

# AUTÓMATOS E LINGUAGENS FORMAIS

(LCC/LMAT)

## 1. Linguagens Formais e Expressões Regulares

Departamento de Matemática

Universidade do Minho

2022/2023

## Definição

- Um **alfabeto** é um conjunto não vazio.  
Habitualmente, notaremos alfabetos por:  $A, A', A_0, \dots$
- Uma **letra** de uma alfabeto  $A$  é um elemento de  $A$ .  
Habitualmente, notaremos letras de alfabetos por:  $a, b, c, a', a_0, \dots$
- Uma **palavra** sobre um alfabeto  $A$  é uma sequência finita de letras de  $A$ , possivelmente vazia.  
Consequentemente, duas palavras são **iguais** quando as respectivas sequências de letras forem iguais.  
Habitualmente, notaremos palavras por:  $u, v, w, x, y, z, u', u_0, \dots$
- Chamaremos **palavra vazia** à sequência vazia de letras, que notaremos por  $\epsilon$ .
- A notação  $a_1 \dots a_n$ , com  $n \geq 1$ , representará uma **palavra não vazia**, cuja primeira letra é  $a_1$ , a segunda  $a_2$ , etc.

Por exemplo,

1, 01, 11, 1001, 00001,  $\epsilon$

são palavras sobre o alfabeto  $\{0, 1\}$ . Por outro lado,

*aba, ab, a, c, bbcac,  $\epsilon$*

são palavras sobre o alfabeto  $\{a, b, c\}$ .

## Definição

- $A^+$  notará o conjunto das palavras não vazias sobre o alfabeto  $A$ , i.e.,

$$A^+ = \{a_1 a_2 \cdots a_n \mid n \in \mathbb{N}, a_1, a_2, \dots, a_n \in A\}.$$

- $A^*$  notará o conjunto das palavras sobre o alfabeto  $A$ , i.e.,

$$A^* = \{\epsilon\} \cup A^+.$$

Por exemplo, sendo  $A = \{a, b\}$ , tem-se:

$$A^* = \{\epsilon, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, bab, bba, bbb, \dots\}.$$

Exercício: mostre que, para  $A$  finito, o conjunto  $A^*$  é numerável.

### Observação

Para provar propriedades sobre palavras é, por vezes, útil ter-se uma **definição indutiva** do conjunto  $A^*$ . Como se pode verificar,  $A^*$  é o conjunto  $X$  definido indutivamente pelas regras:

- (i)  $\epsilon \in X$ ;
  - (ii) Se  $u \in X$  e  $a \in A$ , então  $ua \in X$ ;
- uma vez convencionado que  $a_1 \dots a_n$  abrevia  $\epsilon a_1 \dots a_n$ .

### Observação

A anterior caracterização indutiva de  $A^*$  não só permite obter um **princípio de indução em palavras**, como permite também obter um **princípio de recursão em palavras**.

## Definição

O **comprimento** de uma palavra  $u$  é o comprimento da respetiva sequência de letras, sendo notado por  $|u|$ .

Por exemplo, fixando o alfabeto  $\{a, b, c\}$ :

$$|\epsilon| = 0, |a| = 1, |abc| = 3, |bbab| = 4.$$

## Observação

O comprimento de uma palavra sobre um alfabeto  $A$  corresponde a uma operação de  $A^*$  em  $\mathbb{N}_0$ , que pode ser caracterizada por **recursão em palavras** do seguinte modo:

**1**  $|\epsilon| = 0$ ;

**2**  $|ua| = |u| + 1$ , para todo  $u \in A^*$  e para todo  $a \in A$ .

Exercício: verifique as exemplificações acima utilizando esta caracterização recursiva de comprimento de uma palavra.

## Definição

O **número de ocorrências** de uma letra  $a$  numa palavra  $u$  é notado por  $|u|_a$ .

Por exemplo, fixando o alfabeto  $\{a, b, c\}$ :

$$|baba|_c = 0, \quad |acaba|_a = 3, \quad |acaba|_b = 1.$$

## Proposição

Sejam  $A$  um alfabeto e  $u \in A^*$ . Então:  $|u| = \sum_{a \in A} |u|_a$ .

Exercício: mostre a proposição, recorrendo a **indução em palavras** ou, em alternativa, a **indução no comprimento de palavras**.

