

## Capítulo 6

# Processamento de Linguagens

Algumas das tarefas de processamento de linguagens vão além do simples reconhecimento da linguagens, por exemplo quando falamos de compilação o objectivo é não só o de reconhecimento do texto de entrada como também a sua tradução para uma outra linguagem. Temos então uma tarefa distinta da do reconhecimento e que se designa por tradução.

**Definição 6.1 (Tradução)** *Uma tradução da linguagem  $\mathcal{L}_i$  para a linguagem  $\mathcal{L}_j$ , é uma função  $t : \mathcal{L}_i \rightarrow \mathcal{L}_j$ .*

### 6.1 Analisadores Léxicos

Pode-se estender os reconhecedores baseados em AFD com acções semânticas associadas às regras de reconhecimento, os autómatos que se obtêm são designados por autómatos finitos determinísticos reactivos AFDR.

**Definição 6.2 (Autómato Finito Determinístico Reactivo)** *Um Autómato Finito Determinístico Reactivo é um sêxtuplo  $A = (T, Q, I, F, R, \delta)$  em que:*

- $T$  é o abecedário, ou seja um conjunto de símbolos;
- $Q \neq \emptyset$  é um conjunto finito de estados;
- $I \subseteq Q, I \neq \emptyset$  é um conjunto de estados iniciais;
- $F \subseteq Q, F \neq \emptyset$  é um conjunto de estados finais;
- $R$  é um conjunto de acções semânticas;

---

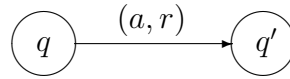
<sup>1</sup>(Versão 1.2)

- $\delta \subseteq (Q \times T) \times (Q \times R)$  é um conjunto de transições de estados.

sendo que:

- $|I| = 1$  (um só estado inicial);
- para todo o  $q \in Q$  e  $a \in T$ , existe no máximo um  $q' \in Q$  e  $r \in R$  tal que  $((q, a), (q', r)) \in \delta$  (uma só escolha possível);
- não existem transições- $\epsilon$ .

Há semelhança de um AFD pode-se representar graficamente um AFDR através de um grafo etiquetado:



o autômato  $D$  transita do estado  $q$  para o estado  $q'$  executando a ação semântica  $r$ , através do reconhecimento do símbolo  $a$ .

Pretende-se então fazer a análise léxica de um dado texto, isto é, reconhecer os símbolos básicos de uma linguagem e convertê-los para uma outra representação. Por exemplo, no caso de uma dada linguagem de programação ter-se-ia:

- reconhecer uma sequência de dígitos e convertê-la para a representação numérica apropriada.
- reconhecer certos símbolos identificando-os como pertencentes à classe de símbolos que não necessitam de tradução, por exemplo os identificadores reservados da linguagem de programação.

Os símbolos básicos, e a informação que lhes é associada pela tradução, são designados por lexemas, os símbolos básicos que não sofrem nenhuma alteração são designados por *símbolos literais*, aqueles que sofrem alguma tipo de transformação são designados por *símbolos não literais*.

Por exemplo:

$$'+' \mapsto \xrightarrow{\text{literal}} ('+', -)$$

$$''23'' \mapsto \xrightarrow{\text{não literal}} (\text{Inteiro}, 23)$$

Ao nível léxico está-se então a construir a tabela de símbolos com a informação de todos os elementos que ocorrem no texto, iremos tratar deste assunto com mais profundidade mais à frente.

## 6.2 Tradução Dirigida pela Sintaxe

Os autómatos reactivos são geralmente usados para a análise léxica do texto, a análise sintáctica pode ser feita através de uma tradução dirigida pela sintaxe, ou por uma tradução dirigida pela semântica, no que se segue vai-se tratar da primeira aproximação.

Na tradução dirigida pela sintaxe a tradução é feita através de reconhecedores do tipo descendente, ou ascendente, baseados em gramáticas independentes do contexto na qual se incluem acções semânticas. Esse tipo de gramáticas são designadas por *gramáticas tradutoras*. Com a utilização deste tipo de gramáticas efectua-se acções semânticas à medida que se faz o reconhecimento.

**Definição 6.3 (Gramáticas Tradutoras)** *Uma gramática tradutora é uma gramática independente do contexto  $G = (V_N, V_T, I, P)$ , em que  $V_T = T \cup A$  é constituído por um conjunto não vazio de símbolos do vocabulário  $T$ , e por um conjunto de acções semânticas  $A$ . Estes dois conjuntos são disjuntos,  $T \cap A = \emptyset$ .*

As gramáticas tradutoras são representadas introduzindo as acções semânticas nas produções. Por exemplo:

Seja  $G = (V_T, V_N, I, P)$  com:

$$\begin{aligned} V_T &= \{', \text{numero}, \text{palavra}\} \cup \{\text{inicio}, \text{fim}, \text{soma}\} \\ V_N &= \{I, \text{Lista}, \text{Elem}\} \\ P &= \left\{ \begin{array}{l} I \longrightarrow \{\text{inicio}\} \text{Lista} \{\text{fim}\} \\ \text{Lista} \longrightarrow \text{Elem} \mid \text{Elem} ', ' \text{Lista} \\ \text{Elem} \longrightarrow \text{numero} \{\text{soma}\} \mid \text{palavra} \end{array} \right\} \end{aligned}$$

Com as seguintes acções semânticas `inicio = somatorio<-0`, `fim = escrever(somatorio)`, e `soma = somatorio<-somatorio+valor(numero)`.

Para obter um tradutor para uma linguagem definida por uma gramática tradutora basta alterar os algoritmos já estudados de modo a incluírem o tratamento das acções semânticas. O *Bison* implementa um tradutor baseado num algoritmo de reconhecimento SLR.

Ver também o capítulo de (Crespo, 1998) dedicado a este tema.

Ver (Aho *et al.*, 1986) capítulo 5.