

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

ACADEMIC HONESTY

Plagiarism and other forms of academic misconduct

- make sure you understand what is not allowed and why
- see, e.g., [Academic Honesty and Integrity at Chalmers](#)

Don'ts

- you are not allowed to cooperate on the written assignments
- you are not allowed to plagiarize or *paraphrase* other peoples work (including the slides for you presentation!)
- You are not allowed to copy otherwise make use of other peoples code.



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

WORD OF WARNING

Do no publish your code on the net!

- it's not prohibited, but ...

... if someone else copies your code you may be subject to disciplinary action, where you have to prove authorship of the code

- quite unnecessary stress even if you did nothing wrong!



I understand the will to share and the concept of freedom of

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

BONUS GRADE

If you pass all labs before the written exam (January 17 2019) your final grade on the course will be one higher than the grade on the written exam.

- given that you passed the exam.
- i.e., the additional grade will not pass you on the course if you failed the written exam



Note that the bonus grade only applies to the first exam attempt, and not re-exams.

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

LAST TIME

Lexical analysis

- string of characters to sequence of tokens

Lexical tokens, token types

- identifiers, keywords, operators, separators, numbers, ...

Specifying token types

- regular expressions

The logo for Cartagena99 features the text "Cartagena99" in a stylized, dark green font. The "99" is significantly larger and more prominent than the "Cartagena" part. The text is set against a light blue and orange gradient background that resembles a stylized wave or a banner.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

TODAY

Context-free grammars

- derivations, derivation trees

Introduction to parsing

- depth first search

Ambiguity

Rewriting grammars

- associativity
- precedence

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

AN EXAMPLE GRAMMAR

Non-terminals

- S, L, E (in capital letters)
- sometimes called syntactic categories

Terminals

- corresponding to the tokens
- *id*, *num* – token types (in italics)
- ; := + () , **print** – lexemes (in bold)

Start non-terminal

- S

Production rules (selected)

```
S → S ; S
S → id := E
S → print ( L )
E → id
E → num
E → E + E
E → ( S , E )
L → E
L → L , E
```

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

DERIVATIONS

A derivation

- starts in the start non-terminal (the start symbol)
- in this case S
- proceeds by replacing one non-terminal according to one of the possible production rules
- until no more non-terminals exist

(1) $S \rightarrow S; S$
(2) $S \rightarrow id := E$
(3) $S \rightarrow \mathbf{print} (L)$
(4) $E \rightarrow id$
(5) $E \rightarrow num$
(6) $E \rightarrow E + E$
(7) $E \rightarrow (S, E)$
(8) $L \rightarrow E$
(9) $L \rightarrow L, E$

$S \rightarrow S; S \rightarrow id := E \rightarrow \mathbf{print} (L)$


CLASES PARTICULARES, TUTORIAS TECNICAS ONLINE
LLAMA O ENVIA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Derive $x := 3; y := (z := 5, x + z)$

- (1) $S \rightarrow S; S$
- (2) $S \rightarrow id := E$
- (3) $S \rightarrow \mathbf{print} (L)$
- (4) $E \rightarrow id$
- (5) $E \rightarrow num$
- (6) $E \rightarrow E + E$
- (7) $E \rightarrow (S, E)$
- (8) $L \rightarrow E$
- (9) $L \rightarrow L, E$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

DERIVATIONS

When a production rule has more than one non-terminal on the right hand side

- $S \rightarrow S; S$

Derivation can continue on either

$$\underline{S} \Rightarrow_1 \underline{S}; S \Rightarrow_{2,5} x:=1; \underline{S} \Rightarrow_{2,5} x:=1; y:=2$$

$$\underline{S} \Rightarrow_1 S; \underline{S} \Rightarrow_{2,5} \underline{S}; y:=2 \Rightarrow_{2,5} x:=1; y:=2$$

Does it matter?

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, bold font. The 'C' is large and blue, while 'artagena99' is in a dark green color. The text is set against a light blue and orange gradient background.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

LEFT AND RIGHT DERIVATIONS

When the left-most non-terminal is always selected we have a left derivation

L: $\underline{S} \Rightarrow_1 \underline{S}; S \Rightarrow_{2,5} x:=1; \underline{S} \Rightarrow_{2,5} x:=1; y:=2$

When the right-most non-terminal is always selected we have a right derivation

R: $\underline{S} \Rightarrow_1 S; \underline{S} \Rightarrow_{2,5} \underline{S}; y:=2 \Rightarrow_{2,5} x:=1; y:=2$

There will always be a right and a left derivation

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

ORDER OF REWRITING

Some types differences in order of rewriting are unimportant

- yet give rise to different derivations
- L: $\underline{S} \Rightarrow_1 \underline{S}; S \Rightarrow_{2,5} x:=1; \underline{S} \Rightarrow_{2,5} x:=1; y:=2$
- R: $\underline{S} \Rightarrow_1 S; \underline{S} \Rightarrow_{2,5} \underline{S}; y:=2 \Rightarrow_{2,5} x:=1; y:=2$

Derivations are linear, and force a *total order* on rewriting steps

How can we relax this?

- *partial order* necessary, certain rewriting steps must occur before other
- other rewriting steps can occur in either order (or in parallel)

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, green, serif font. The '99' is significantly larger and more prominent than the 'Cartagena' part. The text is set against a light blue and white background with a subtle wave-like pattern.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

INDUCED ORDER

The production rules induce an order

For context-free grammars, grammars with

- a single non-terminal on the left hand side of production rules
- an arbitrary number of terminals and non-terminals on the right hand side of production rules

For production rule

$$S \rightarrow S; S$$

consider the derivation

$$S^1 \Rightarrow_1 S^2; S^2 \Rightarrow_1 S^2; S^3; S^3$$

where superscripts denote necessary order

we have that

- all non-terminals on the right hand side of a rule can be rewritten in any order (or in parallel)

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

DERIVATION TREES

For production rule

- $S \rightarrow S; S$

consider the derivation

- $S^1 \Rightarrow_1 S^2; S^2 \Rightarrow_1 S^2; S^3; S^3$

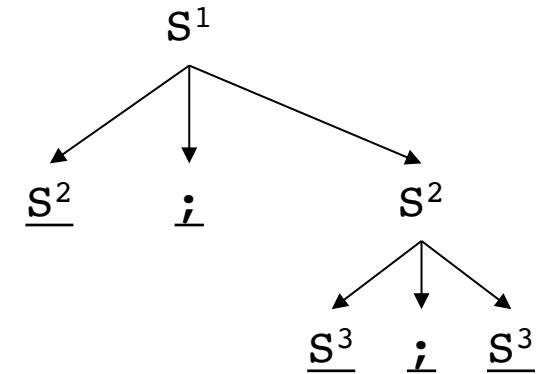
A more reasonable representation is a tree

- the level in the tree represent the necessary order

Rewriting a non-terminal with a production rule

- adds the right hand side of the rule as children to the non-terminal

As an added bonus it is clearer



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Create the derivation tree for $x:=1; y:=2$

- is it unique?

