Context-free languages

Arno Pauly

February 23, 2021

The hierarchy, overview

From simplest to most complicated:

3 Regular languages (having right-linear grammars)

2 Context-free languages (having context-free grammars)

1 Context-sensitive languages (having context-sensitive grammars)

0 Computably enumerable languages (having (unrestricted) grammars)

-1 Arbitrary languages (not necessarily describable by a grammar at all)

Context-free grammars

Definition

A grammar is *context-free*, if the left-hand side of every rule is a single non-terminal.

Potential conventions:

- "Every non-terminal appearing on the right of a rule also appears on the left of a rule."
- "The right hand of a rule has at most 2 symbols."

A context-free language which is not regular

Proposition

The language $\{a^n b^n \mid n \in \mathbb{N}\}$ *is context-free, but not regular.*

Proof.

That it is not regular, we have shown using the pumping lemma.
It is context free because it is generated by the grammar
$S \rightarrow \varepsilon, S \rightarrow aSb$. $\square$

A “real” grammar

Example

Let $\Sigma = \{\text{the}, \text{dog}, \text{cat}, \text{eats}, \text{sleeps}\}$,

$$\mathcal{N} = \{S, \text{NOUN}, \text{NP}, \text{VP}, \text{TRANS-VERB}, \text{INTRANS-VERB}\}$$

and the rules be $S \to$ NP VP, NP $\to$ the NOUN, NOUN $\to$ cat, NOUN $\to$ dog, VP $\to$ INTRANS-VERB, VP $\to$ TRANS-VERB NP, TRANS-VERB $\to$ eats, INTRANS-VERB $\to$ sleeps, INTRANS-VERB $\to$ eats.

Parse trees

Definition

A parse tree for a context-free grammar is a tree labelled by terminals, non-terminals and ε subject to the following rules:

1. The root is labeled S.
2. A vertex is a leaf iff it is labeled by a terminal or ε.
3. For every non-leaf v there is a rule with the label of v on the left, and the concatenation of the child labels on the right.

Theorem

The language generated by a context-free grammar consists exactly of the words built from the terminals on the leaves of finite parse trees, in order.

Parse trees

To show that a word belongs to the language of a context-free grammar, try to construct a parse tree bottom-up.

Closure properties

Theorem

Context free languages are closed under

1. *Union*
2. *Composition*
3. *Kleene-star*
4. *Intersection with regular languages*

Non-closure properties

Theorem

Context free languages are not closed under

1. *Intersection (consider* $\{a^n b^n c^m \mid n, m \in \mathbb{N}\}$ *and* $\{a^n b^m c^m \mid n, m \in \mathbb{N}\}$*)*
2. *Complement (consider* $\{uw \mid |u| = |w| \land u \neq w\}$*)*

What is to come?

- Do we have to leave automata behind? (No)
- How do we prove that a language is not context free?