Alfabetos, linguagens e cadeias

Note-se a diferença conceitual que há entre alfabetos, linguagens e cadeias. Alfabetos são conjuntos, finitos e não-vazios, de símbolos, através de cuja concatenação são obtidas as **cadeias**. Linguagens, por sua vez, são conjuntos, finitos (eventualmente vazios) ou infinitos, de cadeias. Uma cadeia é também denominada **sentença** de uma linguagem, ou simplesmente sentença, no caso de ela pertencer à linguagem em questão. Linguagens são, portanto, coleções de sentenças sobre um dado alfabeto.

Símbolos, alfabeto, cadeias, linguagem

A Figura 1 ilustra a relação entre os conceitos de símbolo, alfabeto, cadeia e linguagem.

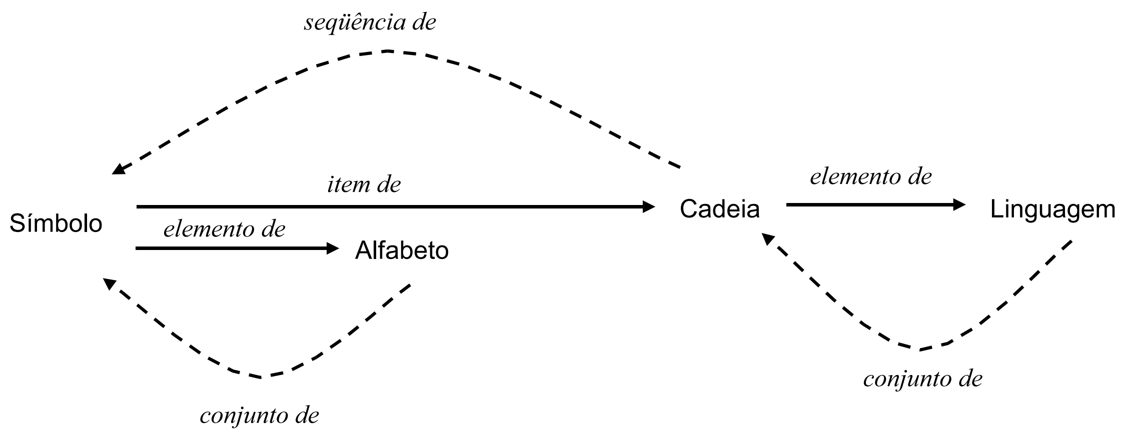
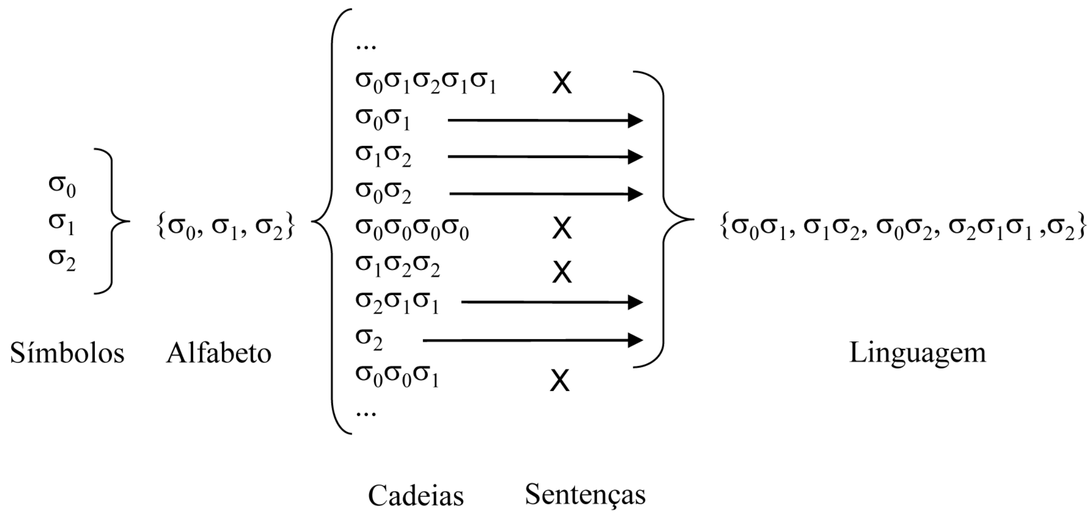


Figura 1: Símbolo, alfabeto, cadeia e linguagem

Símbolos, alfabeto, cadeias, linguagem

Outra maneira de associar significados aos termos “símbolo”, “alfabeto”, “cadeia” e “linguagem” é apresentada na Figura 2, que também ilustra o conceito de “sentença”.