## Definição

A **concatenação** de uma palavra  $u$  com uma palavra  $v$  será notada por  $u \cdot v$  ou, simplesmente, por  $uv$ , sendo dada pela concatenação das respectivas listas. Dito de outro modo:

- se  $u = \epsilon$ , então  $u \cdot v = v$ ;
- se  $u = a_1 \dots a_n$  e  $v = \epsilon$ , então  $u \cdot v = u$ ;
- se  $u = a_1 \dots a_n$  e  $v = b_1 \dots b_m$ , então  $u \cdot v = a_1 \dots a_n b_1 \dots b_m$ .

Por exemplo, para as palavras  $u = abc$  e  $v = aa$  (sobre  $\{a, b, c\}$ ),

$$u \cdot v = abcaa,$$

$$v \cdot u = aaabc,$$

$$u \cdot \epsilon = abc,$$

$$(v \cdot u) \cdot v = aaabcaa = v \cdot (u \cdot v)$$

Exercício: dê uma definição da operação de concatenação  $u \cdot v$  por **recursão na palavra  $v$** .

## Observação

A operação de concatenação de palavras é **associativa**, com **elemento neutro**  $\epsilon$ , pelo que, dado um alfabeto  $A$ ,

$$(A^*, \cdot) \text{ é um monóide,}$$

chamado o **monóide livre gerado por  $A$** .

No entanto, a concatenação de palavras **não é comutativa**.

## Proposição

Para  $u, v, w \in A^*$ , tem-se:

- 1**  $uv = uw \Rightarrow v = w$  (lei do corte à esquerda);
- 2**  $vu = wu \Rightarrow v = w$  (lei do corte à direita);
- 3**  $|uv| = |u| + |v|$  e  $|uv|_a = |u|_a + |v|_a$  (para todo  $a \in A$ ).



## Definição

Sejam  $u \in A^*$  e  $n \in \mathbb{N}_0$ . A **potência- $n$  de  $u$**  corresponderá à “concatenação de  $n$  cópias de  $u$ ”, sendo notada por  $u^n$  e definida recursivamente por:

$$u^n = \begin{cases} \epsilon & \text{se } n = 0 \\ u^{n-1}u & \text{se } n \geq 1. \end{cases}$$

## Proposição

Para toda a palavra  $u$  e para todo  $n, m \in \mathbb{N}_0$ ,

$$u^{n+m} = u^n u^m, \quad (u^n)^m = u^{nm}, \quad |u^n| = n|u|.$$

## Definição

Sejam  $u$  e  $v$  duas palavras de  $A^*$ . Diz-se que:

- $u$  é um **fator** de  $v$  quando existem  $x, y \in A^*$  tais que  $xuy = v$ ;
- $u$  é um **prefixo** de  $v$  quando existe  $y \in A^*$  tal que  $uy = v$ ;
- $u$  é um **sufixo** de  $v$  quando existe  $x \in A^*$  tal que  $xu = v$ .

Por exemplo, sendo  $v = baba$ :

- os fatores de  $v$  são  $\epsilon, b, a, ba, ab, bab, aba, v$ ;
- os prefixos de  $v$  são  $\epsilon, b, ba, bab, v$ ;
- os sufixos de  $v$  são  $\epsilon, a, ba, aba, v$ .

## Definição

A **palavra inversa** de uma palavra  $u \in A^*$  denota-se por  $u^I$  e define-se recursivamente por:

$$u^I = \begin{cases} \epsilon & \text{se } u = \epsilon \\ av^I & \text{se } u = va \quad \text{com } v \in A^* \text{ e } a \in A. \end{cases}$$

## Proposição

Para quaisquer  $a_1, a_2, \dots, a_n \in A$  e  $u, v \in A^*$ ,

$$(a_1 a_2 \cdots a_n)^I = a_n \cdots a_2 a_1,$$

$$(uv)^I = v^I u^I,$$

$$(u^I)^I = u.$$

## Definição

Uma **linguagem** sobre um alfabeto  $A$  é um subconjunto de  $A^*$ .  
Habitualmente, notaremos linguagens por:  $L, K, M, L', L_0, \dots$

São exemplos linguagens sobre  $A = \{a, b\}$ :

$$\emptyset, \{\epsilon\}, \{a\}, A, \{aa, aba, bbb, ababa\}, \{a^m b^n : m, n \in \mathbb{N}\}, A^+, A^*.$$

## Observação

- 1 O conjunto de todas as linguagens sobre o alfabeto  $A$  é  $\mathcal{P}(A^*) = \{L : L \subseteq A^*\}$ .
- 2 Para uma alfabeto finito  $A$ , dado que  $A^*$  é um conjunto infinito (numerável),  $\mathcal{P}(A^*)$  é um conjunto infinito **não numerável**.

