Converting Derivation Trees to Abstract Syntax Trees

The tree produced during a parse of a sentence using a grammar is called a derivation tree. Derivation trees show every step of how we came to understand the structure of a sentence. Oftentimes, we don't need all of the information provided in a derivation. Indeed, sometimes the quantity of detail makes understanding a structure more difficult.

For this reason, we often use an alternative tree form when trying to understand a structure. An abstract syntax tree, or AST, gives us the essential structure of a parsed structure. Many students find ASTs difficult to understand at first, because the rules for converting a derivation tree to an AST vary from one language to another. Nevertheless, an AST always has the following properties:

- All of the interior nodes of an AST are operations.
- All of the leaf nodes of an AST are data.

Consider the lambda calculus expression $\lambda a.(ab)c$. Its derivation tree and corresponding AST are shown side-by-side.

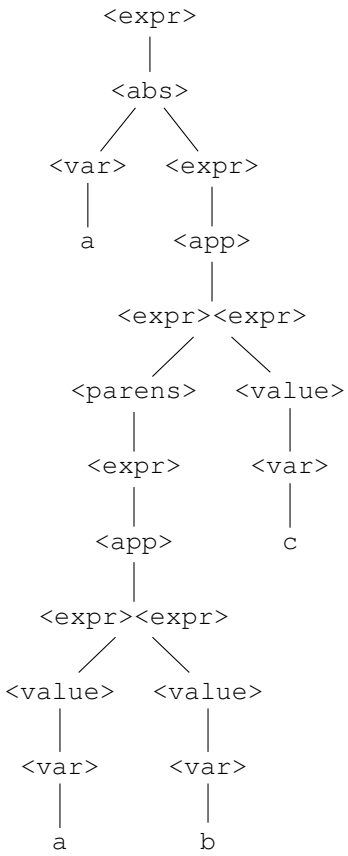

Figure 1: Derivation of $\lambda a.(ab)c$.

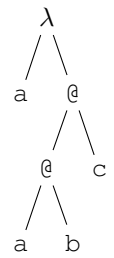

Figure 2: Abstract syntax tree for $\lambda a.(ab)c$.

Lambda Calculus Conversion Rules

In Figure 1 we have a complete record of a parse using the class lambda grammar. Figure 2 is a more compact representation of the same sentence, a small tree that shows only operations and data. What gets discarded depends on the language being analyzed. In the lambda calculus, the only operations are abstraction (λ) and application (`@`). Everything else is data.

With practice, you can derive an AST directly from a lambda expression. If you have trouble seeing how this might be done, start by producing a derivation tree, then try converting the tree into an AST using the rules below. When you are done, discard the topmost `<expr>`.

Figure 3: Variables, where α is a variable like `x`.

Figure 4: Numbers, where v is a number like `1`.

Figure 5: Parentheses, where e is some expression.

Figure 6: Application, where e_1 and e_2 are expressions.

Figure 7: Abstraction, where α is a variable like `x` and e is an expression.