

Stefan Sokołowski

JĘZYKI FORMALNE I METODY KOMPILACJI

wykład 2

Inst. Informatyki Stosowanej, ANS

Elbląg, 2024/2025

Gramatyki prawoliniowe

DEFINICJA:

Gramatyka bezkontekstowa $G = \langle \Sigma, N, P, S \rangle$ jest *prawostronnie liniowa* jeśli ...

Gramatyki prawoliniowe

DEFINICJA:

Gramatyka bezkontekstowa $G = \langle \Sigma, N, P, S \rangle$ jest *prawostronnie liniowa* jeśli ...

praw. liniowe

bezkontekstowe

Gramatyki prawoliniowe

DEFINICJA:

Gramatyka bezkontekstowa $G = \langle \Sigma, N, P, S \rangle$ jest *prawostronnie liniowa* jeśli każda produkcja w P ma postać ...

praw. liniowe

bezkontekstowe

Gramatyki prawoliniowe

DEFINICJA:

Gramatyka bezkontekstowa $G = \langle \Sigma, N, P, S \rangle$ jest *prawostronnie liniowa* jeśli każda produkcja w P ma postać:

- albo $A \rightarrow x$ (czyli: żadnych nieterminali)

gdzie $x \in \Sigma^*$ a $A \in N$.

praw. liniowe

bezkontekstowe

Gramatyki prawoliniowe

DEFINICJA:

Gramatyka bezkontekstowa $G = \langle \Sigma, N, P, S \rangle$ jest *prawostronnie liniowa* jeśli każda produkcja w P ma postać:

- albo $A \rightarrow x$ (czyli: żadnych nieterminali)
- albo $A \rightarrow xB$ (czyli: tylko jeden nieterminal po prawej)

gdzie $x \in \Sigma^*$ a $A, B \in N$.

praw. liniowe

bezkontekstowe

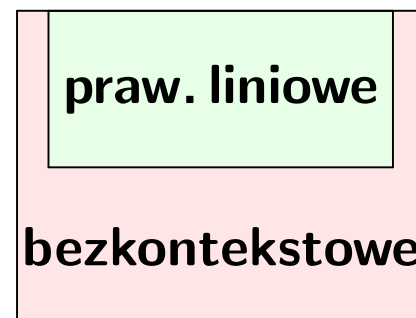
Gramatyki prawoliniowe

DEFINICJA:

Gramatyka bezkontekstowa $G = \langle \Sigma, N, P, S \rangle$ jest *prawostronnie liniowa* jeśli każda produkcja w P ma postać:

- albo $A \rightarrow x$ (czyli: żadnych nieterminali)
- albo $A \rightarrow xB$ (czyli: tylko jeden nieterminal po prawej)

gdzie $x \in \Sigma^*$ a $A, B \in N$.



PRZYKŁAD:

$$E \rightarrow aC$$

$$E \rightarrow bC$$

$$C \rightarrow \lambda$$

$$C \rightarrow aC$$

$$C \rightarrow bC$$

$$C \rightarrow 0C$$

$$C \rightarrow 1C$$

prawoliniowa

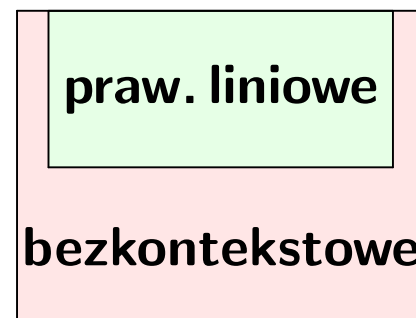
Gramatyki prawoliniowe

DEFINICJA:

Gramatyka bezkontekstowa $G = \langle \Sigma, N, P, S \rangle$ jest *prawostronnie liniowa* jeśli każda produkcja w P ma postać:

- albo $A \rightarrow x$ (czyli: żadnych nieterminali)
- albo $A \rightarrow xB$ (czyli: tylko jeden nieterminal po prawej)

gdzie $x \in \Sigma^*$ a $A, B \in N$.