- (1) $S \rightarrow S; S$
- (2) $S \rightarrow id := E$
- (3) $S \rightarrow \mathbf{print} (L)$
- (4) $E \rightarrow id$
- (5) $E \rightarrow num$
- (6) $E \rightarrow E + E$
- (7) $E \rightarrow (S, E)$
- (8) $L \rightarrow E$
- (9) $L \rightarrow L, E$

Cartagena99

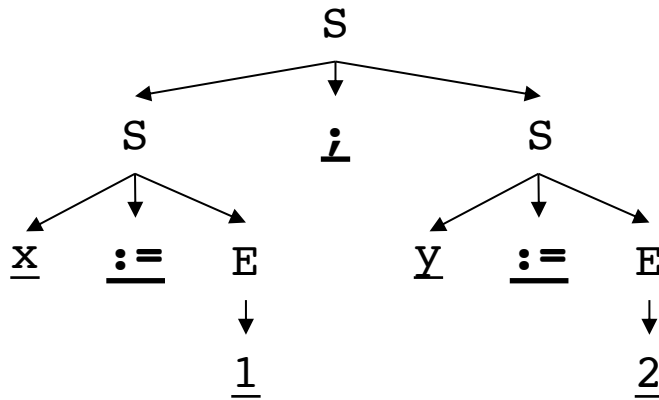
CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Create the derivation tree for $x:=1; y:=2$

- is it unique?
- Yes



- (1) $S \rightarrow S; S$
- (2) $S \rightarrow id := E$
- (3) $S \rightarrow \mathbf{print} (L)$
- (4) $E \rightarrow id$
- (5) $E \rightarrow num$
- (6) $E \rightarrow E + E$
- (7) $E \rightarrow (S, E)$
- (8) $L \rightarrow E$
- (9) $L \rightarrow L, E$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

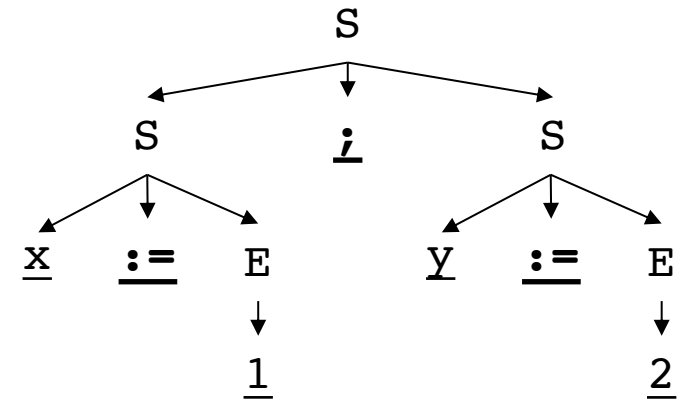
RELATION TO DERIVATIONS

Derivation tree for $x:=1; y:=2$

- result is given by left-to-right traversal of leaves

Left derivations correspond to left-to-right inorder traversal of tree

- $\underline{S} \Rightarrow_1 \underline{S}; S \Rightarrow_2 x:=\underline{E}; S \Rightarrow_5$
 $x:=1; \underline{S} \Rightarrow_2 x:=1; y:=\underline{E} \Rightarrow_5 x:=1; y:=2$



Right derivations correspond to right-to-left inorder traversal of tree

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Create the derivation tree for $x := (y := 1, y + 1)$

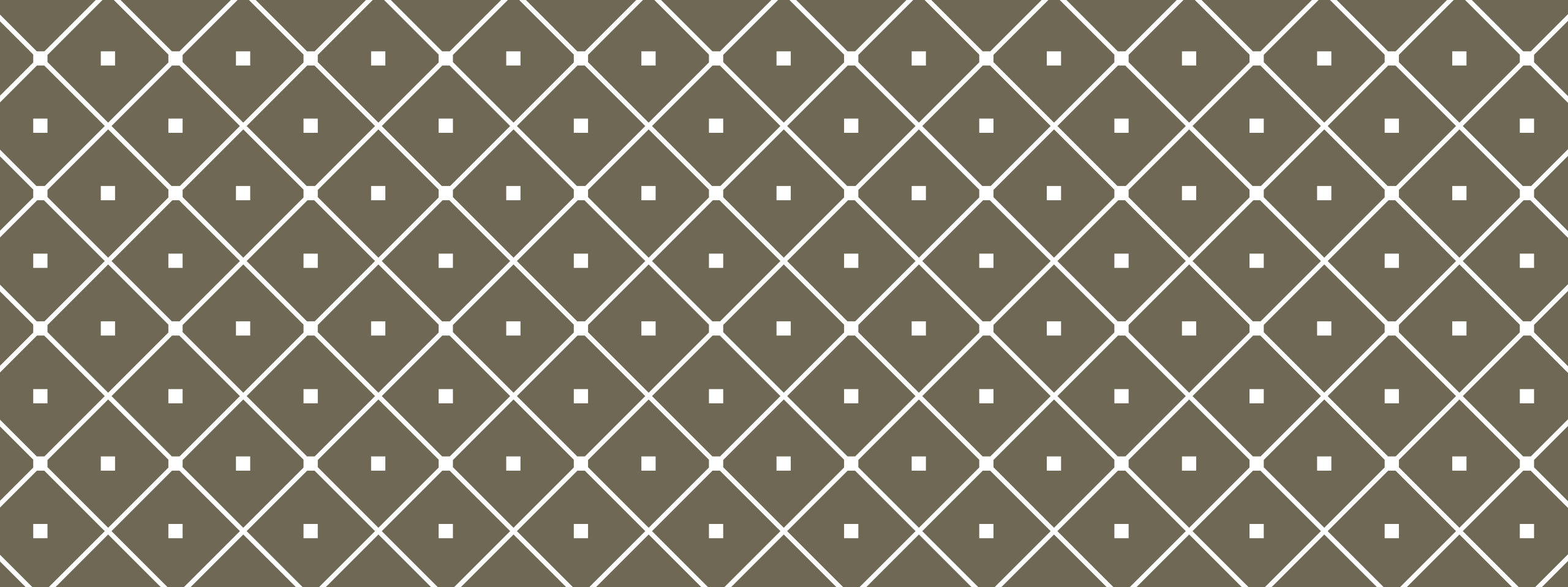
- How many derivations does it represent?