## Exercício

Defina indutivamente as seguintes linguagens:

- a)  $L_0 = \{u \in \{0, 1\}^* : 1 \text{ é fator de } u\}$ ;
- b)  $L_1 = \{u \in \{a, b\}^* : |u|_a = |u|_b\}$ ;
- c)  $L_2 = \{u \in \{0, 1\}^* : |u|_0 \text{ é par}\}$ ;
- d)  $L_3 = \{w \in \{a, b, c\}^* : w = w^I\}$ .

Mostremos, por exemplo, que  $L_1$  é o conjunto  $L$  das palavras sobre o alfabeto  $A = \{a, b\}$  definido indutivamente pelas regras seguintes:

- 1  $\epsilon \in L$ ;
- 2 Se  $w \in L$ , então  $awb \in L$ ;
- 3 Se  $w \in L$ , então  $bwa \in L$ ;
- 4 Se  $w_1, w_2 \in L$ , então  $w_1 w_2 \in L$ .

Para provar a inclusão  $L \subseteq L_1$ , precisaremos do Princípio de indução estrutural associado a esta definição indutiva de  $L$ .

### Princípio de indução estrutural para $L$

Seja  $P(x)$  uma condição sobre  $x \in L$ .

Se:

- 1  $P(\epsilon)$ ;
  - 2 para qualquer  $w \in L$ , se  $P(w)$ , então  $P(awb)$ ;
  - 3 para qualquer  $w \in L$ , se  $P(w)$ , então  $P(bwa)$ ;
  - 4 para quaisquer  $w_1, w_2 \in L$ , se  $P(w_1)$  e  $P(w_2)$ , então  $P(w_1 w_2)$ ;
- então  $P(x)$ , para todo  $x \in L$ .

Mostremos por indução estrutural sobre  $L$  que, para cada  $x \in L$ ,  $|x|_a = |x|_b$ . Para  $x \in L$ , seja  $P(x)$  a condição:  $|x|_a = |x|_b$ .

**1** A propriedade  $P(\epsilon)$  é  $|\epsilon|_a = |\epsilon|_b$ . Ora, como  $|\epsilon|_a = 0 = |\epsilon|_b$ , tem-se  $P(\epsilon)$ .

**2** Seja  $w \in L$  e suponhamos  $P(w)$ , por hipótese de indução. Ou seja, suponhamos que:  $|w|_a = |w|_b$ . Então

$$|awb|_a = |w|_a + 1 = |w|_b + 1 = |awb|_b.$$

Provou-se assim  $P(awb)$ .

**3** Esta condição prova-se de forma análoga à anterior.

**4** Sejam  $w_1, w_2 \in L$  e suponhamos, por H.I.,  $P(w_1)$  e  $P(w_2)$ , isto é:  $|w_1|_a = |w_1|_b$  e  $|w_2|_a = |w_2|_b$ . Logo

$$|w_1 w_2|_a = |w_1|_a + |w_2|_a = |w_1|_b + |w_2|_b = |w_1 w_2|_b.$$

Provou-se assim  $P(w_1 w_2)$ .

Pelo Princípio de indução estrutural para  $L$ , de 1 a 4, conclui-se que  $P(x)$  é verdadeira para todo o  $x \in L$ . Provou-se assim que  $L \subseteq L_1$ .

Mostremos agora a inclusão  $L_1 \subseteq L$ , ou seja, mostremos que, para cada  $x \in L_1$ ,  $x \in L$ . Para  $x \in L_1$ , seja  $Q(x)$  a condição  $x \in L$ .

A prova será feita por **indução no comprimento da palavra  $x$** .

**1 Caso  $|x| = 0$ .** Neste caso  $x = \epsilon$ . Como, pela regra 1 da definição de  $L$ , se tem  $\epsilon \in L$ , segue  $Q(x)$ .

