## LECTURE 2: LANGUAGES, TERMS, FORMULAS

**Warning!** At the beginning of lecture 1, when discussing set-theoretic notions, the warning was made that we would, when need be, use $\langle$ and $\rangle$ to enclose finite ordered sequences in set theory. This is to avoid confusion with the rounded parenthesis that are part of our formal language. This danger of confusion is now imminent, and so the use of $\langle$ and $\rangle$ to enclose finite ordered sequences is now in effect.

In Lecture 2, we introduced first order languages, as well as term and formulas in such a language. The material covered is roughly section 2.1 (pp. 69–79). Below is the slightly different approach that I took in lecture. Before we begin, it is practical to have one new set-theoretic convention in place.

**Definition 0.1.** We let $\mathbb{V}$ denote the (sometimes non-existent[1]) totality of *all* sets. That is, $\mathbb{V}$ is the set-theoretic "universe".

The universe $\mathbb{V}$ must *not* be confused with the set $V$ of variable symbols!

### 1. Terms and formulas

**Notation.** Remember that an expression in a language is any finite sequence (string) of symbols in $\mathcal{L}$ (including logical symbols). Formally, the string $\forall = v_2$ would be written $\langle \forall, =, v_2 \rangle$, but we don't do that. Also, for notational simplicity, we identify the sequence $\langle x \rangle$ of length 1 with $x$. Remember that if $\varepsilon$ and $\delta$ are finite sequences, then we write $\varepsilon\delta$ for the sequence where we append $\delta$ to the end of $\varepsilon$. So if $\varepsilon$ is the expression $(\forall$ and $\delta$ is $v_1)$, then $\varepsilon\delta$ is the expression $(\forall v_1)$. More generally, if $\varepsilon_1, \ldots, \varepsilon_m$ are finite sequences, then $\varepsilon_1 \ldots \varepsilon_m$ is the sequence obtained by concatenating the sequences in the given order.

**Terms.** Let $\mathcal{L}$ be language[2]. In lecture we defined

$$\mathrm{Term}_0^{\mathcal{L}} = \{\langle a \rangle : a \text{ is a variable or constant symbol}\},$$

and then, recursively, that $t \in \mathrm{Term}_{n+1}^{\mathcal{L}}$ if and only if there is some $m$ and an $m$-place function symbol $F$ in $\mathcal{L}$, and terms $t_1, \ldots, t_m \in \bigcup_{0 \le i \le n} \mathrm{Term}_i^{\mathcal{L}}$ such that $t$ is $F t_1 \ldots t_m$. (Formally, we should write $\langle F \rangle t_1 \ldots t_m$ for $F t_1 \ldots t_m$, but, as noted above, we identify $\langle F \rangle$ and $F$ for notational simplicity. Commas are *not* part of our formal language, and so only informally may we write $F(t_1, \ldots, t_n)$ rather than $F t_1 \ldots t_n$.) We finally defined $\mathrm{Term}(\mathcal{L}) = \bigcup_{n \in \omega} \mathrm{Term}_n^{\mathcal{L}}$, and convinced ourselves that $t$ is a term (in $\mathcal{L}$) if and only if $t \in \mathrm{Term}(\mathcal{L})$. The *complexity* of a term $t$ is the least $n$ such that $t \in \mathrm{Term}_n^{\mathcal{L}}$.

---

[1]Those who've heard of Russell's paradox may feel uneasy about $\mathbb{V}$, but rest assured that it will not be used in any potentially dangerous way.

[2]Recall that in lecture we decided to include equality, $=$, as a logical symbol, whereas the book treats it as an optional part of the language. This will not make a big difference, but it slightly alters the structure of some definitions.

**Exercise 1.** Show that $\mathrm{Term}_n^{\mathcal{L}} \subseteq \mathrm{Term}_{n+1}^{\mathcal{L}}$ for $n > 0$. What about $n = 0$? Is it possible for $\mathrm{Term}_n^{\mathcal{L}} = \emptyset$ when $n > 0$? What about $n = 0$? □

**Exercise 2.** Prove the induction principle for terms: If $\Pi$ is a set of terms such that

(1) $\mathrm{Term}_0^{\mathcal{L}} \subseteq \Pi$;
(2) if $F$ is an $m$-place function symbol and $t_1, \dots, t_m \in \Pi$ then $Ft_1 \dots t_m \in \Pi$,

then $\Pi = \mathrm{Term}(\mathcal{L})$. (Hint: Use induction on complexity). □

**Formulas.** In lecture, we defined $\mathrm{Formula}_0^{\mathcal{L}}$, the set of *atomic formulas*, to be the set of $\varphi$ such that $\varphi$ is either $= t_1 t_2$, where $t_1, t_2 \in \mathrm{Term}(\mathcal{L})$, or $\varphi$ is $Rt_1 \dots t_k$ where $R$ is a $k$-place relation symbol, and $t_1, \dots, t_k \in \mathrm{Term}(\mathcal{L})$. We then defined, by recursion on $n \in \omega$, that $\mathrm{Formula}_{n+1}^{\mathcal{L}}$ consists of the $\varphi$ such that one of the following three holds:

(A) $\varphi$ is $(\neg\psi)$ for some $\psi \in \bigcup_{m \leq n} \mathrm{Formula}_m^{\mathcal{L}}$;
(B) $\varphi$ is $(\psi \to \theta)$, where $\psi, \theta \in \bigcup_{m \leq n} \mathrm{Formula}_m^{\mathcal{L}}$;
(C) $\varphi$ is $\forall v_i \psi$ for some $v_i \in V$ and $\psi \in \bigcup_{m \leq n} \mathrm{Formula}_m^{\mathcal{L}}$.