- (1) $S \rightarrow S; S$
- (2) $S \rightarrow id := E$
- (3) $S \rightarrow \mathbf{print} (L)$
- (4) $E \rightarrow id$
- (5) $E \rightarrow num$
- (6) $E \rightarrow E + E$
- (7) $E \rightarrow (S, E)$
- (8) $L \rightarrow E$
- (9) $L \rightarrow L, E$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

CREATION OF DERIVATION TREES

Given a program

- 1+2

and a grammar

$$E \rightarrow T \mid T + E$$
$$T \rightarrow \text{NUM} \mid (E)$$

How can we construct a derivation tree corresponding to the program in the

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

TWO APPROACHES

Top down

Start in S , find derivation to program

- simpler to understand, intuitively corresponds to the derivation process
- slow in general (n^3)
- fast subsets exists (LL(1) – Lecture 3)
- LL(1) easy to implement by hand

Bottom up

Start with program, rewrite backwards to S

- harder to understand
- slow in general (n^3)
- fast subsets exists (LR – Lecture 4)
- very cumbersome to implement by hand
- more used, subsets more expressive than LL(1)

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

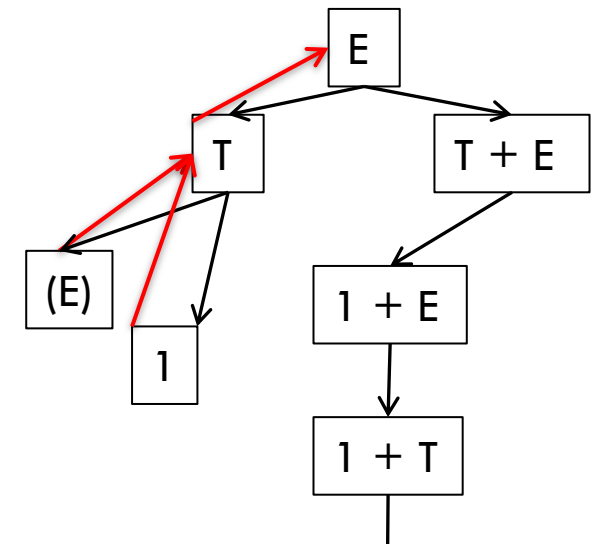
PARSING AS A SEARCH PROBLEM

Push E on stack

1. If stack is empty, fail
 - Let X be top of stack
2. if X is equal to input, done
3. if X is incompatible with input or if X does not contain any more non-terminals
 - backtrack (pop, and continue with 1)
4. Let A be left-most non-terminal in X
5. Select next unused rule P for A in X
 - if no more rules, backtrack
 - mark rule as used for A in X

Try program 1 + 2 in grammar

$$\begin{aligned} E &\rightarrow T \mid T + E \\ T &\rightarrow num \mid (E) \end{aligned}$$



Cartagena99

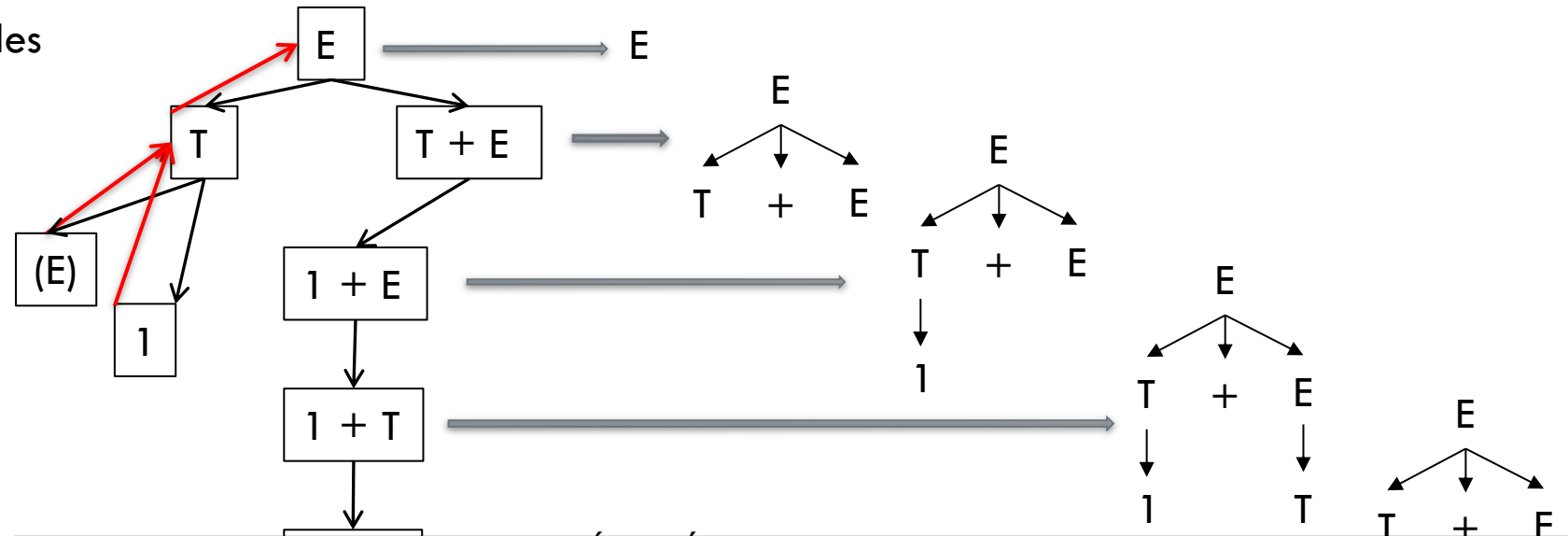
CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