PRZYKŁAD:

$$\begin{array}{lcl} E & \rightarrow & aC \\ E & \rightarrow & bC \\ C & \rightarrow & \lambda \\ C & \rightarrow & aC \\ C & \rightarrow & bC \\ C & \rightarrow & 0C \\ C & \rightarrow & 1C \end{array}$$

prawoliniowa

$$\begin{array}{lcl} E & \rightarrow & a \\ E & \rightarrow & (E) \\ E & \rightarrow & E + E \end{array}$$

bezkontekstowa
ale nie prawoliniowa

Rodzaje gramatyk

PRZYKŁAD:

$$G_1: \begin{array}{lcl} E & \rightarrow & aC \\ E & \rightarrow & bC \\ C & \rightarrow & \lambda \\ C & \rightarrow & aC \\ C & \rightarrow & bC \\ C & \rightarrow & 0C \\ C & \rightarrow & 1C \end{array}$$

Rodzaje gramatyk

PRZYKŁAD:

$$G_1: \begin{array}{lcl} E & \rightarrow & aC \\ E & \rightarrow & bC \\ C & \rightarrow & \lambda \\ C & \rightarrow & aC \\ C & \rightarrow & bC \\ C & \rightarrow & 0C \\ C & \rightarrow & 1C \end{array}$$

$L(G_1)$ = język identyfikatorów

Rodzaje gramatyk

PRZYKŁAD:

$$G_1: \begin{array}{lcl} E & \rightarrow & aC \\ E & \rightarrow & bC \\ C & \rightarrow & \lambda \\ C & \rightarrow & aC \\ C & \rightarrow & bC \\ C & \rightarrow & 0C \\ C & \rightarrow & 1C \end{array}$$

$$G_2: \begin{array}{lcl} I & \rightarrow & a \\ I & \rightarrow & b \\ I & \rightarrow & Ia \\ I & \rightarrow & Ib \\ I & \rightarrow & I0 \\ I & \rightarrow & I1 \end{array}$$

$L(G_1)$ = język identyfikatorów

Rodzaje gramatyk

PRZYKŁAD:

$$G_1:$$

E	$\rightarrow$	aC
E	$\rightarrow$	bC
C	$\rightarrow$	λ
C	$\rightarrow$	aC
C	$\rightarrow$	bC
C	$\rightarrow$	$0C$
C	$\rightarrow$	$1C$

$$G_2:$$

I	$\rightarrow$	a
I	$\rightarrow$	b
I	$\rightarrow$	Ia
I	$\rightarrow$	Ib
I	$\rightarrow$	$I0$
I	$\rightarrow$	$I1$

$$L(G_1) = \text{język identyfikatorów} = L(G_2)$$

Rodzaje gramatyk

PRZYKŁAD:

$$G_1: \begin{array}{lcl} E & \rightarrow & aC \\ E & \rightarrow & bC \\ C & \rightarrow & \lambda \\ C & \rightarrow & aC \\ C & \rightarrow & bC \\ C & \rightarrow & 0C \\ C & \rightarrow & 1C \end{array}$$

$$G_2: \begin{array}{lcl} I & \rightarrow & a \\ I & \rightarrow & b \\ I & \rightarrow & Ia \\ I & \rightarrow & Ib \\ I & \rightarrow & I0 \\ I & \rightarrow & I1 \end{array}$$

$$L(G_1) = \text{język identyfikatorów} = L(G_2)$$

G_1 jest prawostronnie liniowa

Rodzaje gramatyk

PRZYKŁAD:

$$G_1:$$

E	$\rightarrow$	aC
E	$\rightarrow$	bC
C	$\rightarrow$	λ
C	$\rightarrow$	aC
C	$\rightarrow$	bC
C	$\rightarrow$	$0C$
C	$\rightarrow$	$1C$

$$G_2:$$

I	$\rightarrow$	a
I	$\rightarrow$	b
I	$\rightarrow$	Ia
I	$\rightarrow$	Ib
I	$\rightarrow$	$I0$
I	$\rightarrow$	$I1$

$$L(G_1) = \text{język identyfikatorów} = L(G_2)$$

G_1 jest prawostronnie liniowa

G_2 jest bezkontekstowa

ale nie prawostronnie liniowa

Rodzaje gramatyk

PRZYKŁAD:

