Today’s topics

Language Translation
Revising the Syntax Rules and Notation

Upcoming
Language Translation
Generating Code

Reading
Great Ideas, Chapters 9

Importance of Language
- Vehicle for programming
- Use for human and machine communications

Syntax Rules
- First pass already covered
- Need to refine the notation
  - Must be suitable for machine to use
- In addition, need to deal with the meaning
- Also, should see Levels or Layers in dealing with computer
  1. Hardware
  2. Machine language
  3. Assembler
  4. Java (or other high-level language)
  5. Application (e.g. Word, Excel, Filemaker, …)

Language Translation
- Goal is to automatically
- Translate from Java:
  \[ z = x + y; \]
- to Assembler:
  copy ax, x
  add ax, y
  copy z, ax
- What is the meaning we are looking for?
  Machine gives assembler statements meaning because the machine knows what to do with them (after trivial translation to binary). E.g., the machine knows what add means.

Revise Syntactic Rules
- Need to revise Syntactic Production Rules
- New rule:
  R1: \(<n>j -> a sequence of letters and/or digits that begin with a letter
- Replaces (have seen this before):
  R1: \(<name> -> a sequence of letters and/or digits that begin with a letter
- The new R1 says “change \(<n>j into a sequence of letters and/or digits that begin with a letter”
- Use rules to modify strings
  - For syntactic productions, must end up with valid Java programs
Using Syntax Rules

- Examples using R1:
  \( <n>3 \rightarrow x \)
or\( <n>6 \rightarrow data \)
where “n” stand for “name”
- Further use of R1:
  \((<n>3 + <n>6)\)
Replace \(<n>3\) above to get
\((x + <n>6)\)
and \(<n>6\) to get
\((x + data)\)
- More Rules:
  \( R2 : <e>i \rightarrow <n>j \)
Where “e” stands for “expression”
  \( R3: <s>k \rightarrow <n>j = <e>i ; \)
Where “s” stands for “statement”
  It says “\(<s>k\)” can be replaced by “\(<n>j = <e>i ; \)”
- Can now do: \( ans = data ; \)

Using Syntax Rules

- and
  \( R3: <s>_k \rightarrow <n>_j = <e>_i ; \)
  Where “s” stands for “statement”
  It says “\(<s>_k\)” can be replaced by “\(<n>_j = <e>_i ; \)”
- Can now do: \( ans = data ; \)

More Rules

- Need two more rules to make it worthwhile
  \( R4: <e>_i \rightarrow ( <e>_j + <e>_k ) \)
  \( R5: <e>_i \rightarrow ( <e>_j * <e>_k ) \)
These are additional rules for expressions
- Can now handle \( ANS = (X + (Y * Z)) ; \)
  (notice shorthand/simplifications used)

Longer Example

\( ANS = (X + (Y * Z)) ; \)

\begin{align*}
\text{derivation} & \quad \text{rule} \\
&s1 & \quad R3: s1 \rightarrow n2 = e3; \\
n2 & \quad R1: n2 \rightarrow ANS \\
ANS & \quad R4: e3 \rightarrow (e4 + e5); \\
ANS & \quad R2: e4 \rightarrow n6 \\
ANS & \quad R1: n6 \rightarrow X \\
ANS & \quad R5: e5 \rightarrow (e7 * e8); \\
ANS & \quad R2: e7 \rightarrow n9 \\
ANS & \quad R1: n9 \rightarrow Y \\
ANS & \quad R2: e8 \rightarrow n10 \\
ANS & \quad R1: n10 \rightarrow Z \\
\end{align*}
Notes

- **Abbreviations**
  - Just omitted the angle brackets. Could do this because the notation remained *unambiguous*.

- **Role of the subscripts**
  - The subscripts are required to make sure each term is *unique*.
  - Simplest technique is to start subscripts at one and increment every time a different subscript is needed.

- **Simple substitution is all that is required!!!**
  - If you are doing something *more than that*, it is *probably wrong*!

- **The notation and form are important**
  - You will be expected to match them on tests.