WHAT IS THE DERIVATION TREE

The derivation tree lives on the stack

- updated by the selected rules



Cartagena99

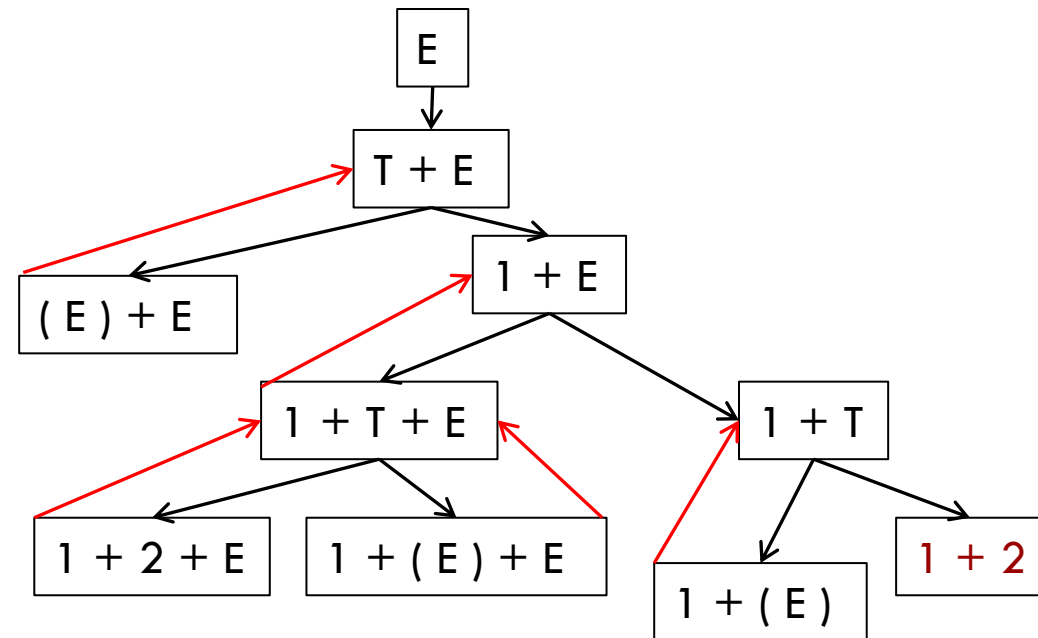
CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

ORDER MATTERS

Consider a small rearrangement of the grammar

$$\begin{array}{l} E \rightarrow T + E \quad | \quad T \\ T \rightarrow (E) \quad | \quad num \end{array}$$



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

IS IT REALLY THIS SIMPLE?

Sadly not!

What about *left recursive* grammars?

- $A \rightarrow Aa$

The left recursion does not have to be immediate, mutually left recursive is problematic too

- $A \rightarrow Bb$

- $B \rightarrow Cc$

- $C \rightarrow Aa$

The logo for Cartagena99 features the text "Cartagena99" in a stylized, teal-colored font. The "99" is significantly larger and more prominent than the "Cartagena" part. The text is set against a light blue and orange gradient background that resembles a stylized wave or a banner.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

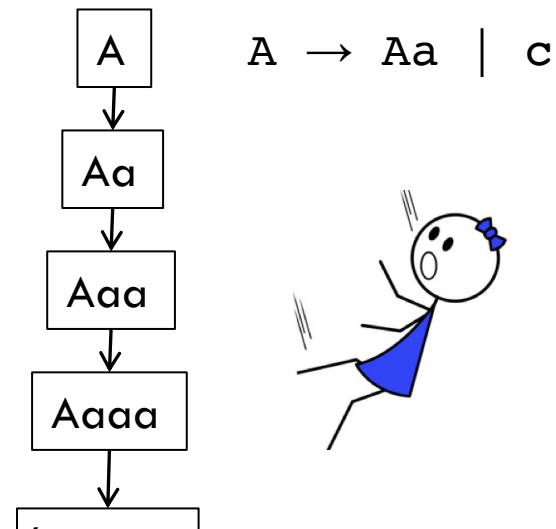
LEFT RECURSION

Naïve search may non-terminate on left-recursive grammars

- it's possible to enhance the algorithm with cycle detection
- not very efficient

Other solution – rewrite grammar to remove left recursion!

Cartagena99



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

COMPLEXITY

The algorithm presented can be seen as a variant of *depth first search*.

The worst case complexity is *exponential* in the length of the input.

- not that hard to construct grammars with this behavior (try!)

