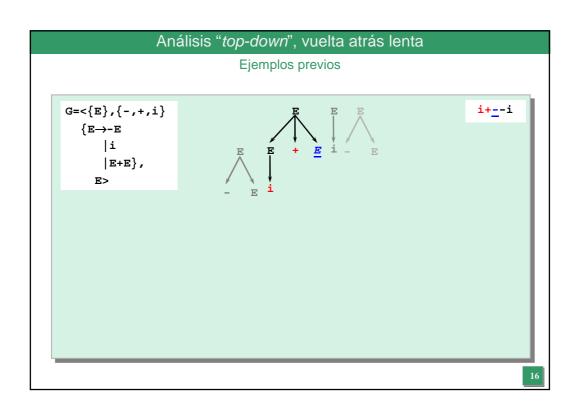
Análisis sintáctico descendente

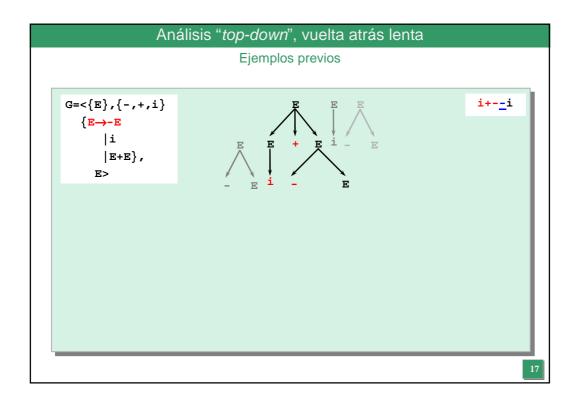
"Top-down" con vuelta atrás lenta: algoritmo

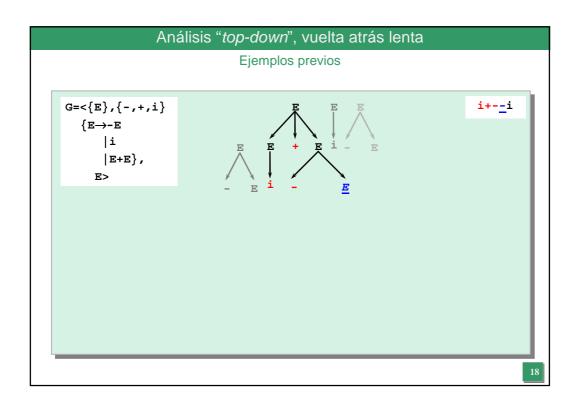
- 1. Símbolo actual = axioma
- 2. Encontrar las reglas que expanden el símbolo actual.
- 3. Si sólo hay una regla, aplicarla.
- 4. Si hay varias reglas, aplicar la primera y mantener las otras en reserva.
- 5. Por cada símbolo de la parte derecha de la regla, repetir desde el paso 2.
- 6. Cuando el análisis falle porque la regla requiere que el siguiente símbolo de la entrada sea un cierto terminal pero no lo es, o la expansión del axioma es correcta pero no cubre toda la cadena de entrada:
 - 1. Deshacer la aplicación de esta regla.
 - 2. Volver al punto de elección de la regla.
 - 3. Elegir la siguiente alternativa.
- 7. El análisis termina con éxito si:
 - 1. Se pueden generar todos los símbolos terminales de la cadena de entrada a partir de la expansión de los no terminales.
 - 2. No se puede analizar la cadena/programa de entrada puesto que es inválida y aunque se han intentado probar todas las reglas en cada paso, no hay ninguna derivación posible.