We let $\mathrm{Formula}(\mathcal{L}) = \bigcup_{n \in \omega} \mathrm{Formula}_n^{\mathcal{L}}$. In lecture we convinced ourselves that $\varphi$ is a formula (also called a *wff* in the book) in $\mathcal{L}$ if and only if $\varphi \in \mathrm{Formula}(\mathcal{L})$. The *complexity* of a formula $\varphi$ is the least $n$ such that $\varphi \in \mathrm{Formula}_n^{\mathcal{L}}$.

**Exercise 3.** Write down the atomic formulas in the language $\mathcal{L} = \{\in\}$, where $\in$ is a binary relation symbol, and the language includes equality.

**Exercise 4.** Show that $\mathrm{Formula}_n^{\mathcal{L}} \subseteq \mathrm{Formula}_{n+1}^{\mathcal{L}}$ for $n > 0$. What happens with $n = 0$? □

**Exercise 5.** Prove the *induction principle for formulas*: If $\Phi \subseteq \mathrm{Formula}(\mathcal{L})$ is a set of formulas such that $\mathrm{Formula}_0^{\mathcal{L}} \subseteq \Phi$ and it holds that

(A) if $\varphi \in \Phi$ then $(\neg\varphi) \in \Phi$;
(B) if $\varphi, \psi \in \Phi$ then $(\varphi \to \psi) \in \Phi$;
(C) if $\varphi \in \Phi$ then for any $v_i \in V$, $\forall v_i \varphi \in \Phi$;

then $\Phi = \mathrm{Formula}(\mathcal{L})$. (Hint: Use induction on complexity.) □

**Exercise 6.** Use the induction principle for formulas to show that in any formula $\varphi$ there is an equal number of start parenthesis and end parenthesis. □

## 2. Really Formalistic People

This section is optional reading. If you plan to go into higher set theory later in life, you probably should read it and do the exercises.

**Exercise 7.** Some people[3] like to be really formal (and some have good reasons to). One such person I know defines a language as follows: A language is a pair $\langle f, p \rangle$ where

(1) $f : \omega \to \mathbb{V}$;
(2) $p : \omega \setminus \{0\} \to \mathbb{V}$;
(3) $(\forall m \in \omega)(\forall n \in \omega)(f(m) \cap p(n) = \emptyset \wedge (m \neq n \implies (f(m) \cap f(n) = \emptyset = p(m) \cap p(m))))$;

[3] We will call such a person a *Really Formalistic Person*, or RFP.

(4) for all $n \in \omega$, $f(n)$ and $p(n)$ are disjoint from $\{11 + 2n : n \in \omega\} \cup \{0, 1, 2, 3, 5, 7, 9\}$;
(5) no function whose domain is in $\omega \setminus \{\emptyset\}$ belongs to any $f(n)$ or $p(n)$.

Explain. (Hint: The same person also defines that the variables $v_1, v_2, \ldots$ officially are the sets $11, 13, 15, \ldots$, that ( is 1, that ) is 3, that $=$ is 9, that $\neg$ is 5, $\to$ is 7, and $\forall$ is 0.) □

Given $g : m \to \mathbb{V}$ and $h : n \to \mathbb{V}$, where $m, n \in \omega$, we let $g^\frown h : m + n \to \mathbb{V}$ be given by

$$g^\frown h(k) = \begin{cases} g(k) & \text{if } k < m, \\ h(j) & \text{if } k = m + j \text{ and } j < n. \end{cases}$$

If $h$ is a finite sequence of finite sequences (that is, informally, $h \in (\mathbb{V}^{<\omega})^{<\omega}$), we define $\operatorname{concat}(h)$, the *concatenation* of $h$, recursively as:

$$\operatorname{concat}(h) = \begin{cases} \emptyset & \text{if } \ell\mathrm{h}(h) = 0, \\ (\operatorname{concat}(h \restriction n))^\frown h(n) & \text{if } \ell\mathrm{h}(h) = n + 1. \end{cases}$$

**Exercise 8.** Let $\mathcal{L} = (f, p)$ be a language in the really formal sense. A RFP defines terms as follows:

$$\mathrm{Term}_0^{\mathcal{L}} = \{\langle a \rangle : a \text{ is a variable or constant symbol}\},$$

and recursively, $\mathrm{Term}_{n+1}^{\mathcal{L}}$ is defined to be the set of all $\operatorname{concat}(h)$ such that for some $k \in \omega \setminus \{0\}$,

(a) $h : k + 1 \to \mathbb{V}$;
(b) $h(0) \in \{\langle a \rangle : a \in f(k)\}$;
(c) $(\forall j < k) h(1 + j) \in \bigcup_{m \leq n} \mathrm{Term}_m^{\mathcal{L}}$.

Then define a term of $\mathcal{L}$ to be any element of $\mathrm{Term}(\mathcal{L}) = \bigcup_{n \in \omega} \mathrm{Term}_n^{\mathcal{L}}$. Explain how this really formal definition relates to our usual definition of a term. □

**Exercise 8*.** Give a definition of formulas and of $\mathrm{Formula}_n^{\mathcal{L}}$ worthy of a RFP. □