Base types and compound types

Languages typically provide a small number of “built-in” types and ways to build compound types out of simpler ones:

- Base types examples: int, bool
- Type builder examples: tuples, lists, records

Base types clutter a language definition; better to make them libraries when possible.

- ML does this to a remarkable extent (e.g., we will soon define away bool and conditionals)

Good to let programmers bind types to type names, just like we bind values to variables.

Compound-type flavors

Conceptually, just a few ways to build compound types:

1. “Each-of”: A \( \downarrow \) contains a \( t_1 \) and a \( t_2 \)
2. “One-of”: A \( t \) contains a \( t_\bot \) or a \( t_2 \)
3. "Self-reference": The definition of \( t \) may refer to \( t \)

Examples:

- int \# bool
- int option
- int list

Remarkable: A lot of data can be described this way.
Convenient to think of as trees.

(optional) jargon: Product types, sum types, recursive types
User-defined types

There are many reasons to define your own types:

1. Using a tuple with 12 fields is incomprehensible
2. Writing down large types is unpleasant, error-prone; computers can help
3. Large programs can use abstract types to be robust to change
   - A couple weeks ahead
4. So the language doesn’t have to “build in” lists and options and ... that aren’t always needed

Datatype

One-of types are less similar across languages

- We’ll discuss OO’s approach to one-of in a few weeks

In ML, we use make a new type with a datatype binding, e.g.:

```plaintext
datatype myTree = SLeaf of string
               | ILeaf of int
               | Node of myTree * myTree
```

Semantics: Extend the environment with three constructors (in part, functions/constants that produce values of type myTree)

So we have a way to build them... what’s missing?

The old way

For lists, we had a way to:

- Test which variant a value was
- Extract the values from value-carrying variants
  - Makes no sense if you have the wrong variant

What would this look like for myTree?

The new way

Rather than add variant-tests and variant-deconstructors, ML has a case expression that uses pattern-matching.

In its simplest form, case has one pattern for each constructor in a datatype and binds one variable for each value carried. Example:

```plaintext
case e of
  Node(t1, t2) => e1
  SLeaf s => e2
  ILeaf i => e3
```

What are the typing rules?
What are the evaluation rules?
Type-checking case

In addition to binding local variables and requiring branches to have the same type, the typing rules for case prevent some run-time errors:

- Exhaustiveness: No test can “fail” (a warning)
- Redundancy: No test can be “impossible” (an error)

So far, case gives us what we need to use datatypes:

- A (combined) way to test variants and extract values (deconstruct)
- Powerful enough to define our own tests and deconstructors

In fact, pattern-matching is far more general and elegant:

- Can use it for datatypes already in the top-level environment
- Can use it for any type (later)
- Can have deep patterns (later)