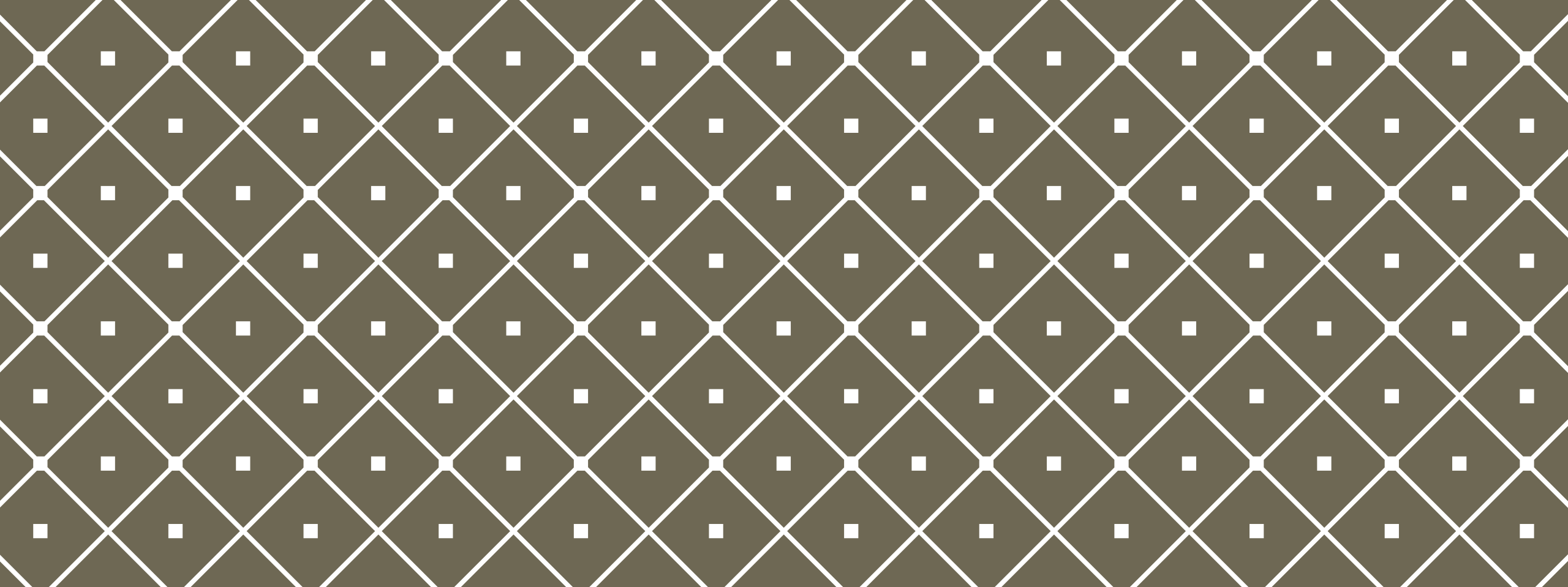
Not very practical

- in the current form

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Consider the grammar of expressions

- create a derivation tree for $10 * x + y$

$$\begin{array}{l} E \rightarrow E + E \\ | E * E \\ | E - E \\ | E / E \\ | (E) \\ | id \\ | num \end{array}$$

Cartagena99

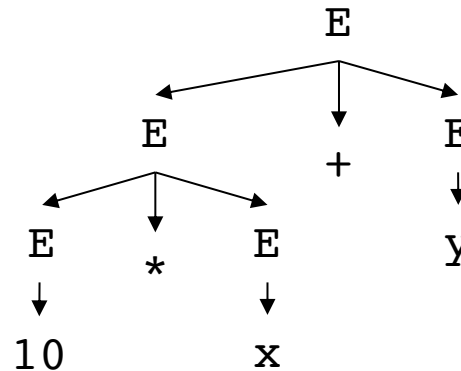
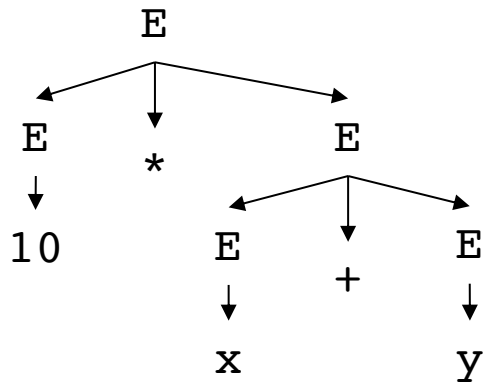
CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Consider the grammar of expressions

- create a derivation tree for $10 * x + y$


$$\begin{array}{l} E \rightarrow E + E \\ | \\ E * E \\ | \\ E - E \\ | \\ E / E \\ | \\ (E) \\ | \\ id \\ | \\ num \end{array}$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

AMBIGUITY

One program, two derivation trees

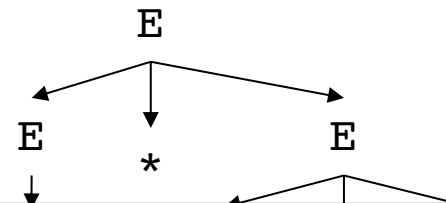
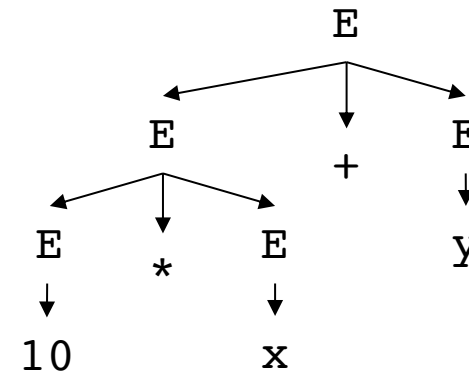
- Does it matter? Yes

Original program

- $10 * x + y$

The trees correspond to

- $(10 * x) + y$



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

AMBIGUITY

A grammar is ambiguous if there exists *at least one* program

- with *at least* two different derivation trees

Problems with ambiguity

- parsing should create abstract syntax tree
- more than one derivation tree means more than one abstract syntax tree
- parsing should be deterministic, i.e., same result for same input
- parser must chose one of the possible trees
- *the choice may change the semantics of the program*

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Is the grammar of straight line programs ambiguous?

$$\begin{aligned} S &\rightarrow S; S \\ S &\rightarrow id := E \\ S &\rightarrow \mathbf{print} (L) \\ E &\rightarrow id \\ E &\rightarrow num \\ E &\rightarrow E + E \\ E &\rightarrow (S, E) \\ L &\rightarrow E \end{aligned}$$

Cartagena99

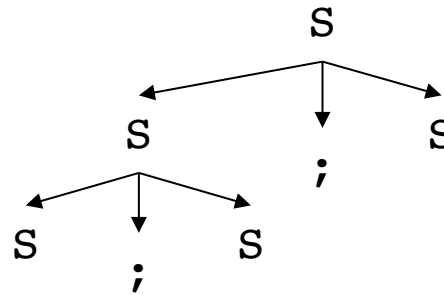
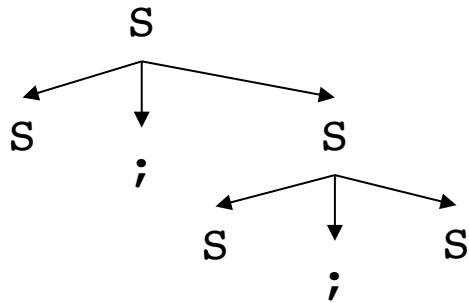
CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Is the grammar of straight line programs ambiguous?