**2 Caso  $|x| > 0$ .** Suponhamos, por H.I., que  $Q(y)$  é verdadeira para todas as palavras  $y \in L_1$  tais que  $|y| < |x|$ . Existem 4 possibilidades:

■  $x = awb$  (ou  $x = bwa$ ) para algum  $w \in A^*$ . Dado que  $x \in L_1$ , tem-se  $|x|_a = |x|_b$ , donde  $|w|_a = |w|_b$ . Portanto  $w \in L_1$  e, pela H.I.,  $w \in L$ .

Daqui resulta, pela regra 2 da definição de  $L$ , que  $x \in L$ . Ou seja, tem-se  $Q(x)$ .

■  $x = awa$  (ou  $x = bwb$ ) para algum  $w \in A^*$ . Dado que  $x \in L_1$ , tem-se  $|x|_a = |x|_b$ . Logo  $x = aw_1w_2a$ , com  $aw_1, w_2a \in L_1$  (porquê?). Pela H.I., tem-se  $aw_1, w_2a \in L$ . Assim, pela regra 4 da definição de  $L$ ,  $x \in L$ . Ou seja, tem-se  $Q(x)$ .

De 1 e 2 segue  $Q(x)$  para cada  $x \in L_1$ , i.e.,  $L_1 \subseteq L$ .



## Definição

Dado que linguagens são conjuntos de palavras, ficam imediatamente definidas para linguagens as diversas operações sobre conjuntos. Em particular, dadas linguagens  $L$  e  $K$  sobre  $A$ :

- 1  $L \cup K = \{u \in A^* : u \in L \text{ ou } u \in K\}$  - **união** de  $L$  e  $K$ .
- 2  $L \cap K = \{u \in A^* : u \in L \text{ e } u \in K\}$  - **interseção** de  $L$  e  $K$ .
- 3  $K \setminus L = \{u \in A^* : u \in K \text{ e } u \notin L\}$  - **complementar** de  $L$  em  $K$ .
- 4  $\bar{L} = A^* \setminus L = \{u \in A^* : u \notin L\}$  - **complementar** de  $L$ .

*E.g.*, para  $A = \{0,1\}$ ,  $L = \{u \in A^* : |u|_0 = 0\}$  e  $K = \{u \in A^* : |u|_1 = 0\}$ :  
 $L \cup K = A^*$ ;  $L \cap K = \emptyset$ ;  $L \setminus K = L$ ;  $\bar{L} = K$ .

## Observação

As anteriores operações em linguagens herdam, de imediato, as propriedades das respetivas operações em conjuntos. Por exemplo: a **união** e a **interseção** de linguagens são operações **associativas**, **comutativas**, que possuem **elemento neutro** e **elemento absorvente**.

## Definição

Dadas linguagens  $L$  e  $K$  sobre  $A$ , a **concatenação** de  $L$  e  $K$  é a linguagem em  $A$ , notada por  $L.K$  ou  $LK$ , dada por:

$$\{u.v : u \in L \text{ e } v \in K\}.$$

Por exemplo, para  $L = \{b, ba\}$  e  $K = \{\epsilon, a\}$ :

$$L.K = \{b, ba, baa\}$$

$$K.L = \{b, ba, ab, aba\}$$

Deste exemplo, pode deduzir-se que a concatenação de linguagens **não é comutativa**.

## Proposição

A concatenação de linguagens é uma operação **associativa**, com **elemento neutro**  $\{\epsilon\}$  e com a **linguagem vazia** como **elemento absorvente**, que **distribui pela união**, i.e., para  $L, K, M \subseteq A^*$ :

$$L.(K \cup M) = (L.K) \cup (L.M) \quad (K \cup M).L = (K.L) \cup (M.L)$$

## Observação

Adiante, uma palavra  $u$  será utilizada muitas vezes como uma abreviatura para  $\{u\}$  (a linguagem cuja única palavra é  $u$ ).

Por exemplo, considerando o alfabeto  $A = \{a, b\}$ ,

$$ab\{aa, bb\} = \{ab\}\{aa, bb\} = \{abaa, abbb\}.$$

## Proposição

Dada uma palavra  $u$  sobre um alfabeto  $A$ :

$$\begin{aligned} uA^* &= \{ux : x \in A^*\}, \\ A^*u &= \{xu : x \in A^*\}, \\ A^*uA^* &= \{xuy : x, y \in A^*\} \end{aligned}$$

são as linguagens cujas palavras têm, respetivamente,  $u$  como **prefixo**, como **sufixo** e como **fator**.

