Linguagens formais

Um conceito importante para definição de computabilidade é o de linguagem formal.
O uso de linguagens viabiliza a comunicação de ideias, conceitos, etc.
Estamos habituados ao uso de linguagens naturais (português, inglês) para nos comunicar.
Nessas linguagens usamos símbolos para construção de palavras que carregam sentido e, em conjunto, formam o que chamamos efetivamente de língua.

As linguagens naturais são carregadas de ambiguidades semânticas que inviabilizam seu uso na definição formal de problemas a serem tratados por máquinas. Dessa forma, optamos por usar **linguagens formais que possuem as seguintes características**:

1. Têm sintaxe bem definida; o que permite sempre verificar se uma sentença pertence ou não à linguagem;
2. Têm semântica precisa, de forma que, toda sentença sempre possui um único significado.

Nós profissionais da computação usamos linguagens formais diariamente. As linguagens de programação, por exemplo, são amostras concretas dessas linguagens.

Como veremos, linguagens formais são usadas na definição de problemas de decisão. Tais problemas estão no cerne de computabilidade. Logo, não há como estudar máquinas e seu poder computacional sem estudar linguagens formais. Contudo, para esse contexto, basta focarmos no aspecto sintático, não sendo necessário estudarmos questões semânticas relacionadas a elas.

A base de qualquer linguagem (formal) é seu alfabeto. Um **alfabeto** é um conjunto finito não vazio de símbolos. Em geral, denotamos um alfabeto por uma letra grega maiúscula (na maioria das vezes Σ ou Γ)

Exemplo:

1. $\Sigma = \{0, \ 1\}$, o alfabeto binário
2. $\Sigma =$ { ✋ , ✊ , ✌ } , o alfabeto do pedra-papel-tesoura
3. $\Sigma = \{ \ a, \ b, \ \ldots, \ z\}$, o alfabeto latino

Uma **palavra** é uma sequência finita de símbolos de um alfabeto Σ. O tamanho (quantidade de símbolos) de uma palavra w é denotado por $|w|$. Uma palavra especial é a **palavra vazia** constituída de zero símbolos e denotada por λ. Ou seja, $|\lambda| = 0$.
Outros exemplos de palavra são:

Exemplo:

1. 0011 construída com o alfabeto binário; $|0011| = 4$.
2. $0^4 = 0000$ construída com o alfabeto $\Sigma = \{0\}$; o expoente aqui denota repetições de símbolos.

3. $0^2 1^2 0^2 = 001100$

☆ O conjunto de todas as palavras construtíveis com um alfabeto Σ é denotado por Σ^* (sigma estrela).

Exemplo:

1. $\{0, 1\}^* = \{\lambda, \ 0, \ 1, \ 00, \ 01, \ 10, \ 11, \ \ldots\}$
2. { ✋ , ✊ , ✌ }* = {λ, ✋ , ✊ , ✌ , ✋ ✊ , ✋ ✌ , ...}

Note que os conjuntos acima são conjuntos de palavras. Logo, embora 0 seja um símbolo do alfabeto, ele representa uma palavra em $\{0, 1\}^*$. O contexto deve deixar evidente se estamos tratando de um símbolo ou uma palavra.

☆ Uma **linguagem** é um subconjunto de palavras de Σ^*.

Exemplo: Considere o alfabeto $\Sigma = \{0, \ 1\}$.

1. $\emptyset$ é a linguagem mais simples que existe; não contém palavras.
2. $\{\lambda\}$ é a linguagem que contém uma única palavra, a palavra vazia.
3. $\{w \in \Sigma^* \mid \ 1 \leq |w| \leq 5\}$ é a linguagem de todas as palavras de tamanho 1 a 5. Ela contém $\sum_{i=1}^{5} 2^i$ palavras.
4. $\{0^n 1^n \ \mid \ n \in \mathbb{N}\}$ é a linguagem de todas as palavras de tamanho par cuja primeira metade só contém 0's e a segunda metade 1's.
5. Σ^* é a linguagem de todas as palavras.

As linguagens de maior interesse são as infinitas. Como linguagens são conjuntos, as operações sobre conjuntos também se aplicam as linguagens. Isso torna mais fácil a especificação de linguagens mais complexas.

Considere L_1 e L_2 duas linguagens definidas respectivamente sobre os alfabetos Σ_1 e Σ_2. As operações sobre conjuntos definem novas linguagens como:

1. $L_1 \cup L_2$ uma linguagem definida sobre $\Sigma_1 \cup \Sigma_2$
2. $L_1 \cap L_2$ uma linguagem sobre $\Sigma_1 \cap \Sigma_2$
3. $L_1 - L_2$ uma linguagem sobre Σ_1
4. $\overline{L_1}$ uma linguagem sobre Σ_1

Veja que as operações $\mathcal{P}(L_1)$ e $\mathcal{P}(\Sigma_1^*)$ também estão definidas e denotam <u>um conjunto</u> de linguagens sobre Σ_1 e <u>o conjunto de todas</u> as linguagens sobre Σ_1.

Exemplo: Sejam $L_1 = \{w \in \{0, 1\}^* \ \mid \ |w| \ \text{é} \ par\}$ e

$L_2 = \{w \in \{0,1\}^* \;\; | \;\; w \;\; começa \;\; com \;\; 0\}$.

1. $L_1 \cup L_2$

2. $L_1 \cap L_2$

3. $\overline{L_1}$

4. $L_1 - L_2$

5. $\mathcal{P}(L_1)$

6. $\mathcal{P}(\{0,1\}^*)$