- Indeed, consider $S; S; S$



$S \rightarrow S; S$
 $S \rightarrow id := E$
 $S \rightarrow \mathbf{print} (L)$
 $E \rightarrow id$
 $E \rightarrow num$
 $E \rightarrow E + E$
 $E \rightarrow (S, E)$
 $L \rightarrow E$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

HANDLING AMBIGUITY

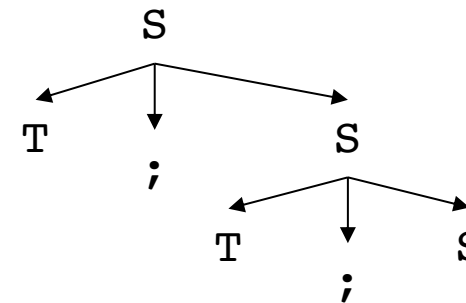
Even though ambiguity doesn't always matter we want to avoid it

We do this by rewriting the grammar to a grammar that is not ambiguous

For the straight line programs we can do this by factoring out the primitive statements

$$S \rightarrow T; S \mid T$$

This forces the derivation trees to a the following form



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

REWRITING EXPRESSIONS

For expressions we must rewrite the grammar to take

- precedence, and
- associativity into account

The first applied rule

- root, lowest precedence

The last applied rule

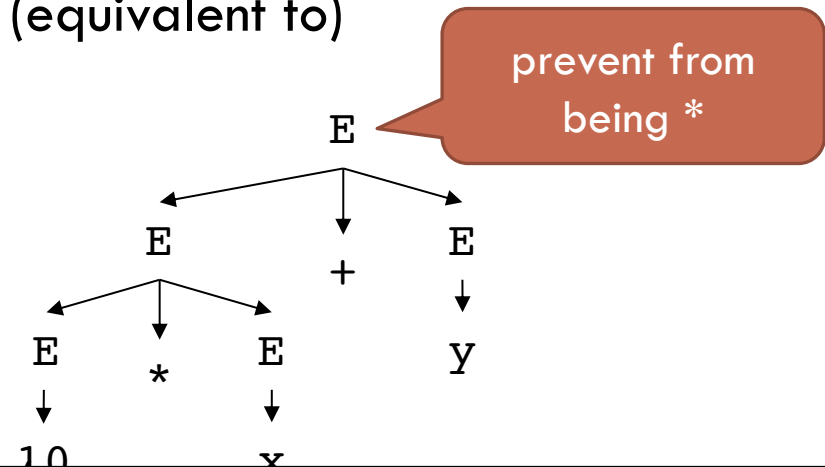
- leaf, highest precedence

How can we encode this in the grammar?

The only possible derivation tree for

$$10 * x + y$$

should be (equivalent to)



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

ASSOCIATIVITY AND PRECEDENCE

Precedence of operators (lowest first)

- + -
- * /

$$a * b + c = (a * b) + c$$

$$a * b - c = (a * b) - c$$

What about associativity?

- + * are (both left and right) associative
- $(a + b) + c = a + (b + c)$
- - / are left associative
- $a - b - c = (a - b) - c$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

REWRITING EXPRESSIONS

The first applied rule

- root, lowest precedence

The last applied rule

- leaf, highest precedence

$$\begin{array}{l} E \rightarrow E + E \\ | \\ E * E \\ | \\ E - E \\ | \\ E / E \\ | \\ \dots \end{array}$$

First applied rule

- earliest in grammar

Last applied rule

- last in grammar

$$\begin{array}{l} E \rightarrow E + E \mid E - E \\ | \\ F \end{array}$$
$$F \rightarrow F * F \mid F / F$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

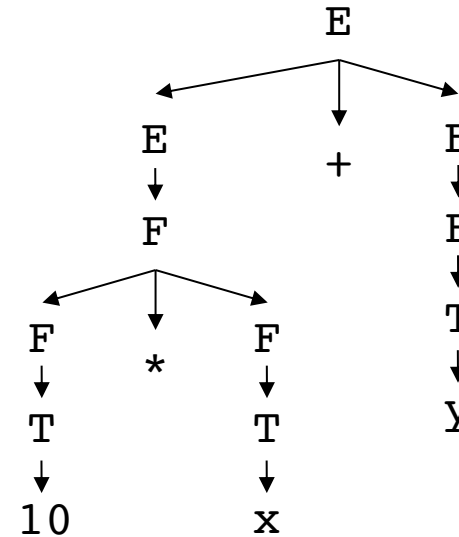
DID IT WORK?

Again, try $10 * x + y$

$E \rightarrow E + E$
| $E - E$
| F

$F \rightarrow F * F$
| F / F
| T

$T \rightarrow (E) \mid id \mid num$



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP: 689 45 44 70

EXERCISE

Is the result unambiguous?

$$\begin{array}{l} E \rightarrow E + E \\ | \quad E - E \\ | \quad F \end{array}$$
$$\begin{array}{l} F \rightarrow F * F \\ | \quad F / F \\ | \quad T \end{array}$$
$$T \rightarrow (E) \mid id \mid num$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Is the result unambiguous?

$$\begin{array}{l} E \rightarrow E + E \\ | \quad E - E \\ | \quad F \end{array}$$
$$\begin{array}{l} F \rightarrow F * F \\ | \quad F / F \\ | \quad T \end{array}$$
$$T \rightarrow (E) \mid id \mid num$$

No, try, e.g., $1 + 2 + 3$

- can be derived left associatively $(1 + 2) + 3$
- can be derived right associatively $1 + (2 + 3)$
- ok, from an operator perspective
- but ambiguous

Same for $*$ and $/$

- but there only the left associative derivation should be possible

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

ASSOCIATIVITY

Left associative

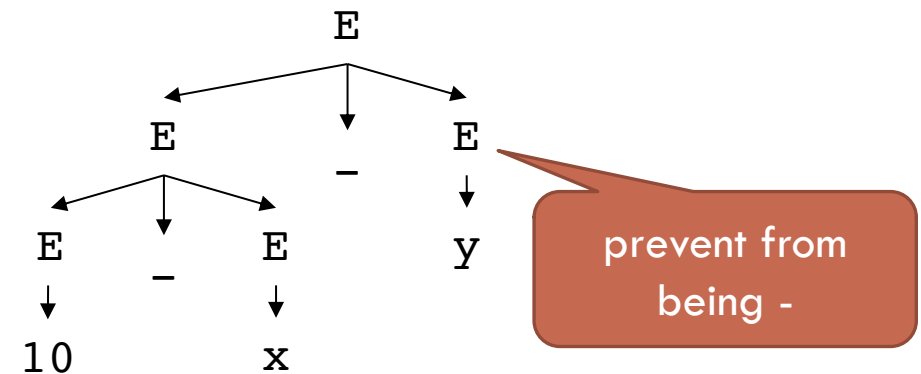
- left recursive trees
- consider $10 - x - y$

Left recursive trees

- left recursive grammar

What about + and *

- does it matter?



