Derivations

Derivation

The construction of a parse tree can be made precise by taking a derivational view, in which production are considered as rewriting rules.

A sentence belongs to a language if there is a derivation from the initial symbol to the sentence.
e.g. $E \to E + E | E * E | - E | (E) | \mathbf{id}$

Kind of derivations

Each sentence can be generated according to two different strategies leftmost and rightmost. Parsers generally return one of this two derivations.

Derivations

Derivation

The construction of a parse tree can be made precise by taking a derivational view, in which production are considered as rewriting rules.

A sentence belongs to a language if there is a derivation from the initial symbol to the sentence.
e.g. $E \rightarrow E + E | E * E | - E | (E) | \mathbf{id}$

Kind of derivations

Each sentence can be generated according to two different strategies leftmost and rightmost. Parsers generally return one of this two derivations.

Ambiguity

A grammar that produces more than one parse tree for some sentence is said to be ambiguos. An ambiguous grammar has more then one left-most derivation or more than one rightmost derivation for the same sentence.

Ambiguity and Precedence of Operators

Using the simplest grammar for expressions let's derive again the parse tree for:

$$\mathbf{id} + \mathbf{id} * \mathbf{id}$$

Now consider the following grammar:

$E \rightarrow E + T|E - T|T$

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

$F \rightarrow (E)|\mathbf{id}$

Use of ambiguos grammar

In some case it can be convenient to use ambiguous grammar, but then it is necessary to define precise disambiguating rules

Ambiguity

A grammar that produces more than one parse tree for some sentence is said to be ambiguos. An ambiguous grammar has more then one left-most derivation or more than one rightmost derivation for the same sentence.

Ambiguity and Precedence of Operators

Using the simplest grammar for expressions let's derive again the parse tree for:

$$\mathbf{id} + \mathbf{id} * \mathbf{id}$$

Now consider the following grammar:

$E \to E + T | E - T | T$

$T \to T * F | T / F | F$

$F \to (E) | \mathbf{id}$

Use of ambiguos grammar

In some case it can be convenient to use ambiguous grammar, but then it is necessary to define precise disambiguating rules

Ambiguity

A grammar that produces more than one parse tree for some sentence is said to be ambiguos. An ambiguous grammar has more then one left-most derivation or more than one rightmost derivation for the same sentence.

Ambiguity and Precedence of Operators

Using the simplest grammar for expressions let's derive again the parse tree for:

$$\mathbf{id} + \mathbf{id} * \mathbf{id}$$

Now consider the following grammar:
$E \rightarrow E + T | E - T | T$
$T \rightarrow T * F | T / F | F$
$F \rightarrow (E) | \mathbf{id}$

Use of ambiguos grammar

In some case it can be convenient to use ambiguous grammar, but then it is necessary to define precise disambiguating rules

Ambiguity

Conditional statements

Consider the following grammar:

$$
\begin{array}{lcl}
\mathit{stmt} & \rightarrow & \textbf{if } \mathit{expr} \textbf{ then } \mathit{stmt} \\
 & | & \textbf{if } \mathit{expr} \textbf{ then } \mathit{stmt} \textbf{ else } \mathit{stmt} \\
 & | & \textbf{other}
\end{array}
$$

decide if the following sentence belongs to the generated language:

$$\textbf{if } E_1 \textbf{ then if } E_2 \textbf{ then } S_1 \textbf{ else } S_2$$

Exercises

Consider the grammar:

$$S \rightarrow SS + | SS * | a$$

and the string $aa + a*$

- Give the leftmost derivation for the string
- Give the rightmost derivation for the string
- Give a parse tree for the string
- Is the grammar ambiguous or unambiguous?
- Describe the language generated by this grammar?

Define grammars for the following languages:

- $\mathscr{L} = \{w \in \{0,1\}^* | w \text{ is palindrom}\}$
- $\mathscr{L} = \{w \in \{0,1\}^* | w \text{ contains the same occurrences of 0 and 1}\}$

Exercises

Consider the grammar:

$$S \rightarrow SS + | SS * | a$$

and the string $aa + a*$

- Give the leftmost derivation for the string
- Give the rightmost derivation for the string
- Give a parse tree for the string
- Is the grammar ambiguous or unambiguous?
- Describe the language generated by this grammar?

Define grammars for the following languages:

- $\mathscr{L} = \{w \in \{0, 1\}^* | w \text{ is palindrom}\}$
- $\mathscr{L} = \{w \in \{0, 1\}^* | w \text{ contains the same occurrences of 0 and 1}\}$

CF grammars are capable to describe the syntax of most, but not all, the programming languages. For instance, the requirement that identifiers must be declared before their usage cannot be expressed in CF grammars.

So what we can do?

Ambiguity

- Many languages admit both ambiguous and unambiguous grammars, while some languages admit only ambiguous grammars.
- A language that only admits ambiguous grammars is called an [illegible], e.g. $\{a^n b^m c^k \mid n = m \text{ or } m = k,\ n, m, k > 0\}$
- A Turing machine cannot decide whether a context-free language is inherently ambiguous or not.

CF grammars are capable to describe the syntax of most, but not all, the programming languages. For instance, the requirement that identifiers must be declared before their usage cannot be expressed in CF grammars.

So what we can do?

Ambiguity

- Many languages admit both ambiguous and unambiguous grammars, while some languages admit only ambiguous grammars
- A language that only admits ambiguous grammars is called an inherently ambiguous language, e.g. $\{a^n b^m c^k \mid n = m \text{ or } m = k; n, m, k \geq 0\}$
- A Turing machine cannot decide whether a context-free language is inherently ambiguous or not

CF grammars are capable to describe the syntax of most, but not all, the programming languages. For instance, the requirement that identifiers must be declared before their usage cannot be expressed in CF grammars.

So what we can do?

Ambiguity

- Many languages admit both ambiguous and unambiguous grammars, while some languages admit only ambiguous grammars
- A language that only admits ambiguous grammars is called an inherently ambiguous language, e.g. $\{a^n b^m c^k \mid n = m \text{ or } m = k; n, m, k \geq 0\}$
- A Turing machine cannot decide whether a context-free language is inherently ambiguous or not

CF grammars are capable to describe the syntax of most, but not all, the programming languages. For instance, the requirement that identifiers must be declared before their usage cannot be expressed in CF grammars.

So what we can do?

Ambiguity

- Many languages admit both ambiguous and unambiguous grammars, while some languages admit only ambiguous grammars
- A language that only admits ambiguous grammars is called an inherently ambiguous language, e.g. $\{a^n b^m c^k \mid n = m \text{ or } m = k; n, m, k \geq 0\}$
- A Turing machine cannot decide whether a context-free language is inherently ambiguous or not

CF grammars are capable to describe the syntax of most, but not all, the programming languages. For instance, the requirement that identifiers must be declared before their usage cannot be expressed in CF grammars.

So what we can do?

Ambiguity

- Many languages admit both ambiguous and unambiguous grammars, while some languages admit only ambiguous grammars
- A language that only admits ambiguous grammars is called an inherently ambiguous language, e.g. $\{a^n b^m c^k \mid n = m \text{ or } m = k; n, m, k \geq 0\}$
- A Turing machine cannot decide whether a context-free language is inherently ambiguous or not