## Definição

Dada uma linguagem  $L$ , a sua **linguagem inversa** será notada por  $L^I$ , sendo dada por:

$$L^I = \{u^I : u \in L\}.$$

Por exemplo, considerando o alfabeto  $\{0, 1\}$ :

- 1 para  $L = \{\epsilon, 0, 01\}$ ,  $L^I = \{\epsilon, 0, 10\}$ ;
- 2 para  $K = \{0^n 1^m : n \in \mathbb{N}_0 \wedge m \in \mathbb{N}\}$ ,  $K^I = \{1^m 0^n : n \in \mathbb{N}_0 \wedge m \in \mathbb{N}\}$ .

## Proposição

Dadas linguagens  $L$  e  $K$ :

- 1  $(L \cup K)^I = L^I \cup K^I$ ;
- 2  $(L.K)^I = K^I.L^I$ ;

## Definição

Sejam  $L$  uma linguagem num alfabeto  $A$  e  $n \in \mathbb{N}_0$ . A **potência  $n$  de  $L$**  é a linguagem em  $A$ , notada por  $L^n$ , dada recursivamente por:

$$\begin{aligned} L^0 &= \{\epsilon\} \\ L^{k+1} &= L^k.L \quad (k \in \mathbb{N}_0) \end{aligned}$$

Por exemplo, para a linguagem  $L = \{a, ab\}$  (sobre o alfabeto  $A = \{a, b\}$ ):

$$\begin{aligned} L^0 &= \{\epsilon\} \\ L^1 &= L^0.L = \{\epsilon\}.\{a, ab\} = \{a, ab\} \\ L^2 &= L^1.L = \{a, ab\}.\{a, ab\} = \{aa, aab, aba, abab\} \end{aligned}$$

## Proposição

Sejam  $u$  uma palavra,  $L$  uma linguagem num alfabeto  $A$  e  $n \in \mathbb{N}_0$ . Então,  $u \in L^n$  se e só se

- 1 (i)  $n = 0$  e  $u = \epsilon$ ; ou
- 2 (ii)  $n \geq 1$  e existem palavras  $u_1, \dots, u_n \in L$  t.q.  $u = u_1 \dots u_n$ .

## Definição

Dada uma linguagem  $L$  num alfabeto  $A$ :

- 1 o fecho positivo de  $L$  é a linguagem em  $A$ , notada por  $L^+$ , dada por:

$$L^+ = \bigcup_{n \geq 1} L^n$$

- 2 o fecho (de Kleene) de  $L$ , também designada a estrela de  $L$ , é a linguagem em  $A$ , notada por  $L^*$ , dada por:

$$L^* = \bigcup_{n \geq 0} L^n = L^+ \cup \{\epsilon\}.$$

Por exemplo, para o alfabeto  $A = \{0, 1\}$  e para a linguagem  $L = \{0, 1\}$  sobre  $A$ :

- 1 para todo  $n \in \mathbb{N}_0$ ,  $L^n = \{u \in A^* : |u| = n\}$  (exercício);
- 2  $L^+ = A^+$  (porquê?);
- 3  $L^* = A^*$  (porquê?).

De facto, os três itens anteriores são válidos para **qualquer** alfabeto  $A$ .

## Proposição

Seja  $L$  uma linguagem. Então,

- 1  $\emptyset^* = \{\epsilon\}$ ,  $\emptyset^+ = \emptyset$ ,  $\{\epsilon\}^* = \{\epsilon\} = \{\epsilon\}^+$ ;
- 2  $L = L^1 \subseteq L^+ \subseteq L^+ \cup \{\epsilon\} = L^*$ ;
- 3  $\epsilon \in L^+$  se e só se  $\epsilon \in L$ ;
- 4  $L^+ = LL^* = L^*L$ .

## Definição

O conjunto das **expressões regulares** sobre um alfabeto  $A$  é o conjunto  $ER(A)$ , de palavras sobre o alfabeto  $A \cup \{\emptyset, \epsilon, (, ), +, \cdot, *\}$ , definido indutivamente por:

- 1  $\emptyset \in ER(A)$  e  $\epsilon \in ER(A)$ ;
- 2  $a \in ER(A)$ , para cada  $a \in A$ ;
- 3 Se  $r, s \in ER(A)$ , então  $(r + s) \in ER(A)$ ;
- 4 Se  $r, s \in ER(A)$ , então  $(r \cdot s) \in ER(A)$ ;
- 5 Se  $r \in ER(A)$ , então  $(r^*) \in ER(A)$ .