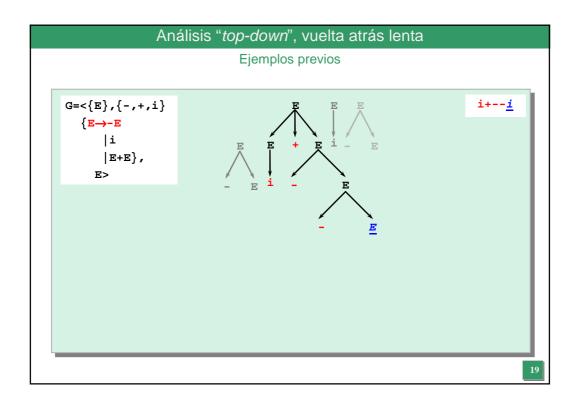


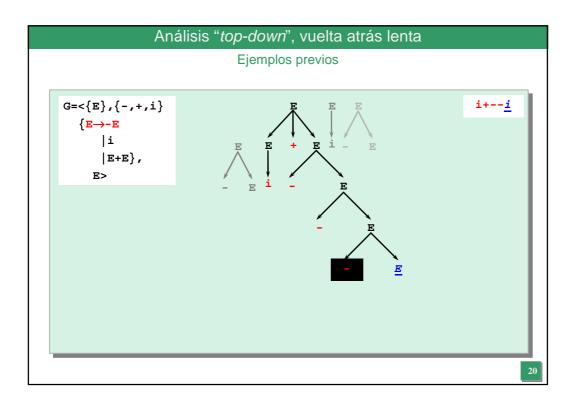


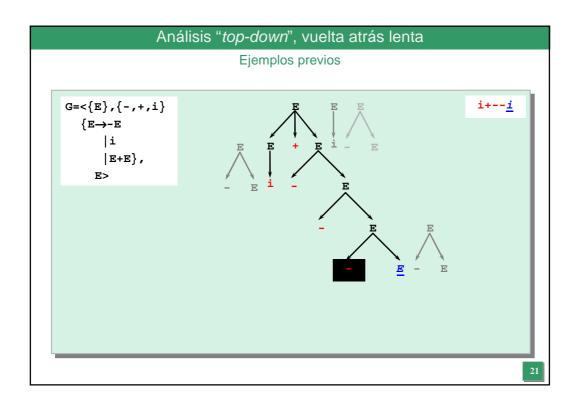


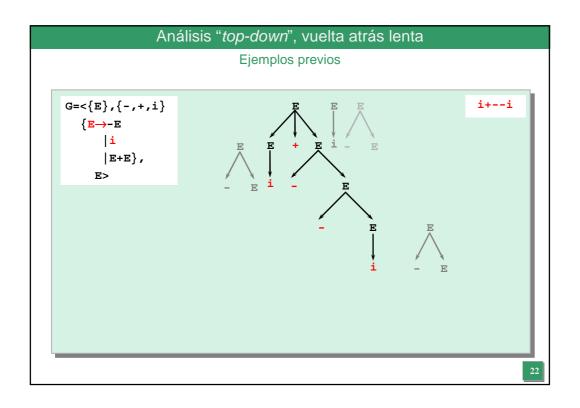


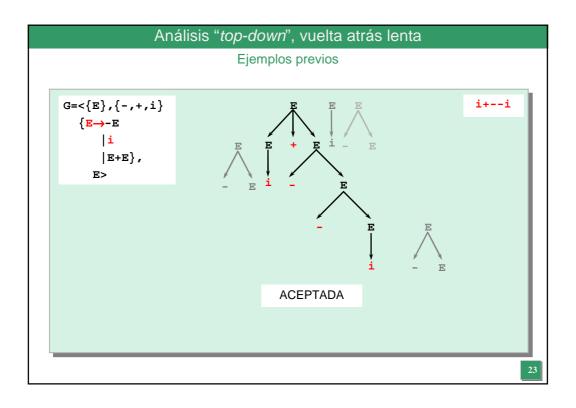


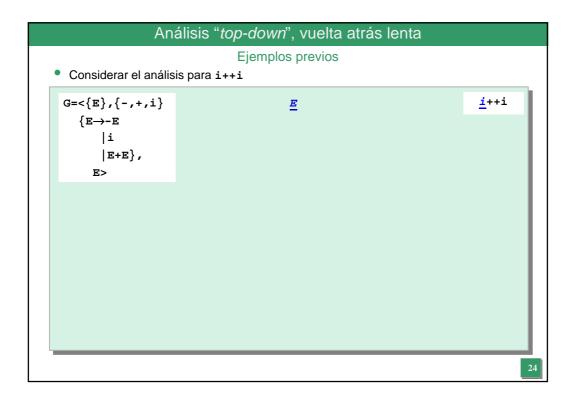


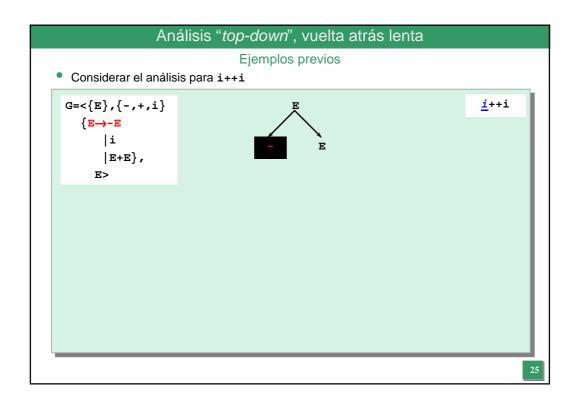


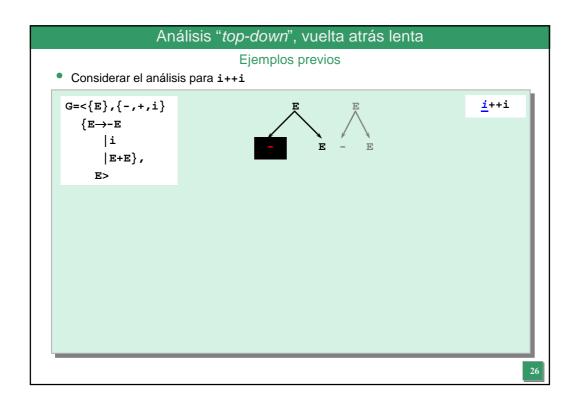


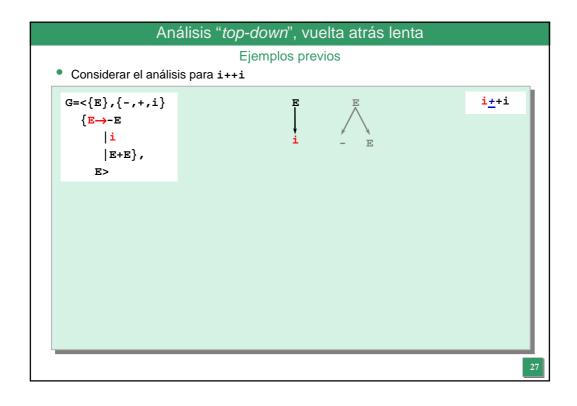


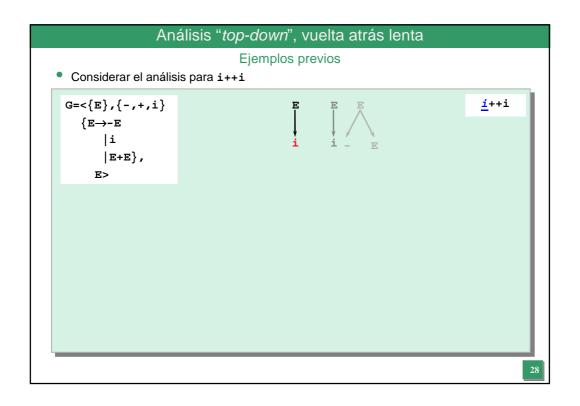


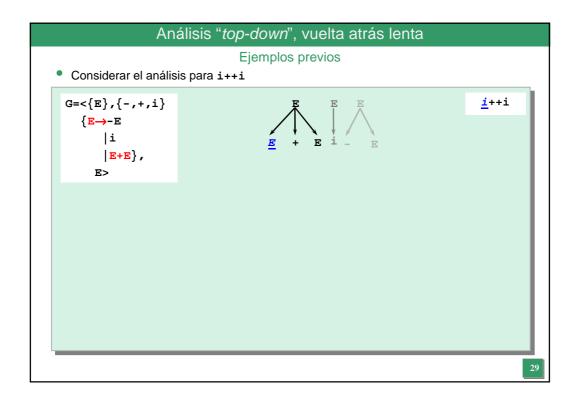


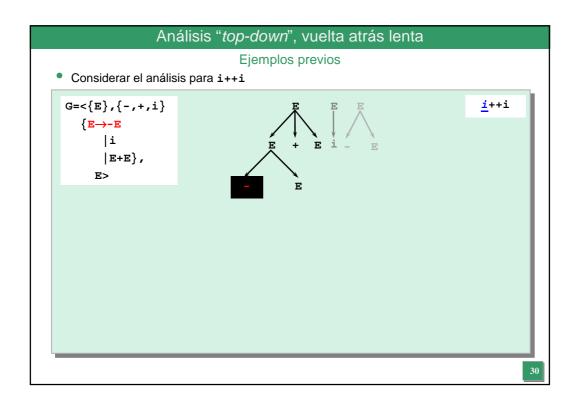


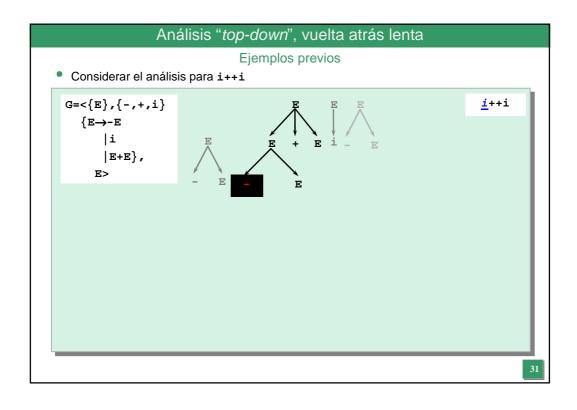


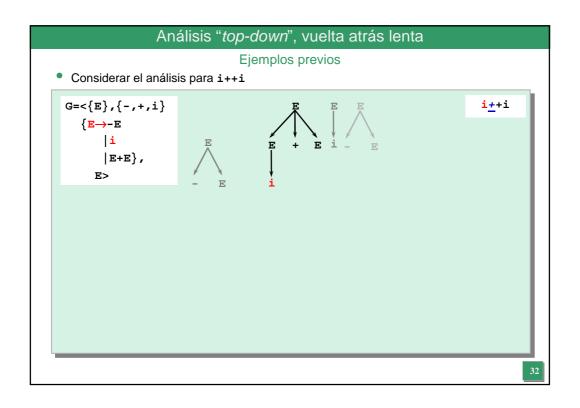


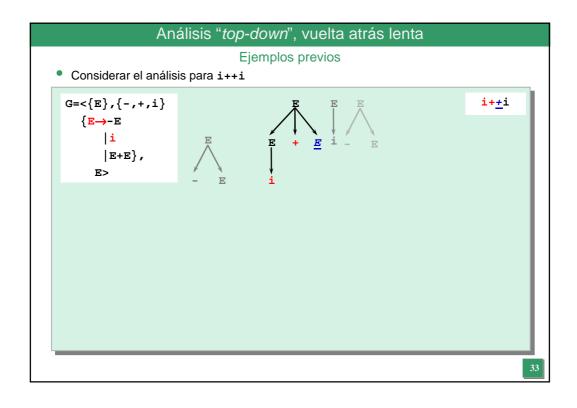


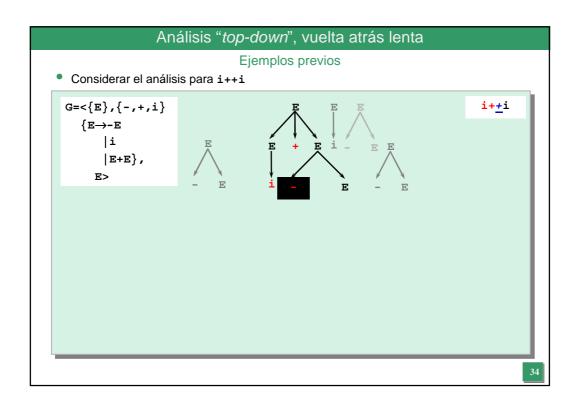


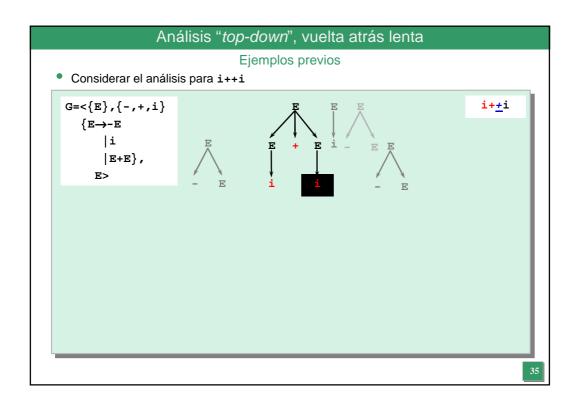


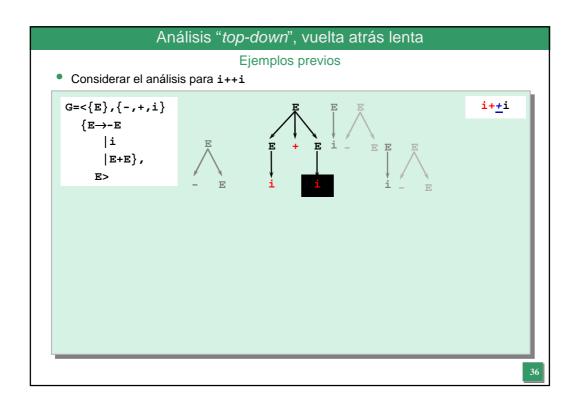


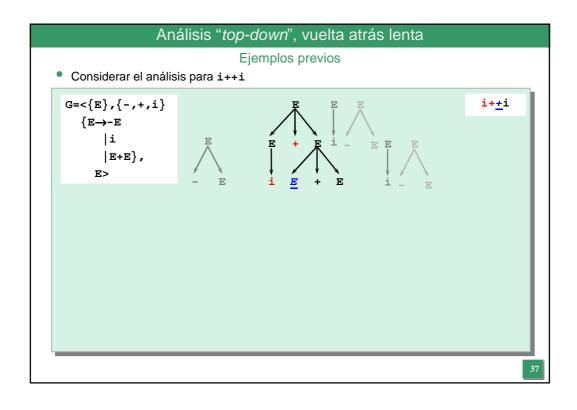


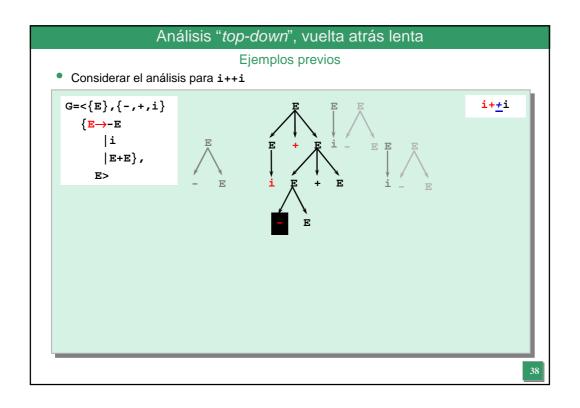


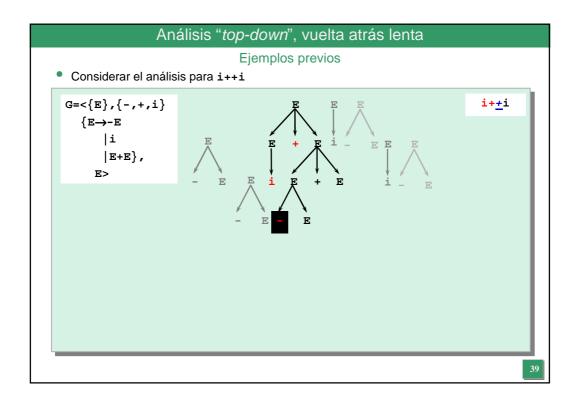


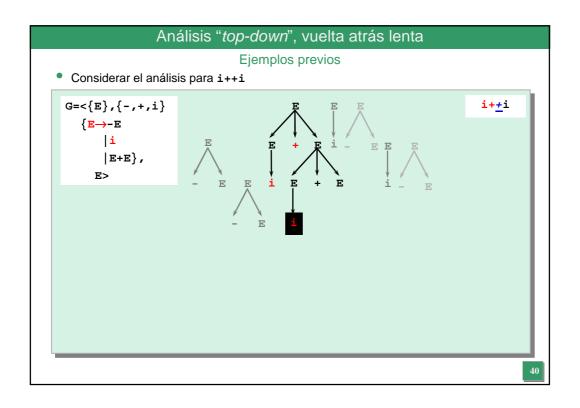


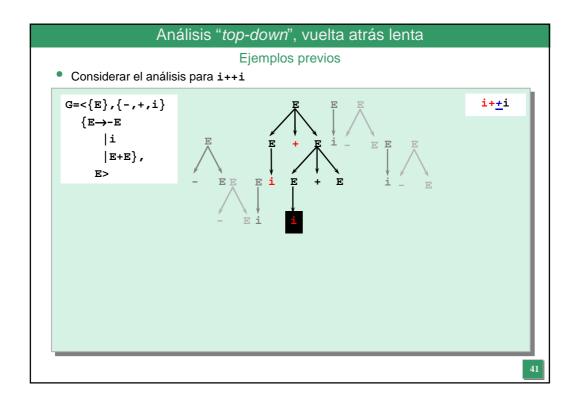


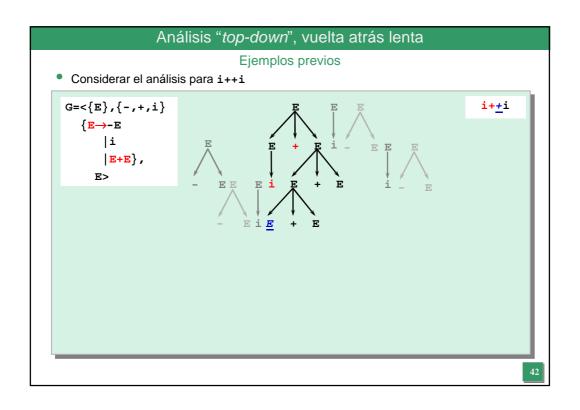


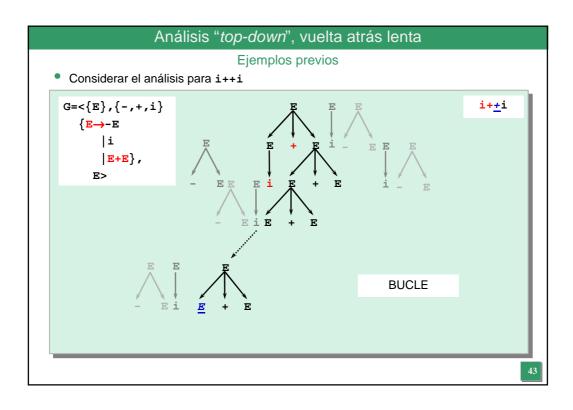












Análisis "top-down", vuelta atrás lenta

Conclusiones

