The Grammar of Syntax

hello@stephsamson.com
github.com/stephsamson

In English…

You still have much to learn.

Much to learn, you still have.

— Yoda

In Lisp-like languages...

(+ 1 2)

~~(1 + 2)~~

What describes this syntax?

Grammar

What is grammar?

A finite set of rules that describes a language's structure.

$$G = (N, T, S, P)$$

$$G = (N\ T\ S\ P)$$

T is a finite nonempty set of terminal symbols.

(+ 1 2)

T = ((+) 1 2)

$$G = (N\ T\ S\ P)$$

N is a finite nonempty set of non-terminal symbols.

$T \cap N = \varnothing.$

(+ 1 2)

N = (LEFT_PAREN, PLUS, RIGHT_PAREN, NUMERAL)

$$G = (N\ T\ S\ P)$$

S is the start symbol.

$$S \in N$$

(+ 1 2)

S → LEFT_PAREN EXPR

$$G = (N\ T\ S\ P)$$

P is a finite set of rewrite rules (also known as productions).

$$\alpha \rightarrow \beta$$

(+ 1 2)

S → LEFT_PAREN EXPR

EXPR → + LIST

LIST → NUM RIGHT_PAREN

NUM → 1 NUM → 2

Backus-Naur Form (BNF)

<symbol> ::= __expression__

<symbol> is a nonterminal

__expression__ consists >0 symbols

Backus-Naur Form

```
<start> ::= <left_paren> <expr>
<expr> ::= <plus> <list>
<list> ::= <num> <list> |
           <list> <right_paren> |
           <right_paren> ε
```

```
<num> ::= 1 | 2      <plus> ::= +
<left_paren> ::= (  <right_paren> ::= )
```

Software that use BNF

ANTLR

Yacc

Racket's parsers

Parsing

Given a grammar and an input, is the input a member of the language generated by that grammar?

Resources

http://www.cs.ucr.edu/~jiang/cs215/tao-new.pdf

Parsing Techniques by Dick Grune

Growing a Language by Guy Steele