G_1 :

E	$\rightarrow$	aC
E	$\rightarrow$	bC
C	$\rightarrow$	λ
C	$\rightarrow$	aC
C	$\rightarrow$	bC
C	$\rightarrow$	$0C$
C	$\rightarrow$	$1C$

G_2 :

I	$\rightarrow$	a
I	$\rightarrow$	b
I	$\rightarrow$	Ia
I	$\rightarrow$	Ib
I	$\rightarrow$	$I0$
I	$\rightarrow$	$I1$

$L(G_1) = \text{język identyfikatorów} = L(G_2)$

G_1 jest prawostronnie liniowa	}	dwie gramatyki różnych typów mogą określać ten sam język
G_2 jest bezkontekstowa		
ale nie prawostronnie liniowa		

Rodzaje gramatyk

PRZYKŁAD:

G_1 :

E	$\rightarrow$	aC
E	$\rightarrow$	bC
C	$\rightarrow$	λ
C	$\rightarrow$	aC
C	$\rightarrow$	bC
C	$\rightarrow$	$0C$
C	$\rightarrow$	$1C$

G_2 :

I	$\rightarrow$	a
I	$\rightarrow$	b
I	$\rightarrow$	Ia
I	$\rightarrow$	Ib
I	$\rightarrow$	$I0$
I	$\rightarrow$	$I1$

$L(G_1) = \text{język identyfikatorów} = L(G_2)$

G_1 jest prawostronnie liniowa G_2 jest bezkontekstowa ale nie prawostronnie liniowa	$\left\{ \begin{array}{l} \text{dwie gramatyki } \textbf{różnych typów} \\ \text{mogą określać ten sam język} \end{array} \right.$
--	--

DEFINICJA:

Język L jest

- *regularny* jeśli ist. gram. **prawostronnie liniowa** G taka, że $L = L(G)$

Rodzaje gramatyk

PRZYKŁAD:

G_1 :

E	$\rightarrow$	aC
E	$\rightarrow$	bC
C	$\rightarrow$	λ
C	$\rightarrow$	aC
C	$\rightarrow$	bC
C	$\rightarrow$	$0C$
C	$\rightarrow$	$1C$

G_2 :

I	$\rightarrow$	a
I	$\rightarrow$	b
I	$\rightarrow$	Ia
I	$\rightarrow$	Ib
I	$\rightarrow$	$I0$
I	$\rightarrow$	$I1$

$L(G_1) = \text{język identyfikatorów} = L(G_2)$

G_1 jest prawostronnie liniowa
 G_2 jest bezkontekstowa
 ale nie prawostronnie liniowa

} dwie gramatyki **różnych typów** mogą określać ten sam język

DEFINICJA:

Język L jest

- *regularny* jeśli ist. gram. **prawostronnie liniowa** G taka, że $L = L(G)$;
- *bezkontekstowy* jeśli ist. gram. **bezkontekstowa** G taka, że $L = L(G)$.

Rodzaje gramatyk

PRZYKŁAD:

G_1 :

E	$\rightarrow$	aC
E	$\rightarrow$	bC
C	$\rightarrow$	λ
C	$\rightarrow$	aC
C	$\rightarrow$	bC
C	$\rightarrow$	$0C$
C	$\rightarrow$	$1C$

G_2 :

I	$\rightarrow$	a
I	$\rightarrow$	b
I	$\rightarrow$	Ia
I	$\rightarrow$	Ib
I	$\rightarrow$	$I0$
I	$\rightarrow$	$I1$

$$L(G_1) = \text{język identyfikatorów} = L(G_2)$$

G_1 jest prawostronnie liniowa
 G_2 jest bezkontekstowa
 ale nie prawostronnie liniowa

} dwie gramatyki **różnych typów**
 mogą określać ten sam język

DEFINICJA:

Język L jest

- *regularny* jeśli ist. gram. **prawostronnie liniowa** G taka, że $L = L(G)$;
- *bezkontekstowy* jeśli ist. gram. **bezkontekstowa** G taka, że $L = L(G)$.