- Como el alumno habrá observado, la presencia en la gramática de reglas recursivas por la izquierda hace que este tipo de análisis encuentre bucles infinitos de los que no puede salir.
- Hay maneras de solucionar este problema (entre otros) y construir analizadores sintácticos descendentes. Este tipo de analizadores suele llamarse LL(k)
- Uno de los objetivos de este tema es demostrar que se pueden construir fácilmente analizadores de tipo LL(1) (la entrada se lee desde la izquierda left las derivaciones en el árbol se hacen de izquierda left a derecha y se necesita conocer sólo 1 símbolo de la entrada por anticipado) si la gramática se puede expresar en forma adecuada

8.3

Análisis descendente determinista: LL(1) (sin vuelta atrás)

45

Análisis "top-down" selectivo

Concepto

- También se llama "sin vuelta atrás", en oposición a la técnica descrita al principio del tema.
- También se llama "descenso recursivo" por la técnica de paso automático de la gramática LL(1) al analizador.
- Intuitivamente, la razón de su eficiencia reside en que las partes derechas de cada no terminal pueden considerarse "indexadas" por el siguiente terminal.
- De hecho con LL(1) se puede construir una tabla de doble entrada (no terminales / terminales) en la que, cada celda, está ocupada solamente por una regla que es la única opción que se puede aplicar cuando en el "árbol sintáctico" corresponda tratar el no terminal y en el recorrido del programa toque tratar el terminal.

8.3.1

Gramática LL(1) mediante la construcción de la tabla de análisis descendente

47

Análisis LL(1) con tabla de análisis

Algoritmo de cálculo

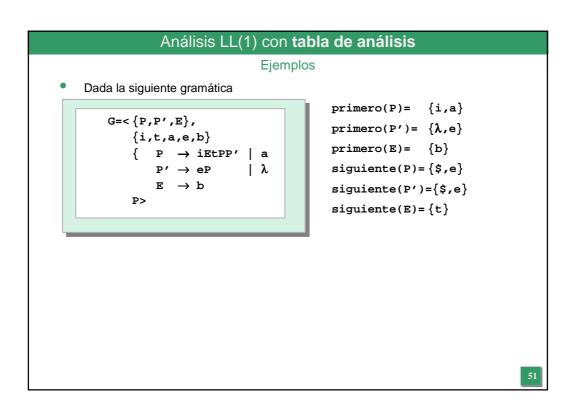
• Describiremos ahora el algoritmo de cálculo de la tabla de análisis $T \in M_{|\Sigma^n| \times |\Sigma^n| + 1}$ (matriz con una filas para los no terminales y columnas para los terminales añadiendo el delimitador \$):

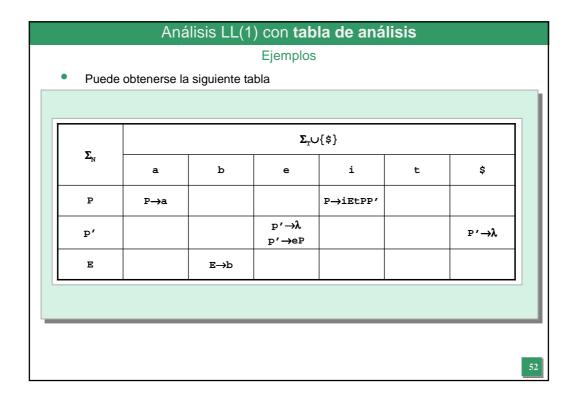
Se aplica los siguientes pasos

- 1. $\forall A \rightarrow \alpha \in P$ realizar el proceso
 - 1. $\forall a \in \text{primero}(\alpha) \cap \Sigma_{\text{T}} \qquad A \rightarrow \alpha \in T[A,a].$
 - 2. Si $\lambda \in \text{primero}(\alpha)$
 - Entonces \forall besigniente(A) $A \rightarrow \alpha \in T[A,b]$ (obsérvese que b puede ser también \$)