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

REWRITING EXPRESSIONS

Precedence

- + -
- * /

$$E \rightarrow E + F \mid E - F \mid F$$
$$F \rightarrow F * T \mid F / T \mid T$$
$$T \rightarrow (E) \mid id \mid num$$

Associativity

- + * left associative (by choice)
- / - left associative (by necessity)

Cartagena99

CLASES PARTICULARES, TUTORIAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

LEFT RECURSION

... but left associative operators give left recursive grammars ...

$$E \rightarrow E + F \mid E - F \mid F$$
$$F \rightarrow F * T \mid F / T \mid T$$
$$T \rightarrow (E) \mid id \mid num$$

Most unfortunate for predictive recursive descent parser

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70



LEFT FACTORING

Consider the following simplified expression grammar

$$E \rightarrow E + F \mid F$$

Possible derivations are

- F
- F + F
- F + F + F
- ...

Is there a way to express the same language

- that is not left recursive?

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

LEFT FACTORING

Consider the following simplified expression grammar

$$E \rightarrow E + F \mid F$$

Possible derivations are

- F
- F + F
- F + F + F
- ...

Is there a way to express the same language

- that is not left recursive?

Yes!

$$E \rightarrow F T$$
$$T \rightarrow + F T \mid \lambda$$

Same derivations, but

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

Cartagena99

LEFT FACTORING

Immediate left recursion, general case

$$A \rightarrow A \alpha \mid \beta$$

Rewrite as follows

$$\begin{aligned} A &::= \beta T \\ T &::= \alpha T \mid \lambda \end{aligned}$$

Compare

$$E \rightarrow E + F \mid F$$

and

$$\begin{aligned} E &\rightarrow F T \\ T &\rightarrow + F T \mid \lambda \end{aligned}$$

α and β are meta variables

- that match against the grammatical rules

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Immediate left recursion, general case

$$A \rightarrow A \alpha \mid \beta$$

Rewrite as follows

$$\begin{aligned} A &::= \beta T \\ T &::= \alpha T \mid \lambda \end{aligned}$$

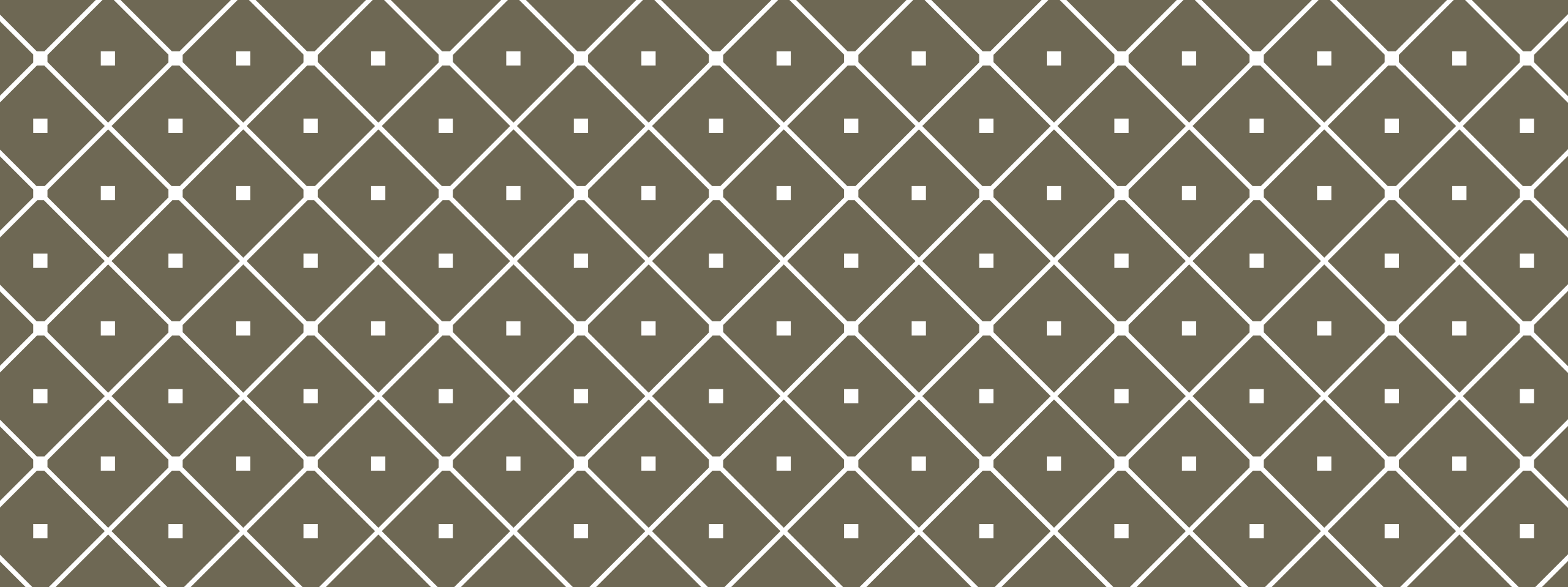
Left factorize the grammar for comma separated lists of numbers

$$L \rightarrow L, \textit{num} \mid \lambda$$

The logo for Cartagena99, featuring the text "Cartagena99" in a stylized font with a blue and orange gradient background.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