Exemplo

Exemplo 2.1
O símbolo a é elemento do alfabeto $\{a\}$ e também um item da cadeia aaa, que por sua vez é elemento da linguagem $\{aaa\}$. Por outro lado, a linguagem $\{aaa\}$ é um conjunto que contém a cadeia aaa, a cadeia aaa é uma seqüência de símbolos a e o alfabeto $\{a\}$ contém o símbolo a. A Figura 3 ilustra esses conceitos, conforme a Figura 1.

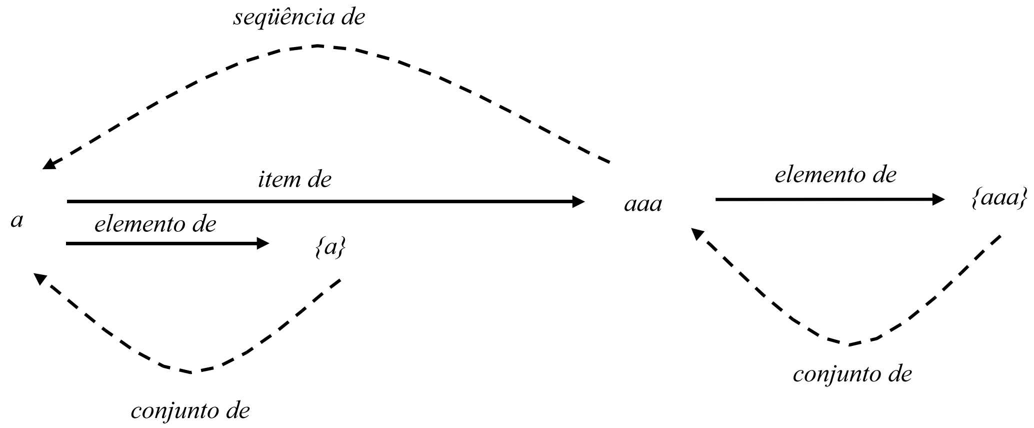


Figura 3: $a, \{a\}, aaa, \{aaa\}$

Exemplo

Exemplo 2.2
A Figura 4 ilustra uma aplicação dos conceitos da Figura 2 ao alfabeto $\{a, b\}$. A linguagem apresentada é, naturalmente, apenas uma das inúmeras que podem ser criadas a partir desse alfabeto.

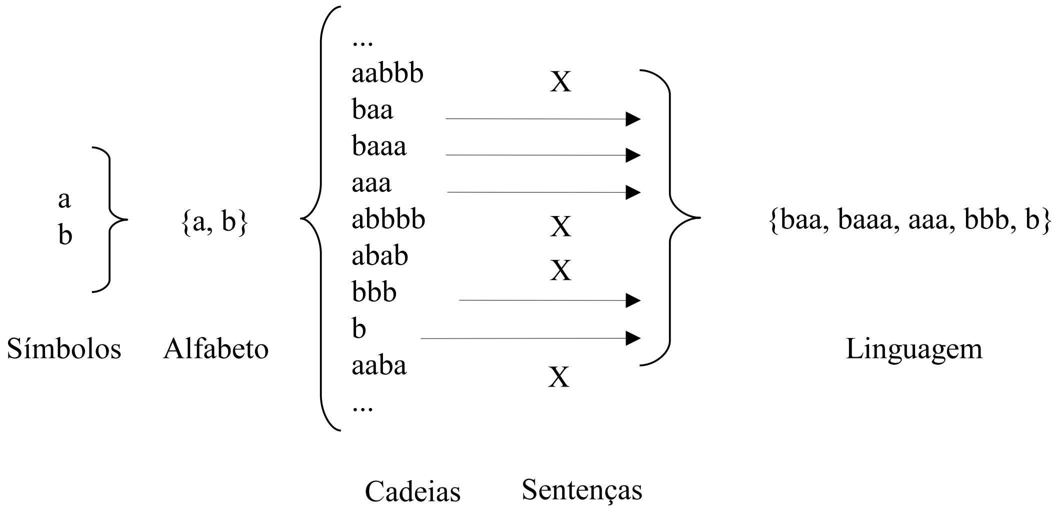


Figura 4: Símbolos a e b, cadeias, sentenças e linguagem

Fechamento reflexivo e transitivo

A definição de uma linguagem pode, portanto, ser formulada de maneira mais rigorosa com o auxílio da operação de fechamento reflexivo e transitivo: sendo uma linguagem qualquer coleção de cadeias sobre um determinado alfabeto Σ, e como Σ^* contém todas as possíveis cadeias sobre Σ, então toda e qualquer linguagem L sobre um alfabeto Σ sempre poderá ser definida como sendo um subconjunto de Σ^*, ou seja, $L \subseteq \Sigma^*$.

