How Can We Define Languages in Computer Science?

Several scientific methods have been developed to precisely define the syntactic rules of languages.

Backus-Naur Form (BNF):

A **meta-language** used to describe the syntax of languages.

<name>	meta-symbol
::=	definition
|	alternative
\({expression}_{min}^{max}\)	repetition (minimum and maximum repetitions can be specified using subscripts)
GOTO	terminal symbol (in quotes for clarity, can also use apostrophes instead)

1. Example: Syntax of License Plates

Let’s start with a few typical examples and try to generalize:

ABC-935, GHT-234, HSD-333, AI-BB-654

Syntax definition: $$ <license\_plate>::=<newType> | <oldType> \\ <oldType>::=\{<letter>\}_3^3 − \{<number>\}_3^3 \\ <newType>::=\{<letter>\}_2^2 − \{<letter>\}_2^2 − \{<number>\}_3^3 \\ <letter>::=A|B|C...|Z \\ <number>::=0|1|2|3|4|5|6|7|8|9 \\ $$

2. Example: Syntax of Phone Calls in Hungary

Let’s list a few examples and try to generalize:

062012345, +36301234567, 0680460046

Syntax definition: $$ <phone call>::= \{<prefix>\}_0^1 <city> <customer> \\ <prefix>::=\{+\}_0^1 36|06 \\ <city>::=\{<number>\}_1^2\\ <customer>::=\{<number>\}_6^7\\ <number>::=0|1|2|3|4|5|6|7|8|9 \\ $$

3. Example: How can we describe the BNF formula using itself?

\[\langle \text{BN formula} \rangle ::= \langle \text{rule} \rangle \] \[\langle \text{rule} \rangle ::= \langle \text{identifier} \rangle "::=" \langle \text{expression} \rangle \] \[\langle \text{identifier} \rangle ::= "\langle" \langle \text{letter} \rangle \mid \left\lbrace \langle \text{letter} \rangle \mid \langle \text{digit} \rangle \right\rbrace "\rangle" \] \[\langle \text{expression} \rangle ::= \langle \text{term} \rangle \left\lbrace "|" \langle \text{term} \rangle \right\rbrace \] \[\langle \text{term} \rangle ::= \langle \text{factor} \rangle \left\lbrace \langle \text{factor} \rangle \right\rbrace \] \[\langle \text{factor} \rangle ::= \langle \text{identifier} \rangle \mid \langle \text{terminal_symbol} \rangle \] \[\langle \text{terminal_symbol} \rangle ::= \left\lbrace \langle \text{character} \rangle \rangle

\right\rbrace \] \[\langle \text{letter} \rangle ::= \langle \text{uppercase} \rangle \mid \langle \text{lowercase} \rangle \] \[\langle \text{uppercase} \rangle ::= A \mid B \mid C \dots Z \] \[\langle \text{lowercase} \rangle ::= a \mid b \mid c \dots z \] \[\langle \text{digit} \rangle ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9 \]