Análisis LL(1) con tabla de análisis **Ejemplos** En la gramática de los ejemplos anteriores G'=<{E,E',T,T',F},{-,*,i} primero(E)= {i,(} $\{ E \rightarrow TE'$ $primero(E') = \{+,\lambda\}$ primero(T)= {i,(} $E' \rightarrow +TE' \mid \lambda$ T→FT′ primero(T')= $\{*,\lambda\}$ $T' \rightarrow *FT' | \lambda$ primero(F)= {i,(} $F \rightarrow i \mid (E) \}$, $siguiente(E) = {),\$}$ siguiente(E')={),\$} E> siguiente(T)={),\$,+} siguiente(T')={),\$,+} siguiente(F)={),\$,+,*}

Análisis LL(1) con tabla de análisis **Ejemplos** En la gramática de los ejemplos anteriores $\Sigma_{\scriptscriptstyle T} \cup \{\$\}$ $\Sigma_{\!\scriptscriptstyle N}$ id E E→TE′ $E \rightarrow TE'$ E' $E' \rightarrow +TE'$ $E' \rightarrow \lambda$ $E' \rightarrow \lambda$ Т $T \rightarrow FT'$ $T \rightarrow FT'$ T' $T' \rightarrow \lambda$ $T' \rightarrow *FT'$ ${\tt T'} {\to} {\lambda}$ ${\tt T'} {\to} {\lambda}$ $F \rightarrow id$ $F \rightarrow (E)$





Análisis LL(1) con tabla de análisis

Definición

Podemos definir las gramáticas LL(1) como aquéllas que cumplen

Que la tabla de análisis construida con el algoritmo explicado anteriormente es determinista (todas las casillas tienen, a lo más, una regla)

Ejemplos

- Es fácil deducir que
 - La gramática del primer ejemplo de tabla es LL(1)
 - La segunda no lo es.

53

8.3.2

Uso de la tabla de análisis descendente

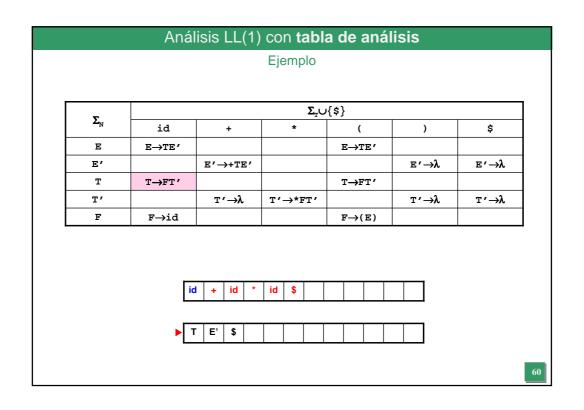
Análisis LL(1) con tabla de análisis Algoritmo de análisis A partir de la tabla de análisis puede utilizarse el siguiente algoritmo AnalisisLL(1)(char * cadena."\$") char * simbolo_actual; pila pila_analisis; simbolo_actual = cadena; push(pila_analisis, `\$'); push(pila_analisis, S); /* S es el axioma */ while (cima(pila_analisis) != `\$') { if (terminal(cima(pila_analisis)) || cima(pila_analisis) == `\$') if (cima(pila_analisis) == *simbolo_actual) { pop(pila_analisis); simbolo_actual++; } else salir_error(); else if (M[cima(pila_analisis), *simbolo_actual] != null) { pop(pila_analisis); push_cadena(*)(pila_analisis, reverse(parte_derecha(M[cima(pila_analisis), *simbolo_actual]))); else salir_error(); } (*) push_cadena(pila, cadena) inserta en la pila una cadena

Análisis LL(1) con tabla de análisis Ejemplo $\Sigma_{T} \cup \{\$\}$ $\Sigma_{_{\! N}}$ id) \$ E E→TE′ $E \rightarrow TE'$ E' E'→+TE' $E' \rightarrow \lambda$ Т $T \rightarrow FT'$ $T \rightarrow FT'$ Т' т′→λ $T' \rightarrow *FT'$ $T' \rightarrow \lambda$ **T′**→λ F $F \rightarrow id$ $F \rightarrow (E)$ id \$ Ε

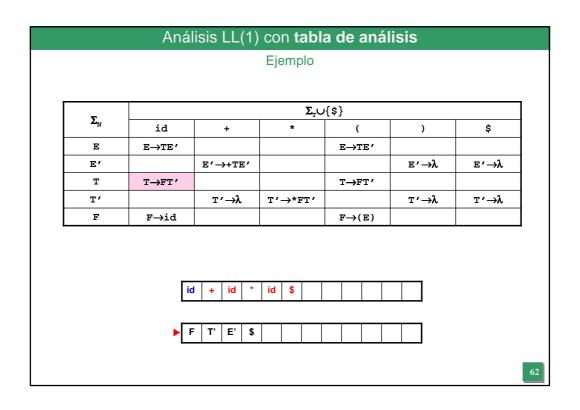
| | | | Ejemplo | · | | |
|--------------------|-----------|--------------|---|-------|---------------|--------------|
| | | | Ljempio | | | |
| | | | | | | |
| | 1 | | 7 | (A) | | |
| $\Sigma_{_{ m N}}$ | | T | $\Sigma_{\scriptscriptstyle m T}$ \cup | | | |
| | id | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | E ′→λ | E′→7 |
| T | T→FT′ | | | T→FT′ | | |
| T' | | T ′→λ | T'→*FT' | | Τ′ →λ | Τ′→ <i>î</i> |
| F | F→id | | | F→(E) | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | io | + id * | id \$ | | $\overline{}$ | |
| | ic | d + id * | id \$ | | | |
| | ic | 1 + id * | id \$ | | | |
| | io ▶ [| | id \$ | | | |

| | | | Ejemplo | | | |
|---|-----------|--------------|---------|-------|--------------|------|
| | | | Ejempio | | | |
| | | | | | | |
| | _ | | 7 | {\$} | | |
| $\mathbf{\Sigma}_{\!\scriptscriptstyle \mathrm{N}}$ | id | Τ . | * | | | |
| | 10 | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | Ε ′→λ | E'→7 |
| T | T→FT′ | | | T→FT′ | | |
| T' | | Τ ′→λ | T'→*FT' | | Τ′ →λ | т′→й |
| F | F→id | | | F→(E) | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | г. | | 1 | | | |
| | i i | d + id * | id \$ | | | |
| | | | | | | |
| | Ľ | | | | | |
| | <u> </u> | <u> </u> | | | | |
| | L" ▶[: | \$ | | | | |