Habitualmente, notaremos expressões regulares por:  $r, s, r', r_0, \dots$

Por exemplo, sendo  $A = \{a, b\}$ , são expressões regulares sobre  $A$ :

$$\emptyset, \epsilon, a, b, (a + b), ((b \cdot \emptyset) + \epsilon), ((a + b)^*), ((a^*) + (b^*)).$$



## Observação

A notação das expressões regulares pode ser abreviada da seguinte forma:

- o símbolo  $\cdot$  pode ser omitido;
- podem omitir-se parênteses desnecessários usando associatividade para as operações  $+$  e  $\cdot$ , e considerando que  $*$  tem a maior prioridade e que  $\cdot$  tem prioridade em relação a  $+$ ;
- para  $r \in ER(A)$  e  $n \in \mathbb{N}_0$ , a abreviatura  $r^n$  é definida recursivamente, por:
  - $r^0 = \epsilon$ ,  $r^1 = r$  e, para  $n \geq 2$ ,  $r^n = (r^{n-1} \cdot r)$ ;
  - $r^+ = (r \cdot (r^*))$ .

Por exemplo,

- $a^+b^2$  é uma abreviatura de  $((a \cdot (a^*)) \cdot (b \cdot b))$ ;
- $\emptyset^*a + (b + \epsilon)b$  abrevia  $((\emptyset^*) \cdot a) + (((b + \epsilon) \cdot b))$ .

## Definição

A cada expressão regular  $r$  sobre um alfabeto  $A$  associa-se uma linguagem  $\mathcal{L}(r)$  sobre  $A$ , dita a **linguagem representada por  $r$**  ou a **linguagem de  $r$** . A função

$$\begin{aligned}\mathcal{L} : ER(A) &\rightarrow \mathcal{P}(A^*) \\ r &\mapsto \mathcal{L}(r)\end{aligned}$$

é definida recursivamente por:

- 1  $\mathcal{L}(\emptyset) = \emptyset$  e  $\mathcal{L}(\epsilon) = \{\epsilon\}$ ;
- 2  $\mathcal{L}(a) = \{a\}$ , para cada  $a \in A$ ;
- 3  $\mathcal{L}((r + s)) = \mathcal{L}(r) \cup \mathcal{L}(s)$ , para quaisquer  $r, s \in ER(A)$ ;
- 4  $\mathcal{L}((r \cdot s)) = \mathcal{L}(r) \cdot \mathcal{L}(s)$ , para quaisquer  $r, s \in ER(A)$ ;
- 5  $\mathcal{L}((r)^*) = \mathcal{L}(r)^*$ , para cada  $r \in ER(A)$ .

Por exemplo, sendo  $A = \{a, b\}$ , tem-se:

$$1 \quad \mathcal{L}((b + \epsilon)a) = \mathcal{L}(b + \epsilon)\mathcal{L}(a) = (\mathcal{L}(b) \cup \mathcal{L}(\epsilon))\{a\} = (\{b\} \cup \{\epsilon\})\{a\} = \{b, \epsilon\}\{a\} = \{ba, a\};$$

$$2 \quad \mathcal{L}(a^*) = \mathcal{L}(a)^* = \{a\}^* = \{a^n : n \in \mathbb{N}_0\};$$

$$3 \quad \mathcal{L}(a^*(a^3 + b)) = \{a\}^*\{a^3, b\} = \{a^m : m \geq 3\} \cup \{a^n b : n \in \mathbb{N}_0\};$$

$$4 \quad \mathcal{L}(a^* + b^*) = \mathcal{L}(a)^* \cup \mathcal{L}(b)^* = \{a\}^* \cup \{b\}^* = \{a^n : n \in \mathbb{N}_0\} \cup \{b^n : n \in \mathbb{N}_0\};$$

$$5 \quad \mathcal{L}((a + b)^*) = \mathcal{L}(a + b)^* = (\{a\} \cup \{b\})^* = \{a, b\}^* = A^*, \text{ ou seja, o conjunto de todas as palavras sobre o alfabeto } A;$$

$$6 \quad \mathcal{L}((a + b)^* aba(a + b)^*) = A^* aba A^* \text{ é a linguagem das palavras que têm } aba \text{ como fator.}$$

## Definição

- 1 Uma **linguagem**  $L$  sobre um alfabeto  $A$  diz-se **regular** quando pode ser representada por alguma expressão regular sobre  $A$ , ou seja:

$$\exists r \in ER(A). L = \mathcal{L}(r).$$

- 2 O **conjunto das linguagens regulares** sobre um alfabeto  $A$  será denotado por  $Reg(A)$ .