PRZYKŁAD: Język identyfikatorów $L(G_2)$ jest regularny.

Rodzaje gramatyk

ZADANIE:

$$G_1: \begin{array}{lcl} A & \rightarrow & b \\ A & \rightarrow & c \\ A & \rightarrow & AA \end{array}$$

← gramatyka **nie**prawoliniowa

Czy $L(G_1)$ jest regularny?

Rodzaje gramatyk

ZADANIE:

$$G_1: \begin{array}{l} A \rightarrow b \\ A \rightarrow c \\ A \rightarrow AA \end{array}$$

← gramatyka **nie**prawoliniowa

Czy $L(G_1)$ jest regularny?

Rozwiązanie:

$L(G_1)$ = zbiór niepustych napisów nad $\{b, c\}$

Rodzaje gramatyk

ZADANIE:

$$G_1: \begin{array}{lcl} A & \rightarrow & b \\ A & \rightarrow & c \\ A & \rightarrow & AA \end{array}$$

← gramatyka **nie**prawoliniowa

Czy $L(G_1)$ jest regularny?

Rozwiązanie:

$L(G_1)$ = zbiór niepustych napisów nad $\{b, c\}$

$$G_2: \begin{array}{lcl} A & \rightarrow & b \\ A & \rightarrow & c \\ A & \rightarrow & bA \\ A & \rightarrow & cA \end{array}$$

Rodzaje gramatyk

ZADANIE:

$$G_1: \begin{array}{l} A \rightarrow b \\ A \rightarrow c \\ A \rightarrow AA \end{array}$$

← gramatyka **nie**prawoliniowa

Czy $L(G_1)$ jest regularny?

Rozwiązanie:

$L(G_1) =$ zbiór niepustych napisów nad $\{b, c\}$

$L(G_2) =$ zbiór niepustych napisów nad $\{b, c\}$

$$G_2: \begin{array}{l} A \rightarrow b \\ A \rightarrow c \\ A \rightarrow bA \\ A \rightarrow cA \end{array}$$

Rodzaje gramatyk

ZADANIE:

$$G_1: \begin{array}{l} A \rightarrow b \\ A \rightarrow c \\ A \rightarrow AA \end{array}$$

← gramatyka **nie**prawoliniowa

Czy $L(G_1)$ jest regularny?

Rozwiązanie:

$L(G_1)$ = zbiór niepustych napisów nad $\{b, c\}$

$L(G_2)$ = zbiór niepustych napisów nad $\{b, c\}$

$L(G_1) = L(G_2)$ i gramatyka G_2 jest prawoliniowa.

$$G_2: \begin{array}{l} A \rightarrow b \\ A \rightarrow c \\ A \rightarrow bA \\ A \rightarrow cA \end{array}$$

Rodzaje gramatyk

ZADANIE:

$$G_1: \begin{array}{lcl} A & \rightarrow & b \\ A & \rightarrow & c \\ A & \rightarrow & AA \end{array}$$

← gramatyka **nie**prawoliniowa

Czy $L(G_1)$ jest regularny?

Rozwiązanie:

$L(G_1)$ = zbiór niepustych napisów nad $\{b, c\}$

$L(G_2)$ = zbiór niepustych napisów nad $\{b, c\}$

$$G_2: \begin{array}{lcl} A & \rightarrow & b \\ A & \rightarrow & c \\ A & \rightarrow & bA \\ A & \rightarrow & cA \end{array}$$

$L(G_1) = L(G_2)$ i gramatyka G_2 jest prawoliniowa.

Ponieważ istnieje prawoliniowa gramatyka G_2 **równoważna** gramatyce G_1 (czyli generująca ten sam język), więc $L(G_1)$ **jest regularny**.

