9.1 Skolemization: Getting rid of the $\exists$’s

Idea: Once $\varphi$ is in RPF, we will convert $\varphi$ to it’s Skolemization: $\varphi_S$. To do this, we replace existentially quantified variables with a new function symbol applied to the variables that have been previously universally quantified.

We will see:

**Theorem 9.1** (Trevor & Skolem’s Theorem) For any $\varphi \in L(\Sigma)$, we can construct universal formula $\varphi_S \in L(\Sigma')$ such that $\varphi$ and $\varphi_S$ are equi-satisfiable, i.e., $\varphi \in \text{FO-SAT} \iff \varphi_S \in \text{FO-SAT}$. ($\Sigma'$ is the result of adding some new Skolem function symbols to $\Sigma$.)

$\varphi_S$ will usually not be equivalent to $\varphi$, but it will be the case that $(\varphi_S \to \varphi) \in \text{FO-VALID}.$

**Example 9.2** Let $\text{od2}(x) \equiv \exists y \exists z \forall w (E(x,y) \land E(x,z) \land y \neq z \land (\neg E(x,w) \lor w = y \lor w = z)).$ The formula $\text{od2}(x)$ has $x$ as a free variable and says that vertex $x$ has out-degree 2. To Skolemize $\text{od2}(x)$, we will get rid of the “$\exists y$” and replace all occurrences of $y$ by the term $f_y(x)$ where $f_y$ is a new unary function symbol. Similarly for “$\exists z$”

\[
\text{od2}(x) \equiv \forall w (E(x,f_y(x)) \land E(x,f_z(x)) \land f_y(x) \neq f_z(x) \land (\neg E(x,w) \lor w = f_y(x) \lor w = f_z(x))). \quad \square
\]

**Example 9.3** $\alpha \equiv \exists x \forall y (x + y = y).$ Note that $\alpha$ says that there is an identity element for addition. $\alpha_S \equiv \forall y (c + y = y).$ This says that $c$ is the identity element for addition. Note that $\alpha_S$ implies $\alpha$ and it is more specific: not only is there an identity element for addition, but we pick a particular one and give it a name, the new Skolem constant symbol, $c.$ \square

**Example 9.4** $\beta \equiv \forall x \exists y E(x,y).$ $\beta$ says that every vertex has an outgoing edge.

$\beta_S \equiv \forall x E(x,f(x))$

Here $f$ is a new unary Skolem function symbol. Observe that $\beta_S \to \beta.$ $\beta_S$ is more specific than $\beta.$ Not only does every vertex have at least one outgoing edge, but the function $f$ picks one. \square

For any $\varphi \in L(\Sigma)$, we produce $\varphi_S \in L(\Sigma')$ which is universal and in RPF-CNF form and is equi-satisfiable with $\varphi$, i.e, $\varphi \in \text{FO-SAT} \iff \varphi_S \in \text{FO-SAT}.$

$$\varphi_s \equiv \forall x_1 \cdots x_k \bigwedge_{i=1}^{a} \bigvee_{j=1}^{b} \ell_{ij}$$

9.2 Herbrand Theory:

We will assume that our vocabulary, $\Sigma$, always includes at least one constant symbol, $c$. For example, let $\Sigma = (R_1^{r_1}, \ldots, R_s^{r_s} ; c, c_2, \ldots, c_k, f_1^{r_1}, \ldots, f_t^{r_t}).$ recall that we defined term $(\Sigma)$ by induction. A term is closed if it has no variables. Since every vocabulary has the constant symbol, $c$, $c \in \text{closedTerm}(\Sigma)$, so $\text{closedTerm}(\Sigma) \neq \emptyset.$
Example 9.5 \( \Sigma_{\text{st-graph}} = (E^2; s, t) \). \( \text{closedTerm}(\Sigma_{\text{st-graph}}) = \{s, t\} \).

Example 9.6 Example: \( \Sigma_{\text{#th}} = (\leq^2; 0, 1, +^2[\text{infix}], *^2[\text{infix}]) \). \( \text{closedTerm}(\Sigma_{\text{#th}}) = \{0, 1, 0 + 0, 0 + 1, 0 \ast 0, (1 + 1) \ast (1 + 1), (1 + 1) \ast (1 + 1) + 1, \ldots\} \).

Recall that each term \( t \in \text{term}(\Sigma) \) is a sequence of symbols. Each structure \( A \in \text{STRUC}[\Sigma] \) interprets the term \( t \) as an element of its universe: \( t^A \in |A| \).

Definition 9.7 The Herbrand Universe for \( (\Sigma) \) \( \equiv \text{closedTerm}(\Sigma) \).

Definition 9.8 Def. A Herbrand Structure \( H \in \text{STRUC}[\Sigma] \) has the Herbrand Universe as its universe, i.e., \( |H| = \text{closedTerm}(\Sigma) \). Furthermore, for every function symbol \( f^r \in \Sigma \), and every \( t_1, \ldots, t_r \in |H| \), we have,

Gizem’s Condition: \( f^H(t_1, \ldots, t_r) = f(t_1, \ldots, t_r) \).

