Introduction to Context-Free Grammars

Ian Ludden

Learning Objectives

By the end of this lesson, you will be able to:

Learning Objectives

By the end of this lesson, you will be able to:

- Correctly interpret the definition of a context-free grammar: variables, rules, start symbols, terminal symbols.

Learning Objectives

By the end of this lesson, you will be able to:

- Correctly interpret the definition of a context-free grammar: variables, rules, start symbols, terminal symbols.
- Correctly interpret a grammar rule, e.g., what does ε or a vertical bar mean on the right hand side of a rule.

Motivation, i.e., why you should care

Motivation, i.e., why you should care

- Applications to natural language processing (NLP), a field with lots of recent applications of deep learning (Link to article)

Motivation, i.e., why you should care

- Applications to natural language processing (NLP), a field with lots of recent applications of deep learning (Link to article)
- Context-free grammars will be covered in more depth in CS 374

Motivation, i.e., why you should care

- Applications to natural language processing (NLP), a field with lots of recent applications of deep learning (Link to article)
- Context-free grammars will be covered in more depth in CS 374
- Good practice for inductive proofs on trees

Formal Definition

Formal Definition

Definition

A ***context-free grammar*** $G = (\Sigma, \Gamma, R, S)$ is a structure defined by:

- A finite set Σ of *symbols*, or ***terminals***

Formal Definition

Definition

A ***context-free grammar*** $G = (\Sigma, \Gamma, R, S)$ is a structure defined by:

- A finite set Σ of *symbols*, or ***terminals***
- A finite set Γ of ***non-terminals*** (disjoint from Σ)

Formal Definition

Definition

A ***context-free grammar*** $G = (\Sigma, \Gamma, R, S)$ is a structure defined by:

- A finite set Σ of *symbols*, or ***terminals***
- A finite set Γ of ***non-terminals*** (disjoint from Σ)
- A finite set R of ***production rules***

Formal Definition

Definition

A ***context-free grammar*** $G = (\Sigma, \Gamma, R, S)$ is a structure defined by:

- A finite set Σ of *symbols*, or ***terminals***
- A finite set Γ of ***non-terminals*** (disjoint from Σ)
- A finite set R of ***production rules***
- A finite set S of ***start symbols*** (often just one)

Formal Definition

Definition

A ***context-free grammar*** $G = (\Sigma, \Gamma, R, S)$ is a structure defined by:

- A finite set Σ of *symbols*, or ***terminals***
- A finite set Γ of ***non-terminals*** (disjoint from Σ)
- A finite set R of ***production rules***
- A finite set S of ***start symbols*** (often just one)

$$
\begin{aligned}
\langle\text{sentence}\rangle &\to \langle\text{noun phrase}\rangle\langle\text{verb phrase}\rangle\langle\text{noun phrase}\rangle \\
\langle\text{noun phrase}\rangle &\to \langle\text{adjective phrase}\rangle\langle\text{noun}\rangle \\
\langle\text{adj. phrase}\rangle &\to \langle\text{article}\rangle \mid \langle\text{possessive}\rangle \mid \langle\text{adjective phrase}\rangle\langle\text{adjective}\rangle \\
\langle\text{verb phrase}\rangle &\to \langle\text{verb}\rangle \mid \langle\text{adverb}\rangle\langle\text{verb phrase}\rangle
\end{aligned}
$$

$$
\begin{aligned}
\langle\text{noun}\rangle &\to \text{dog} \mid \text{trousers} \mid \text{daughter} \mid \text{nose} \mid \text{homework} \mid \text{time lord} \mid \text{pony} \mid \cdots \\
\langle\text{article}\rangle &\to \text{the} \mid \text{a} \mid \text{some} \mid \text{every} \mid \text{that} \mid \cdots \\
\langle\text{possessive}\rangle &\to \langle\text{noun phrase}\rangle\text{'s} \mid \text{my} \mid \text{your} \mid \text{his} \mid \text{her} \mid \cdots \\
\langle\text{adjective}\rangle &\to \text{friendly} \mid \text{furious} \mid \text{moist} \mid \text{green} \mid \text{severed} \mid \text{timey-wimey} \mid \text{little} \mid \cdots \\
\langle\text{verb}\rangle &\to \text{ate} \mid \text{found} \mid \text{wrote} \mid \text{killed} \mid \text{mangled} \mid \text{saved} \mid \text{invented} \mid \text{broke} \mid \cdots \\
\langle\text{adverb}\rangle &\to \text{squarely} \mid \text{incompetently} \mid \text{barely} \mid \text{sort of} \mid \text{awkwardly} \mid \text{totally} \mid \cdots
\end{aligned}
$$

More Examples

Example 1: Binary Strings

Start symbol is S, terminals are 0, 1, and ε, rules are:

$$S \to 0S$$
$$S \to 1S$$
$$S \to \varepsilon$$

More Examples

Example 1: Binary Strings

Start symbol is S, terminals are 0, 1, and ε, rules are:

$$S \to 0S$$
$$S \to 1S$$
$$S \to \varepsilon$$

Example 2: Simple Arithmetic Expressions

Start symbols are E and V (also non-terminals), terminals are x, y, $+$, and $\times$, rules are:

$$E \to E + V \mid E \times V \mid V + V \mid V \times V$$
$$V \to x \mid y$$

Recap: Learning Objectives

By the end of this lesson, you will be able to:

- Correctly interpret the definition of a context-free grammar: variables, rules, start symbols, terminal symbols.
- Correctly interpret a grammar rule, e.g., what does ε or a vertical bar mean on the right hand side of a rule.