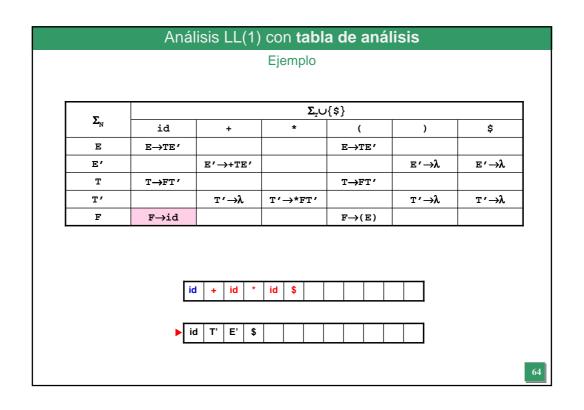
| | | | Ejemplo | | | |
|--|-------|--------------|---|-------|--------------|------|
| | | | LJempio | | | |
| | | | | | | |
| | | | $\Sigma_{\scriptscriptstyle m T} \cup$ | ι e J | | |
| $\mathbf{\Sigma}_{\!\scriptscriptstyle{\mathrm{N}}}$ | | Ι. | * | | ` | |
| | id | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | Ε′ →λ | E′→λ |
| T | T→FT′ | | | T→FT′ | | |
| T' | | T′ →λ | T'→*FT' | | T′ →λ | T′→λ |
| F | F→id | | | F→(E) | | |
| | _ | | | | | |
| | | | | | | |
| | ic | t + id * | id \$ | | | |
| | ic | d + id * | id \$ | | | |
| | ic | | id \$ | | | |



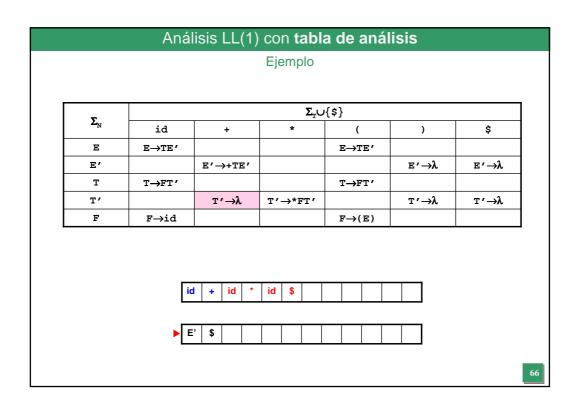
| | | | Ejemplo | <u> </u> | | |
|--------------------|-------|--------------|---|------------|--------------|--------------|
| | | | Ejempio | | | |
| | | | | | | |
| | 1 | | Σ | .(A) | | |
| $\Sigma_{_{ m N}}$ | | _ | $\Sigma_{\scriptscriptstyle m T}$ \cup | /{\$} □ | I | T |
| _N | id | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | E′→λ | E'→ |
| T | T→FT′ | | | T→FT′ | | |
| T' | | Τ′ →λ | T'→*FT' | | T′ →λ | Τ′→ <i>î</i> |
| | | | | | | |
| F | F→id | | | F→(E) | | |
| F | F→id | | | F→(E) | | |
| F | F→id | | | F→(E) | | |
| F | F→id | | | F→(E) | | |
| F | | id * | i.u e | F→(E) | | |
| F | F→id | 1 + id * | id \$ | F→(E) | | |
| F | | 1 + id * | id \$ | F→(E) | | |
| F | | | id \$ | F→(E) | | |



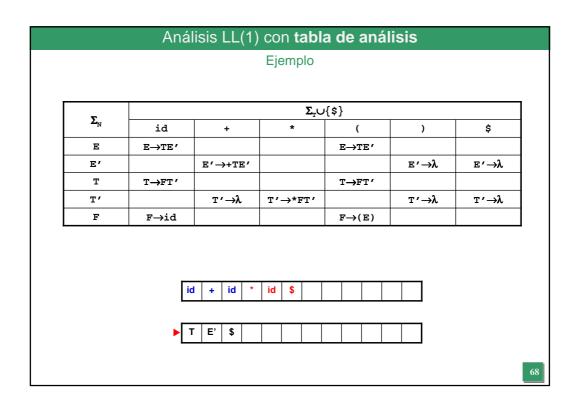
| | | | Ejemplo | | | |
|--------------------|----------|--------------|------------------------------------|---------------|--------------|------|
| | | | | | | |
| | | | | | | |
| ~ | | | $\Sigma_{\scriptscriptstyle m T}$ | \ {\$} | | |
| $\Sigma_{_{ m N}}$ | id | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | E ′→λ | E'→ |
| T | T→FT′ | | | T→FT′ | | |
| T' | | T′ →λ | T'→*FT' | | T ′→λ | T′→/ |
| F | F→id | | | F→(E) | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | io | d + id * | id \$ | | | |
| | <u> </u> | | | | | |
| | | | | | | |
| | _ | | | | | |
| | ▶ 1 | " E' \$ | | | | |



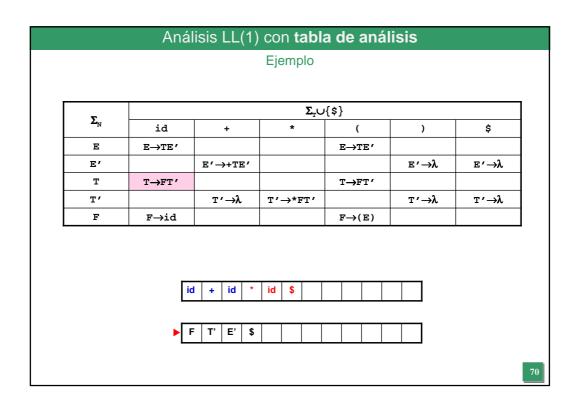
| | | | Ejemplo | | | |
|--------------------|-------|--------------|------------------------------------|-------|--------------|--------------|
| | | | Ejempio | | | |
| | | | | | | |
| | | | T | .(A) | | |
| $\Sigma_{_{ m N}}$ | | T | $\Sigma_{\scriptscriptstyle m T}$ | | Γ | T . |
| N | id | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | E′→λ | Ε'→ λ |
| T | T→FT′ | | | T→FT′ | | |
| T' | | T ′→λ | T'→*FT' | | T ′→λ | T′→λ |
| F | F→id | | | F→(E) | | |
| | _ | • | - | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | [i | d + id * | id \$ | | | |
| | [| d + id * | id \$ | | | |
| | [| d + id * | id \$ | | | |
| | _ | d + id * | id \$ | | | |