Proposition 9.9 Let \( H \) be a Herbrand Structure of vocabulary \( \Sigma \) and let \( t \) be any closed term of \( \Sigma \). Then \( t^H = t \).

Proof: By induction on \( t \).

Definition 9.10 Let \( \varphi \in \mathcal{L}(\Sigma) \). If \( H \) is a Herbrand structure and \( H \models \varphi \) then we say that \( H \) is a Herbrand model of \( \varphi \).

Example 9.11 \( \Sigma_{\text{st-graph}} \) has exactly 16 Herbrand structures: \( H_0, \ldots, H_{15} \) all with universe \( \{s, t\} \), and with the 16 possible interpretations of \( E \) on a two-element universe. For example, let \( E^{H_2} = \{(t, s)\} \). Note that \( H_2 \models \forall x \, (\neg E(x, x)) \). Thus \( H_2 \) is a Herbrand model of \( \forall x \, (\neg E(x, x)) \).

Let \( \beta = \exists x \exists y \, (R(x) \land \neg R(y)) \). Let \( A_2 = (\{0, 1\}, R^{A_2} = \{1\}) \). Note that \( A_2 \models \beta \). However, \( H_0 \models \neg \beta \) and \( H_1 \models \neg \beta \). Thus, all the Herbrand structures of vocabulary \( \Sigma_0 \) satisfy \( \neg \beta \). Thus, \( \beta \) is satisfiable, but has no Herbrand model with vocabulary \( \Sigma_0 \).

Theorem 9.12 (Herbrand’s Theorem) Let \( \Sigma \) be a vocabulary s.t. \( \text{closedTerm}(\Sigma) \neq \emptyset \). Let \( \varphi \in \mathcal{L}(\Sigma) \) be a universal sentence, \( \varphi = \forall x_1 \cdots x_k(\alpha) \) where \( \alpha \) is quantifier free. Then

\( (\varphi \in \text{FO-SAT}) \iff (\varphi \text{ has a Herbrand model}) \).

Proof: \( \Leftarrow \): If \( \varphi \) has a Herbrand model, \( H \models \varphi \), then \( \varphi \) has a model, so it is satisfiable.

\( \Rightarrow \): Assume that \( \varphi \in \text{FO-SAT} \) and let \( A \models \varphi \).

Goal: define a Herbrand structure, \( H \), s.t. \( H \models \varphi \).

Two parts of building \( H \) are trivial, i.e., we must have that \( |H| = \text{closedTerm}(\Sigma) \) and for all \( f^r \in \Sigma \) and \( t_1, \ldots, t_r \in \text{closedTerm}(\Sigma) \), \( f^H(t_1, \ldots, t_r) = f(t_1, \ldots, t_r) \).

What remains to be defined is \( R^H \) for each \( R^a \in \Sigma \). For these definitions, we just ask \( A \):

\[
R^H \overset{\text{def}}{=} \{(t_1, \ldots, t_a) \in |H|^a \mid A \models R(t_1, \ldots, t_a)\} .
\]
**Claim.** For all closed quantifier-free \( \beta \in L(\Sigma) \), \( H \models \beta \iff A \models \beta \).

**Proof:** By induction on \( \beta \).

**base case:** \( \beta = R(t_1, \ldots, t_a) \). By (\( * \)), \( H \models \beta \iff A \models \beta \).

**inductive case:** Assume true for \( \beta_1 \) and \( \beta_2 \).

\[
H \models \lnot \beta_1 \iff H \not\models \beta_1 \iff A \not\models \beta_1 \iff A \models \lnot \beta_1
\]

\[
H \models \beta_1 \lor \beta_2 \iff (H \models \beta_1) \text{ or } (H \models \beta_2) \iff (A \models \beta_1) \text{ or } (A \models \beta_2) \iff A \models \beta_1 \lor \beta_2.
\]

\( \square \)

Finally we show that \( H \models \forall x_1 \cdots x_k (\alpha) \).

Since \( A \models \forall x_1 \cdots x_k (\alpha) \), it follows that for all closed terms, \( t_1, \ldots, t_k \), \( A[t_1^A/x_1, \ldots, t_k^A/x_k] \models \alpha \).

Thus, by the Translation Lemma, for all closed terms, \( t_1, \ldots, t_k \), \( A \models \alpha[t_1^A/x_1, \ldots, t_k^A/x_k] \).

Thus, by our Claim, for all closed terms, \( t_1, \ldots, t_k \), \( H \models \alpha[t_1^A/x_1, \ldots, t_k^A/x_k] \).

Thus, by the Translation Lemma, for all closed terms, \( t_1, \ldots, t_k \), \( H[t_1/x_1, \ldots, t_k/x_k] \models \alpha \).

Thus, for all \( t_1, \ldots, t_k \in |H| \), \( H[t_1/x_1, \ldots, t_k/x_k] \models \alpha \).

Thus, by Tarski’s Definition of Truth, \( H \models \forall x_1 \cdots x_k (\alpha) \). \( \square \)