Lecture 2: Parsing Propositional Formulas

1 Review of Syntax

Our alphabet of symbols consists of:

- PROP, a set of atomic propositions
- $\neg$, a unary connective, and $\{\wedge, \vee, \rightarrow, \leftrightarrow\}$, a set of binary connectives.
- '(' and ')', left and right parenthesis.

Any string built from these symbols is called an *expression*. E.g., '$)p \rightarrow \vee q$' is an expression. *Formulas* are expressions with a specific structure. We define FORM, the set of all formulas in two different but equivalent ways.

1.1 Definition of FORM

Given a set PROP of propositions, we define the set FORM to be the smallest set satisfying the following two properties.

Basis: PROP $\subseteq$ FORM, and

Closure: $\varphi, \psi \in$ FORM implies $(\neg\varphi) \in$ FORM and $(\varphi \circ \psi) \in$ FORM for any binary connective $\circ$.

Alternatively, we can define FORM using induction as follows:

$$\text{FORM}_0 = \text{PROP}$$

and, for each $i \geq 0$,

$$\text{FORM}_{i+1} = \text{FORM}_i \cup \{(\neg\varphi), (\varphi \circ \psi) \mid \varphi, \psi \in \text{FORM}_i\}.$$

We set

$$\text{FORM}' = \bigcup_{i=0}^{\infty} \text{FORM}_i,$$

which is equivalent to FORM (see notes from previous lecture).

Lemma 1. FORM = FORM$'$.

2 Parsing and Unique Readability

Consider the formula $((p \to q) \to ((q \to r) \to (p \to r)))$. If you are given an expression, how can you tell if it is valid? Parsing lets us examine an expression to determine validity. The following will help us formalize this problem.

Definition 1. *Formulas are either atomic or composite.*

- *The set of atomic formulas = Prop.*
- *For a composite formula* $(\neg\varphi)$

  $\neg$ *is the primary connective*

  φ *is the immediate sub-formula*

- *For a composite formula* $(\varphi \circ \psi)$

  $\circ$ *is the primary connective*

  φ, ψ *are immediate sub-formulas*

We would like to show that our language is not ambiguous, i.e. that there is only one way to read a formula in FORM.

Theorem 1. Unique Readability *Every composite formula has a unique primary connective and well-defined immediate sub-formulas.*

To prove this, we first need the following definition:

Definition 2 (Prefix)**.** *An expression* α *is a prefix of an expression* β *if there exists an expression* γ *such that* $\beta = \alpha\gamma$*.* α *is a* strict prefix *if* $\gamma \neq \epsilon$*.* α *is a nonempty prefix if* $\alpha \neq \epsilon$*.*

Proof of Unique Readability: We prove this using structural induction.

Basis: φ is an atomic proposition. Since atomic propositions are uniquely readable, this statement is trivially true.

Inductive Step: We must consider two cases.

1. Suppose φ is $(\neg\theta)$.
   - φ is not atomic
   - Suppose φ is also $(\neg\theta')$. It follows that $\theta = \theta'$, by string matching. So θ is unique.
   - φ is also $(\alpha \circ \beta)$. So α, which is a formula, must start with $\neg$. This is impossible, because a formula is either atomic or starts with (.
2. Suppose φ is $(\alpha \circ \beta)$.
   - φ is not atomic

- φ is not $(\neg\beta)$, because the formula α cannot start with $\neg$.
- Suppose φ is also $(\alpha' \circ' \beta')$. Then
  - either $\alpha = \alpha', \circ = \circ'$ and $\beta = \beta'$
  - or α is a strict prefix of α', or α' is a strict prefix of α. If we could show that this situation is impossible, then our proof would be complete.

Let's attempt to prove the following lemma directly:

Lemma 2. Prefix Lemma *A strict prefix θ of a formula φ is not a formula.*

Proof. We use structural induction

Basis: φ is atomic. This is trivially true because the only strict prefix of an atomic proposition is the empty string ϵ, which is not a formula.

Inductive step:

- φ is $(\neg\psi)$.

  θ is ϵ, which is not a formula.

  θ is (, which is not a formula..

  θ is $(\neg$, which is not a formula.

  θ is $(\neg\psi'$ where ψ' is a strict, nonempty prefix of φ. We are stuck here! We need to distinguish a prefix of a formula from a formula.