Rodzaje gramatyk

Język zadany gramatyką **prawoliniową**
jest regularny.

Rodzaje gramatyk

Język zadany gramatyką **prawoliniową**
jest regularny.

Język zadany bezkontekstową gramatyką **nieprawoliniową**
może nie być regularny

Rodzaje gramatyk

Język zadany gramatyką **prawoliniową**
jest regularny.

Język zadany bezkontekstową gramatyką **nieprawoliniową**
może nie być regularny,
może być regularny

Rodzaje gramatyk

Język zadany gramatyką **prawoliniową**
jest regularny.

Język zadany bezkontekstową gramatyką **nieprawoliniową**
może nie być regularny,
może być regularny,
— potrzebne jest staranniejsze zbadanie.

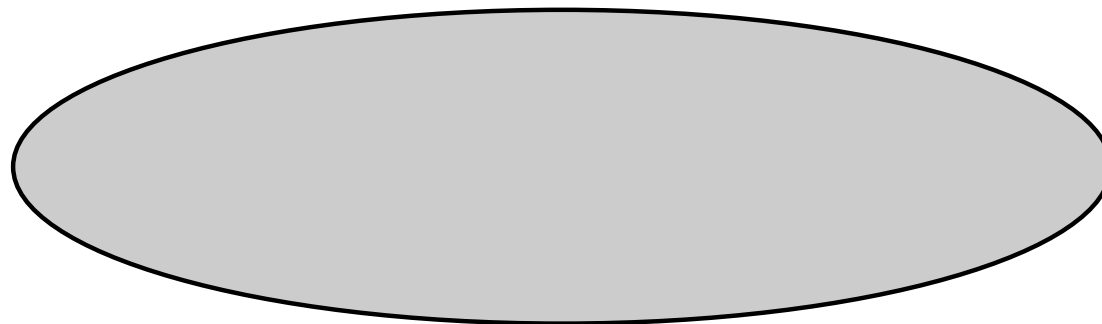
Hierarchia języków



Noam Chomsky, ur. 1928, MIT
„the father of modern linguistics”