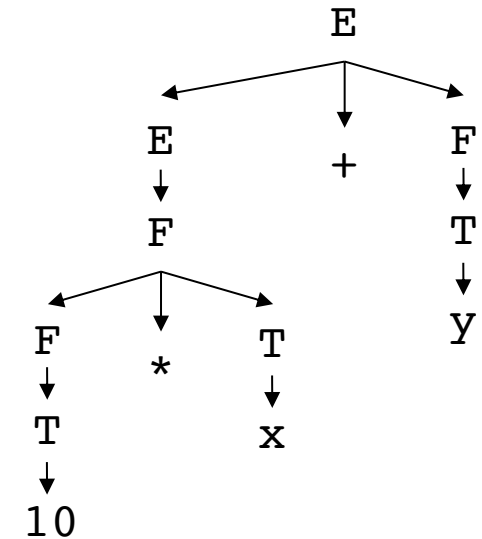
ABSTRACT SYNTAX

Consider our rewritten expression grammar

$$\begin{aligned} E &\rightarrow E + F \mid E - F \mid F \\ F &\rightarrow F * T \mid F / T \mid T \\ T &\rightarrow (E) \mid id \mid num \end{aligned}$$

The derivation trees contain a lot of unnecessary information

- separators (punctuation) only present to disambiguate
- ()



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

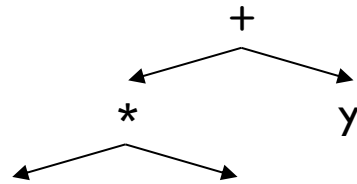
ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

ABSTRACT SYNTAX

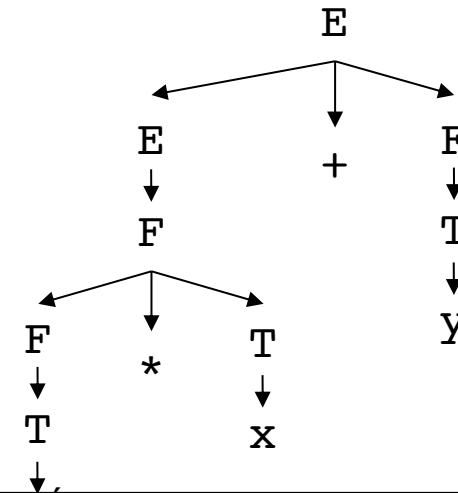
Later compiler stages are only interested in information that is important for the semantics of the program

For $10 * x + y$

Abstract syntax tree



Derivation tree/concrete syntax tree



Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

DESIGNING ABSTRACT SYNTAX

Keep only the essentials

- ambiguities not an issue – will be generated from parsing unambiguous grammars
- remove all syntax that only guides the parsing process

For the expressions, start in the grammar that was before the encoding of associativity and precedence

$$E \rightarrow E + E \mid E * E \mid E - E \mid E / E \mid (E) \mid id \mid num$$

Remove unnecessary decorations

$$E \rightarrow E + E \mid E * E \mid E - E \mid E / E \mid id \mid num$$
The logo for Cartagena99, featuring the text 'Cartagena99' in a stylized font with a blue and orange gradient background.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

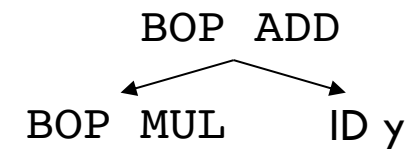
DESIGNING ABSTRACT SYNTAX

Give proper names (easier to map to code)

$$E \rightarrow \text{ADD } E \ E \mid \text{MUL } E \ E \mid \text{SUB } E \ E \mid \text{DIV } E \ E \mid \text{ID } id \\ \mid \text{NUM } num$$

Each syntactic category corresponds to one superclass and each production corresponds to one class

Contract operators to reduce number of needed classes

$$E \rightarrow \text{BOP } OP \ E \ E \mid \text{ID } id \mid \text{NUM } num$$


Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

MAPPING TO CODE

Chose longer names!

- only to illustrate mapping

$E \rightarrow BOP \ OP \ E \ E$

| ID *id*

| NUM *num*

$OP \rightarrow ADD \ | \ SUB \ | \ MUL \ | \ DIV$

```
public class E {}
```

```
public class BOP : E {  
    public OP op;  
    public E left, right;  
    public enum OP { ADD, SUB, MUL, DIV }  
}
```

```
public class ID : E {  
    public string id;  
}
```

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Create abstract syntax for

$$S \rightarrow S; S$$
$$S \rightarrow id := E$$
$$S \rightarrow \mathbf{print} (L)$$
$$E \rightarrow id$$
$$E \rightarrow num$$
$$E \rightarrow E + E$$
$$E \rightarrow (S, E)$$
$$L \rightarrow E$$
$$L \rightarrow L, E$$
The logo for Cartagena99 features the text "Cartagena99" in a stylized, green, serif font. The "99" is significantly larger and more prominent than the "Cartagena" part. The text is set against a light blue and orange gradient background that resembles a stylized wave or a banner.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Create abstract syntax for

```
S → S; S
S → id := E
S → print ( L )
E → id
E → num
E → E + E
E → ( S, E )
L → E
L → L, E
```

Remove all extra decoration

```
S → S; S
S → id := E
S → print L
E → id
E → num
E → E + E
E → S, E
L → E
L → L, E
```

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

EXERCISE

Replace list encodings with actual lists

$P \rightarrow [S]$
 $S \rightarrow id := E$
 $S \rightarrow \mathbf{print} [L]$
 $E \rightarrow id$
 $E \rightarrow num$
 $E \rightarrow E + E$
 $E \rightarrow [S], E$

Notice new syntactic category P for programs

Introduce proper names

$P \rightarrow \mathbf{Prog} [S]$
 $S \rightarrow \mathbf{Asn} id E$
 $S \rightarrow \mathbf{Print} [L]$
 $E \rightarrow \mathbf{Var} id$
 $E \rightarrow \mathbf{Num} num$
 $E \rightarrow \mathbf{Edd} E E$
 $E \rightarrow \mathbf{Let} [S] E$

Only contains the essentials, and translates

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

NEXT TIME

Predictive recursive descent!

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, dark green font. The '99' is significantly larger and more prominent than the 'Cartagena' part. The text is set against a light blue and orange gradient background that resembles a stylized wave or a banner.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70