Por exemplo, **todas as linguagens do slide anterior** são regulares (dado serem iguais a  $\mathcal{L}(r)$ , para cada uma das expressões regulares consideradas).

São também exemplos de linguagens regulares  $\emptyset$  e  $\{\epsilon\}$  já que

$$\emptyset = \mathcal{L}(\emptyset) \quad \text{e} \quad \{\epsilon\} = \mathcal{L}(\epsilon).$$

## Observação

Alternativamente, o conjunto  $Reg(A)$  das linguagens regulares sobre  $A$  pode ser definido indutivamente por:

- 1  $\emptyset, \{\epsilon\} \in Reg(A)$ ;
- 2  $\{a\} \in Reg(A)$ , para todo  $a \in A$ ;
- 3  $Reg(A)$  é fechado para as operações de **união**, **concatenação** e **fecho de Kleene**, ou seja,  
se  $L, K \in Reg(A)$  então  $L \cup K, L \cdot K, L^* \in Reg(A)$ .

## Observação

Note-se que, se  $L$  é uma linguagem regular sobre um alfabeto  $A$ , então  $L^+ = L^*L$  também é uma linguagem regular sobre  $A$ .

## Observação

- 1 Para um alfabeto **finito**  $A = \{a_1, \dots, a_n\}$ :
  - a linguagem  $A$  é regular:  $A = \{a_1\} \cup \dots \cup \{a_n\} = \mathcal{L}(a_1 + \dots + a_n)$ ;
  - a linguagem  $A^*$  é regular:  $A^* = \{a_1, \dots, a_n\}^* = \mathcal{L}((a_1 + \dots + a_n)^*)$
- 2 Para qualquer palavra  $u$  sobre um alfabeto,  $\{u\}$  é uma linguagem regular sobre  $A$ . De facto:
  - caso  $u = \epsilon$ :  $\{u\} = \{\epsilon\} = \mathcal{L}(\epsilon)$ ;
  - caso  $u = a_1 a_2 \dots a_n$ , com  $a_1, a_2, \dots, a_n \in A$ :

$$\{u\} = \{a_1\}\{a_2\} \dots \{a_n\} = \mathcal{L}(a_1 a_2 \dots a_n) = \mathcal{L}(u).$$

- 3 Toda a **linguagem finita**  $L$ , sobre um alfabeto  $A$ , é regular. De facto:
  - caso  $L = \emptyset$ :  $L = \emptyset = \mathcal{L}(\emptyset)$ ;
  - caso  $L = \{u_1, \dots, u_k\}$ , com  $k \geq 1$ ,  $u_i \in A^*$ :

$$L = \{u_1\} \cup \dots \cup \{u_k\} = \mathcal{L}(u_1 + \dots + u_k).$$

## Proposição

Existem linguagens que **não** são regulares.

A proposição segue por razões de cardinalidade. De facto:

- por um lado, o conjunto  $\mathcal{P}(A^*)$  das linguagens sobre um alfabeto  $A$  finito, com duas ou mais letras, é **infinito não numerável**;
- por outro, o conjunto  $ER(A)$  das expressões regulares sobre um alfabeto  $A$  finito é **numerável**.

Assim, há linguagens que não poderão corresponder à representação de qualquer expressão regular.

## Observação

Prova-se que **não** são regulares as linguagens (sobre  $A = \{0, 1\}$ ):

- 1  $\{0^p : p > 0 \text{ primo}\}$
- 2  $\{0^n 1^n : n \in \mathbb{N}_0\}$
- 3  $\{u \in A^* : u^I = u\}$

Note-se que expressões regulares distintas podem representar a mesma linguagem.

Por exemplo, as expressões regulares

$$(a + b)^* \quad \text{e} \quad (a + b)(a + b)^* + \epsilon$$

representam a mesma linguagem, nomeadamente  $\{a, b\}^*$ .