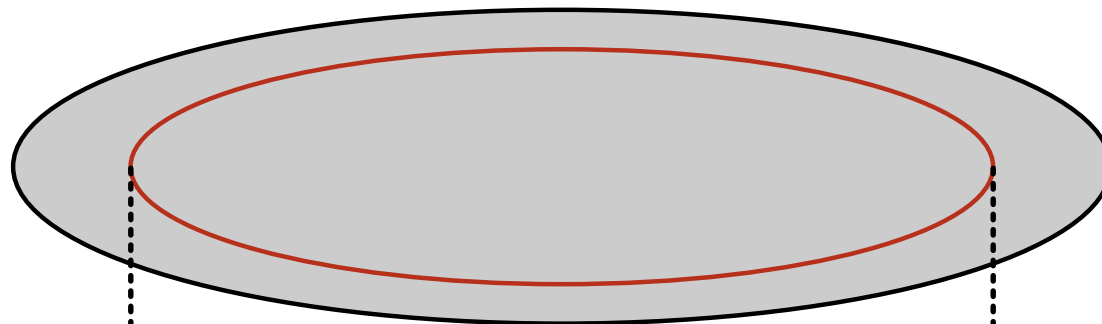
Hierarchia języków

języki

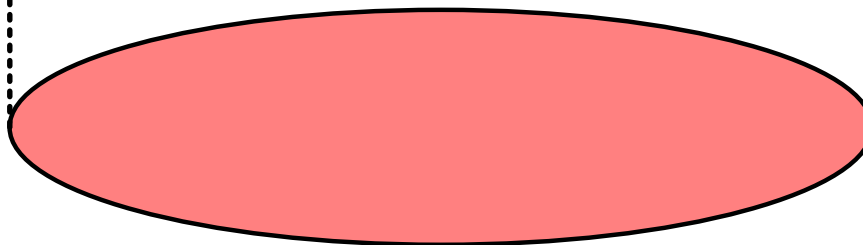


