

Setul de reguli corespunzătoare sintaxei unui limbaj trebuie descris folosind un alt limbaj de descriere, numit *metalimbaj*. Folosirea unui limbaj natural, de exemplu, limba română sau limba engleză, ca metalimbaj are dezavantajul de a fi ambiguu, fiind posibile diferite interpretări. Studiul formal al sintaxei aplicat unui limbaj de programare a înregistrat un progres notoriu în anii '60, când a apărut specificarea limbajului Algol60 (Naur, 1960, 1963), care folosea o notație simplă și elegantă cunoscută sub numele de *Backus-Naur-Form (BNF)* [Nau60].

Backus-Naur-Form folosește următoarele obiecte:

- simboluri de bază (primitive) sau simboluri terminale, în general cuvinte cheie, operatori, separatori ai limbajului;
- variabile metalingvistice - simboluri neterminale, a căror semnificații corespund construcțiilor limbajului și se notează între paranteze unghiulare $\langle \rangle$;
- conective metalingvistice $::=$ (prin definiție egal) și $|$ (alternative).

Exemplul 1.1. Numerele întregi pot fi descrise în BNF astfel:

$$\begin{aligned}\langle \text{intreg} \rangle & ::= \langle \text{intreg_fara_semn} \rangle | \\ & \quad + \langle \text{intreg_fara_semn} \rangle | \\ & \quad - \langle \text{intreg_fara_semn} \rangle \\ \langle \text{intreg_fara_semn} \rangle & ::= \langle \text{cifra} \rangle | \\ & \quad \langle \text{cifra} \rangle \langle \text{intreg_fara_semn} \rangle \\ \langle \text{cifra} \rangle & ::= 0|1|\dots|9\end{aligned}$$

Notățiile folosite în EBNF² de către Wirth sunt:

- neterminalele sunt scrise ca și cuvinte (fără paranteze unghiulare: Intregfarasemn în loc de <intreg_fara_semn>)
- terminale sunt notate între ghilimele, de exemplu "BEGIN" (și de obicei cu majuscule)
- | este folosit, ca în BNF, pentru a nota alternanța
- parantezele rotunde (și) sunt folosite pentru a nota gruparea
- parantezele drepte [și] notează apariția opțională (zero ori sau o dată) a unui simbol sau grup de simboluri
- acoladele { și } notează repetarea opțională (o dată sau mai multe ori) a unui simbol sau grup de simboluri
- = este folosit în locul lui ::=
- punctul . este folosit ca sfârșit de regulă
- (* *) sunt folosite pentru comentarii
- ε, secvența vidă, poate fi descrisă prin notația []
- spațiile sunt, în general, nesemnificative.

Exemplul 1.2. Numerele întregi descrise în notație EBNF:

Intreg = Semn Intregfarasemn .

Intregfarasemn = cifra { cifra } .

Semn = ["+" | "-"] .

cifra = "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9" .