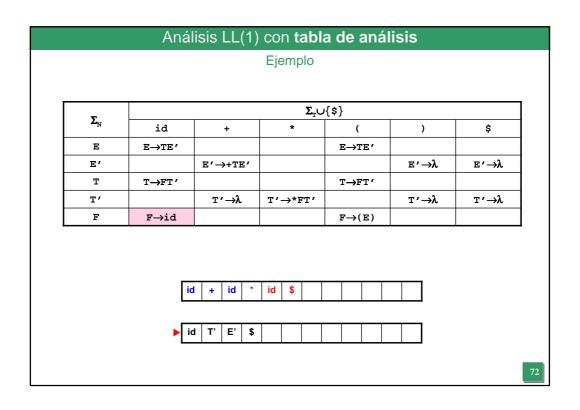
| | | | Ejemplo | | | |
|--------------------|-------|-----------------------|---|-------|---------------|------|
| | | | Ejempio | | | |
| | | | | | | |
| | | | $\Sigma_{\scriptscriptstyle m T}$ \cup | Jėl | | |
| $\Sigma_{_{\! N}}$ | id | 1 . | * | | , | |
| | | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | E ′→λ | E′→λ |
| Т | T→FT′ | | | T→FT′ | | |
| T' | | T′→λ | T'→*FT' | | T′→λ | T′→λ |
| F | F→id | | | F→(E) | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | i i | d + id * | id \$ | | $\overline{}$ | |
| | i | d + id * | id \$ | | | |
| | i i | d + id * | id \$ | | | |
| | | d + id * + T E' \$ | id \$ | | | |

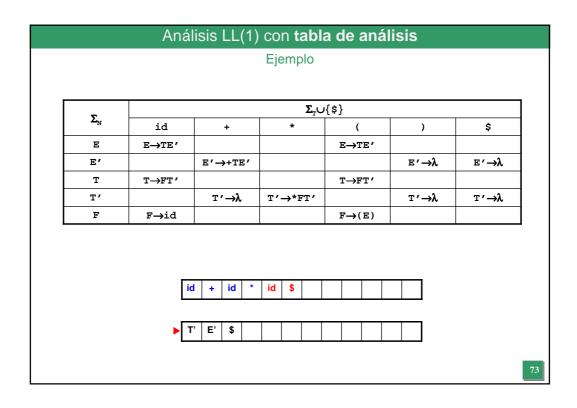


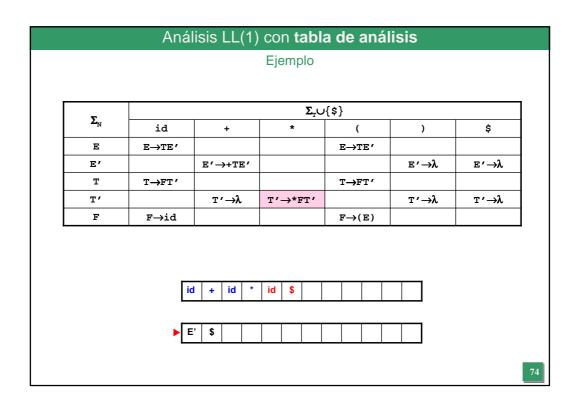
| | | | Ejemplo | | | |
|--------------------|-------|--------------|---------------------------|-------|--------------|--------------|
| | | | Ejempio | | | |
| | | | | | | |
| | | | | (A) | | |
| $\Sigma_{_{\! N}}$ | _ | ı | $\Sigma_{_{ m T}}$ \cup | | Γ | Ι . |
| N | id | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | E′→λ | Ε'→ λ |
| T | T→FT′ | | | T→FT′ | | |
| T' | | T′ →λ | T'→*FT' | | T ′→λ | T′→λ |
| F | F→id | | | F→(E) | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | _ | | | | | |
| | ic | 1 + id * | id \$ | | \Box | |
| | ic | 1 + id * | id \$ | | | |
| | | | id \$ | | | |
| | ic | | id \$ | | | |



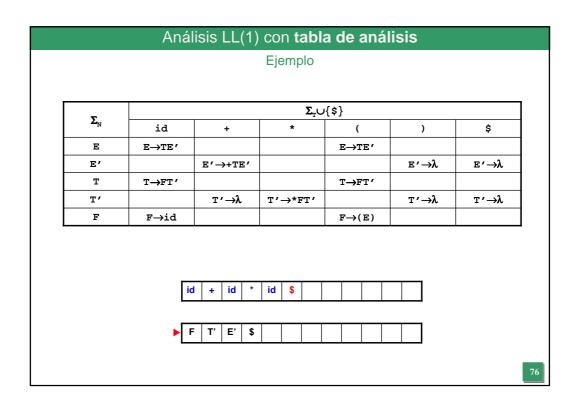
| | | | Ejemplo | | | |
|--------------------|-------|--------------|------------------------------------|-------|--------------|------|
| | | | Ejempio | | | |
| | | | | | | |
| | 1 | | ~ | .(A) | | |
| $\Sigma_{_{ m N}}$ | | 1 | $\Sigma_{\scriptscriptstyle m T}$ | | Γ | Ι . |
| N | id | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | E′→λ | E'→7 |
| T | T→FT′ | | | T→FT′ | | |
| T' | | T ′→λ | T'→*FT' | | T ′→λ | T′→λ |
| F | F→id | | | F→(E) | | |
| - | | | | | | |
| | | | | | | |
| | | | | | | |
| - | | | | | | |
| | | d + id * | id \$ | | | |
| | į | d + id * | id \$ | | | |
| | i | d + id * | id \$ | | | |
| | L | d + id * | id \$ | | | |