Hierarchia języków

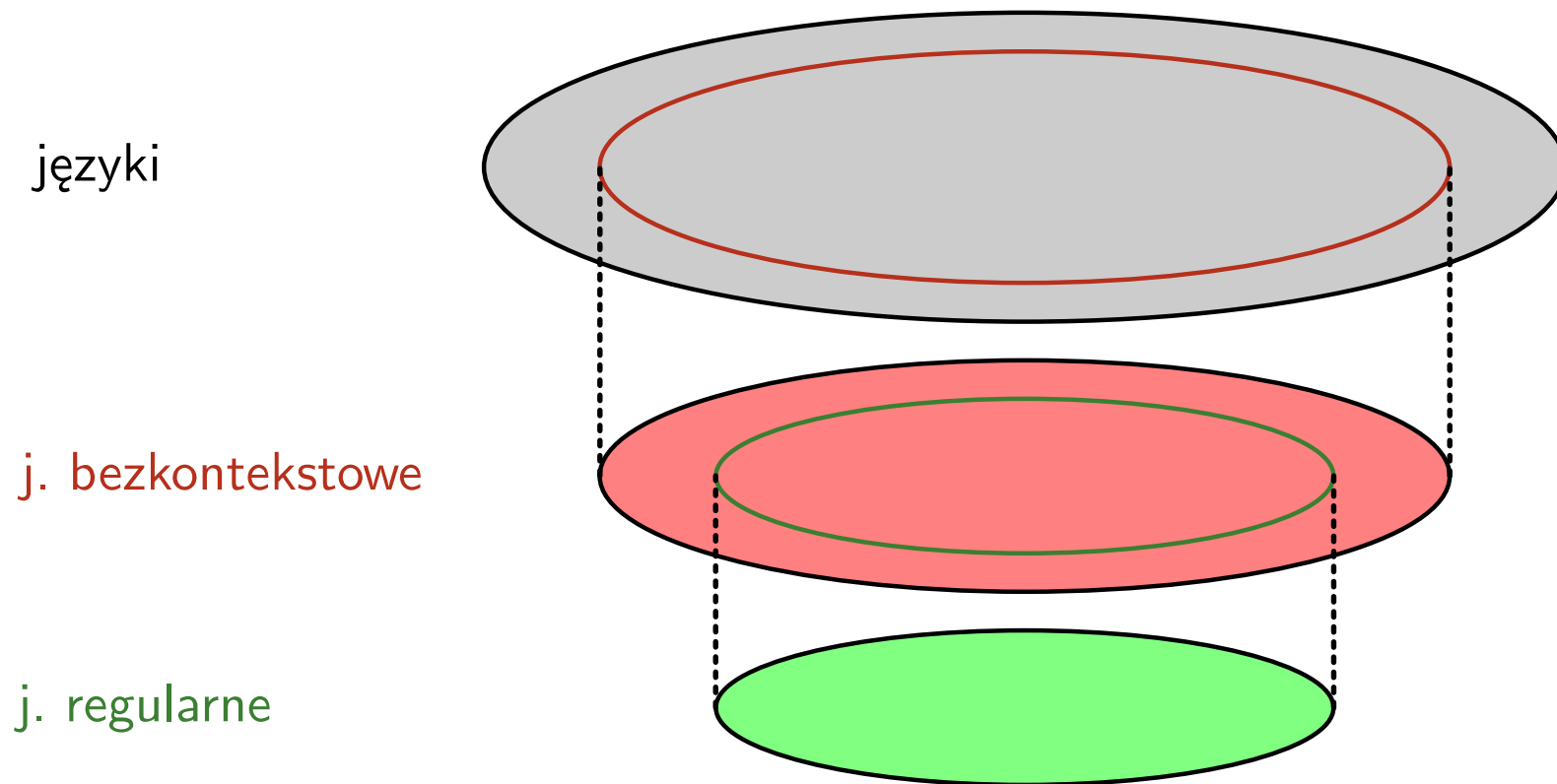
języki



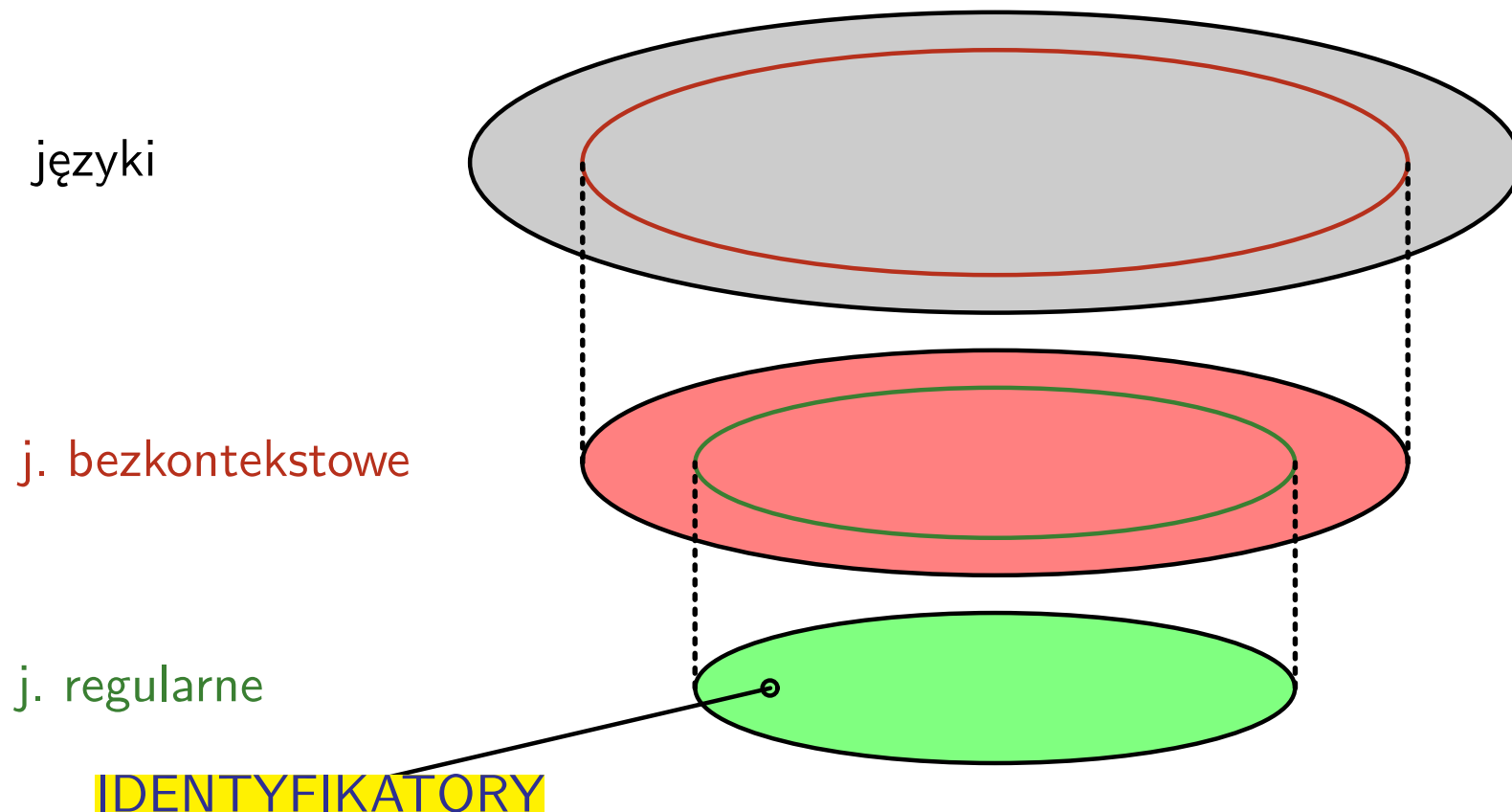