### Definição

- Diremos que  $r$  é menor ou igual que  $s$ , escrevendo,  $r \leq s$ , quando  $\mathcal{L}(r) \subseteq \mathcal{L}(s)$ .
- Diremos que  $r$  é equivalente a  $s$  ou, simplesmente, que  $r$  é igual a  $s$ , escrevendo, respetivamente,  $r \equiv s$  e  $r = s$ , quando  $r \leq s$  e  $s \leq r$ , ou seja, quando  $\mathcal{L}(r) = \mathcal{L}(s)$ .

### Observação

Dado um alfabeto  $A$ ,  $(ER(A), \leq)$  constitui um conjunto parcialmente ordenado. (Porquê?)



Por exemplo, considerando o alfabeto  $A = \{a, b\}$ , pode escrever-se:

- 1  $a \leq a + b$ , mas  $a + b \not\leq a$ , pelo que  $a \neq a + b$ ;
- 2  $a + b \leq (a + b)^+ \leq (a + b)^*$  e consequentemente  $a + b \leq (a + b)^*$  (porquê?);
- 3  $(a + b)^* aa(a + b)^* \leq (a + b)^* a(a + b)^*$  (porquê?);
- 4  $(a + b)^* \leq (a + b)(a + b)^* + \epsilon$ ;
- 5  $(a + b)(a + b)^* + \epsilon \leq (a + b)^*$ ;
- 6  $(a + b)^* = (a + b)(a + b)^* + \epsilon$ ;

## Observação

Adiante, por norma, dadas expressões regulares  $r, s \in ER(A)$ , a notação  $r = s$  significará que  $r$  é **equivalente** a  $s$ , ou seja,  $\mathcal{L}(r) = \mathcal{L}(s)$  (e não que as palavras  $r$  e  $s$ , sobre o alfabeto  $A \cup \{\emptyset, \epsilon, (, ), +, \cdot, *\}$ , são sequências de letras de iguais).

## Proposição

Sejam  $r$ ,  $s$  e  $t$  expressões regulares sobre um alfabeto  $A$ . Então,

$$(i) \quad (r + s) + t = r + (s + t);$$

$$(ii) \quad r + \emptyset = \emptyset + r = r;$$

$$(iii) \quad r + s = s + r;$$

$$(iv) \quad r + r = r;$$

$$(v) \quad r\emptyset = \emptyset r = \emptyset;$$

$$(vi) \quad r\epsilon = \epsilon r = r;$$

$$(vii) \quad (rs)t = r(st);$$

$$(viii) \quad r(s + t) = rs + rt;$$

$$(ix) \quad (r + s)t = rt + st;$$

$$(x) \quad \emptyset^* = \epsilon^* = \epsilon;$$

$$(xi) \quad r^* = r^*r^* = (r^*)^* = (\epsilon + r)^* = r^+ + \epsilon;$$

$$(xii) \quad r^+ = rr^* = r^*r;$$

$$(xiii) \quad (r+s)^* = (r^*+s^*)^* = (r^*s^*)^* = (r^*s)^*r^*;$$

$$(xiv) \quad r(sr)^* = (rs)^*r;$$

$$(xv) \quad (r^*s)^* = (r + s)^*s + \emptyset;$$

$$(xvi) \quad (rs^*)^* = r(r + s)^* + \epsilon.$$

## Definição

Uma **equação linear à direita** sobre expressões regulares é uma equação do tipo

$$X = rX + s$$

na qual  $r, s \in ER(A)$  são expressões regulares e  $X$  é dita a (expressão regular) **indeterminada** ou **incógnita**.

Habitualmente, usaremos  $X, Y, X_1, \dots$  para representar expressões regulares indeterminadas

Exemplos, de tais equações lineares são:

$$X = aX + a$$

$$Y = (a + b^*)Y + (a + \epsilon)$$

## Definição

Diz-se que uma expressão regular  $t \in ER(A)$  é uma **solução** da equação  $X = rX + s$  quando  $t = rt + s$ .

Por exemplo, uma solução da equação

$$X = aX + a$$

é  $a^+$ , dado que se tem  $a^+ = aa^+ + a$ , uma vez que:

$$\mathcal{L}(a^+) = \{a^n : n \in \mathbb{N}\} = \{a^n : n \geq 2\} \cup \{a\} = \mathcal{L}(aa^+) \cup \mathcal{L}(a) = \mathcal{L}(aa^+ + a).$$

$(a + b)^*$  é uma solução da equação  $Y = (a + b^*)Y + (a + \epsilon)$ .  
(Porquê?)