□

The straightforward approach failed in this case. The problem was that the statement we were trying to prove using induction was *too weak*. An induction proof depends critically on using the induction hypothesis in proving the inductive step, and the hypothesis is simply a limited version of the original statement we are trying to prove. So sometimes a stronger statement is easier to prove, because it makes the induction hypothesis more powerful.

We thus attempt another proof. Before we proceed, we need a few more definitions:

Definition 3. Length and Parenthesis Counting
The length of an expression α, denoted $length(\alpha)$, is the number of letters in the expression. The number of left parenthesis '(' that occur in α is denoted by $\ell(\alpha)$, and the number of right parenthesis ')' is denoted by $r(\alpha)$.

The length of a formula is the same as its length as an expression, but it can also be defined inductively as follows:

- $length(p) = 1$, *where* $p \in \mathrm{PROP}$.
- $length((\neg\varphi)) = 3 + length(\varphi)$
- $length((\varphi \circ \psi)) = 3 + length(\varphi) + length(\psi)$

Lemma 3. *If φ a formula then $\ell(\varphi) = r(\varphi)$.*

Proof. By induction.

Base case: $\varphi \in \text{PROP}$. Then $\ell(\varphi) = r(\varphi) = 0$.

Inductive step:

1. φ is $(\neg\theta)$. Then, by IH, $\ell(\varphi) = 1 + \ell(\theta) = 1 + r(\theta) = r(\varphi)$.
2. φ is $(\theta \circ \psi)$. Then, by IH, $\ell(\varphi) = 1 + \ell(\theta) + \ell(\psi) = 1 + r(\theta) + r(\psi) = r(\varphi)$.

□

Next, we'll show that if you stop reading a formula too early, you end up with too many '('s. The number of parenthesis is a distinguishing characteristic of formulas.

Lemma 4. *If θ is a nonempty strict prefix of a formula φ, then $\ell(\theta) > r(\theta)$.*

Proof. By induction.

Basis: $\varphi = p$, where $p \in \text{PROP}$. Then ϵ and p are the only prefixes of φ. Since no nonempty strict prefix exists, the statement is trivially true.

Inductive step:

1. φ is $(\neg\psi)$.
   (a) θ is '('. The statement is obviously true.
   (b) θ is '$(\neg$'. The statement is obviously true.
   (c) θ is '$(\neg\psi_1$' where ψ_1 is a strict prefix of formula ψ. Then, by IH, $\ell(\theta) = 1 + \ell(\psi_1) > 1 + r(\psi_1) > r(\psi_1) = r(\theta)$.
   (d) θ is '$(\neg\psi$'. Then, by Lemma 3, $\ell(\theta) = 1 + \ell(\psi) = 1 + r(\psi) > r(\psi) = r(\theta)$.
2. φ is $(\alpha \circ \beta)$.
   (a) $\theta =$ '('. The statement is obviously true.
   (b) $\theta =$ '$(\alpha_1$', where α_1 is a strict non-empty prefix of α. Then, by IH, $\ell(\theta) = 1 + \ell(\alpha_1) > 1 + r(\alpha_1) > r(\alpha_1) = r(\theta)$.
   (c) $\theta =$ '$(\alpha$'. Then, by Lemma 3, $\ell(\theta) = 1 + \ell(\alpha) = 1 + r(\alpha) = 1 + r(\theta) > r(\theta)$.
   (d) $\theta =$ '$(\alpha\circ$'. The proof is same as the previous case.
   (e) $\theta =$ '$(\alpha \circ \beta_1$', where β_1 is a strict non-empty prefix of β. Then, by IH and Lemma 3, $\ell(\theta) = 1 + \ell(\alpha) + \ell(\beta_1) = 1 + r(\alpha) + \ell(\beta_1) > 1 + r(\alpha) + r(\beta_1) = 1 + r(\theta) > r(\theta)$.
   (f) $\theta =$ '$(\alpha \circ \beta$'. Then, by Lemma 3, $\ell(\theta) = 1 + \ell(\alpha) + \ell(\beta) = 1 + r(\alpha) + r(\beta) = 1 + r(\theta) > r(\theta)$.

□

We now have the tools to prove the Prefix Lemma directly.

Proof of Prefix Lemma: Let θ be a strict prefix of formula φ. Then, by Lemma 4, $\ell(\theta) > r(\theta)$. If θ is itself a formula, then, by Lemma 3, we get $\ell(\theta) = r(\theta)$, which is a contradiction. Therefore θ cannot be a formula.

This completes the proof of unique readability.