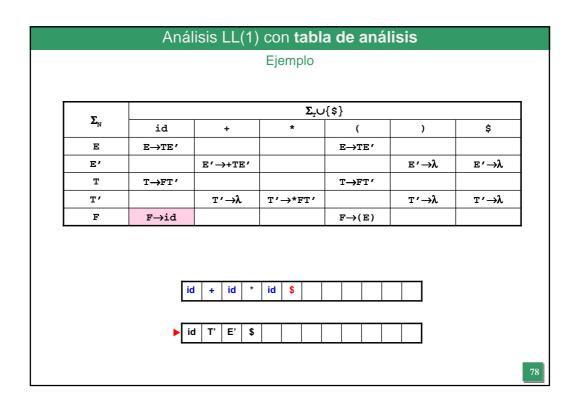


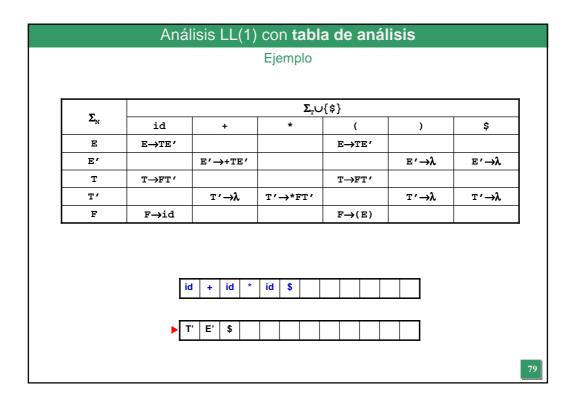


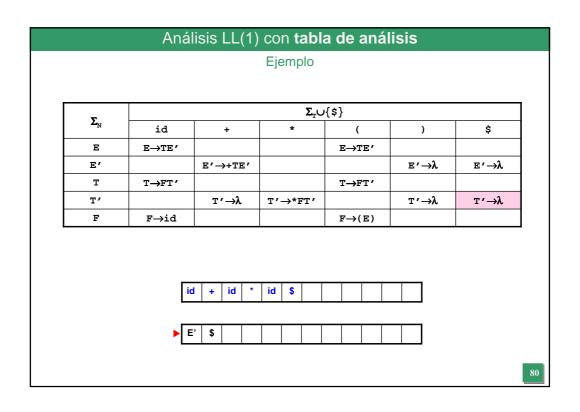
| | | | Ejemplo | | | |
|--------------------|----------------|--------------|----------------------|-------|---------------|--------------|
| | | | Ljempio | | | |
| | | | | | | |
| | | | 7 | lej. | | |
| $\Sigma_{_{\! N}}$ | <u> </u> | T | $\Sigma_{ m T} \cup$ | | | |
| | id | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | E ′→λ | E'→7 |
| T | T→FT′ | | | T→FT′ | | |
| T' | | T ′→λ | T'→*FT' | | T ′→λ | Τ′→ / |
| F | F→id | | | F→(E) | | |
| | • | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | T _i | d + id * | id \$ | | $\overline{}$ | |
| | ie | d + id * | id \$ | | | |
| | io | d + id * | id \$ | | | |
| | | d + id * | id \$ | | | |



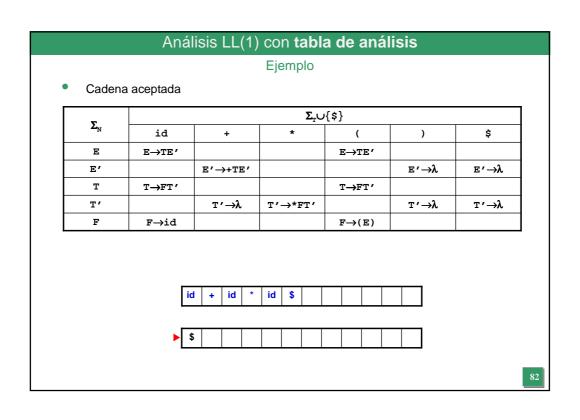
| | | | Ejemplo | | | |
|--------------------|-------|-----------------|------------------------------------|-------|--------------|------|
| | | | Ljempio | | | |
| | | | | | | |
| | | | $\Sigma_{\scriptscriptstyle m T}$ | le) | | |
| $\Sigma_{_{ m N}}$ | | Ι. | * | | , | |
| | id | + | * | (|) | \$ |
| E | E→TE′ | | | E→TE′ | | |
| E' | | E'→+TE' | | | E′→λ | E′→λ |
| T | T→FT′ | | | T→FT′ | | |
| T' | | T ′→λ | T'→*FT' | | Τ′ →λ | T′→λ |
| F | F→id | | | F→(E) | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | T. | 4 , i.d * | ia e I | | | |
| | i | d + id * | id \$ | | | |
| | i | d + id * | id \$ | | | |
| | [i | | id \$ | | | |



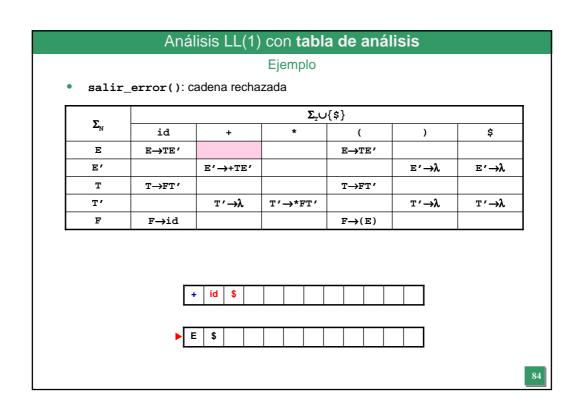




Análisis LL(1) con tabla de análisis Ejemplo $\Sigma_{\scriptscriptstyle m T} \cup \{\, \$\, \}$ $\boldsymbol{\Sigma}_{\!\scriptscriptstyle N}$ id) \$ E $E \rightarrow TE'$ $E \rightarrow TE'$ E'→+TE' $E' \rightarrow \lambda$ $E' \rightarrow \lambda$ Т T→FT′ T→FT' T' ${\tt T'} {\to} {\lambda}$ $T' \rightarrow *FT'$ ${\tt T'} {\to} {\lambda}$ T'→λ F ${\tt F}{\rightarrow}{\tt id}$ $F \rightarrow (E)$ + id id \$ 81



Análisis LL(1) con tabla de análisis Ejemplo $\Sigma_{\scriptscriptstyle m T} \cup \{\, \$\, \}$ $\Sigma_{\!_{N}}$ id) \$ (E $E \rightarrow TE'$ $E \rightarrow TE'$ $E' \rightarrow +TE'$ $E' \rightarrow \lambda$ $E' \rightarrow \lambda$ Т T→FT′ T→FT′ T' ${\tt T'} {\to} {\lambda}$ $T' \rightarrow *FT'$ ${\tt T'} {\to} {\lambda}$ T'→λ F ${\tt F}{\rightarrow}{\tt id}$ $F \rightarrow (E)$ id



Análisis sintáctico

Bibliografía

- [Alf] "Teoría de Autómatas y lenguajes formales" M. Alfonseca y otros
- [Hop] "Introducción a la teoría de autómatas, lenguajes y computación" Hopcroft, J.; Motwani, R.; Ullman, J.
- [Aho] "Compiladores. Principios, técnicas y herramientas" A. V. Aho; R. Sefthi; J. D. Ullman