“Maior” linguagem

Diz-se que a **maior** linguagem que se pode definir sobre um alfabeto Σ, observando-se um conjunto qualquer P de propriedades, corresponde ao conjunto de **todas** as cadeias $w \in \Sigma^*$ tais que w satisfaz simultaneamente a **todas** as propriedades $p_i \in P$.

De uma forma geral, sempre que for feita uma referência a uma determinada linguagem L cujas cadeias satisfaçam a um certo conjunto de propriedades P, estará implícita (a menos de ressalva em contrário) a condição de que se trata da maior linguagem definida sobre L, cujas cadeias satisfaçam o conjunto de propriedades P.

“Maior” linguagem

Um caso particular importante a se considerar é a linguagem cujo conjunto P de propriedades seja o menos restritivo possível, considerando toda e qualquer cadeia de qualquer comprimento (finito) como sendo válida. Assim, a maior linguagem dentre todas as que podem ser definidas sobre um alfabeto Σ qualquer é $L = \Sigma^*$ (note-se, neste caso, que a única propriedade a ser satisfeita pelas cadeias de L é que elas “sejam definidas sobre Σ”, ou seja, obtidas a partir da simples justaposição de símbolos de Σ). Qualquer outra linguagem definida sobre esse mesmo alfabeto corresponderá obrigatoriamente a um subconjunto (eventualmente próprio) de Σ^*.

“Menor” linguagem

Por outro lado, a **menor** linguagem que pode ser definida sobre um alfabeto Σ qualquer é $\emptyset$, ou seja, a linguagem vazia ou a linguagem composta por zero sentenças.

Todas as linguagens

Finalmente, como o conjunto de todos os subconjuntos possíveis de serem obtidos a partir de Σ^* é 2^{Σ^*}, tem-se que 2^{Σ^*} representa o conjunto de **todas** as linguagens que podem ser definidas sobre o alfabeto Σ.

Menor, maior, todas

Em resumo:

- $\emptyset$ é o conjunto constituído por zero cadeias e corresponde à menor linguagem que se pode definir sobre um alfabeto Σ qualquer;
- Σ^* é o conjunto de todas as cadeias possíveis de serem construídas sobre Σ e corresponde à maior de todas as linguagens que pode ser definida sobre Σ;
- 2^{Σ^*} é o conjunto de todos os subconjuntos possíveis de serem obtidos a partir de Σ^*, e corresponde ao conjunto formado por todas as possíveis linguagens que podem ser definidas sobre Σ. Observe-se que $\emptyset \in 2^{\Sigma^*}$, e também que $\Sigma^* \in 2^{\Sigma^*}$.

Exemplos

Exemplo 2.4
Seja $\Sigma = \{a,b,c\}$ e P o conjunto formado pela única propriedade "todas as cadeias são iniciadas com o símbolo a". Então:

- A linguagem $L_0 = \emptyset$ é a menor linguagem que pode ser definida sobre Σ;
- A linguagem $L_1 = \{a, ab, ac, abc, acb\}$ é finita e observa P;
- A linguagem $L_2 = \{a\}\{a\}^*\{b\}^*\{c\}^*$ é infinita e observa P;
- A linguagem $L_3 = \{a\}\{a,b,c\}^*$ é infinita, observa P e, dentre todas as que observam P, trata-se da maior linguagem, pois não existe nenhuma outra cadeia em Σ^* que satisfaça a P e não pertença a L_3;
- $L_0 \subseteq \Sigma^*, L_1 \subseteq \Sigma^*, L_2 \subseteq \Sigma^*, L_3 \subseteq \Sigma^*$;
- $L_0 \in 2^{\Sigma^*}, L_1 \in 2^{\Sigma^*}, L_2 \in 2^{\Sigma^*}, L_3 \in 2^{\Sigma^*}$;
- Além de L_0, L_1, L_2 e L_3, existem inúmeras outras linguagens que podem ser definidas sobre Σ.