j. bezkontekstowe



Hierarchia języków

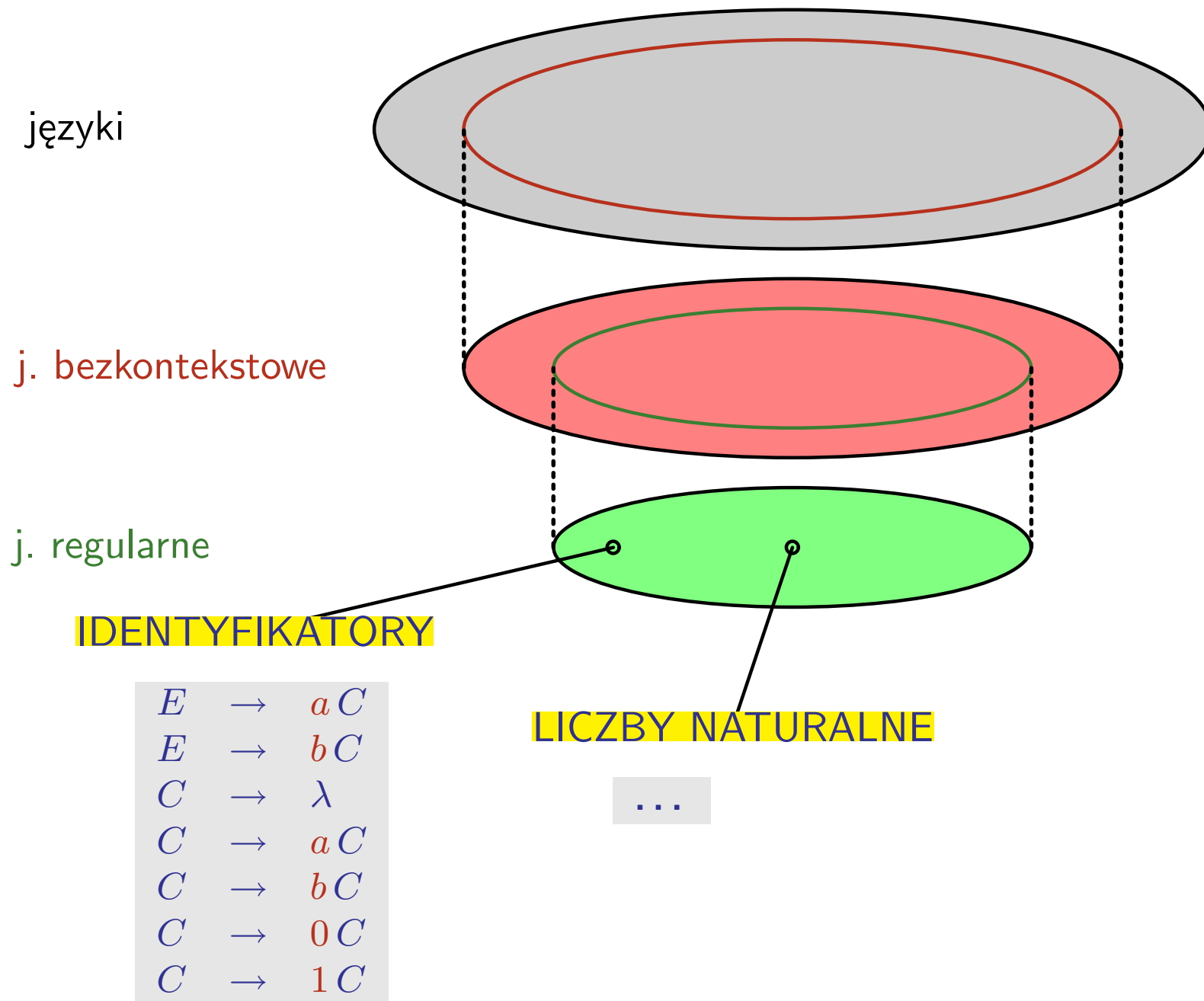


Hierarchia języków

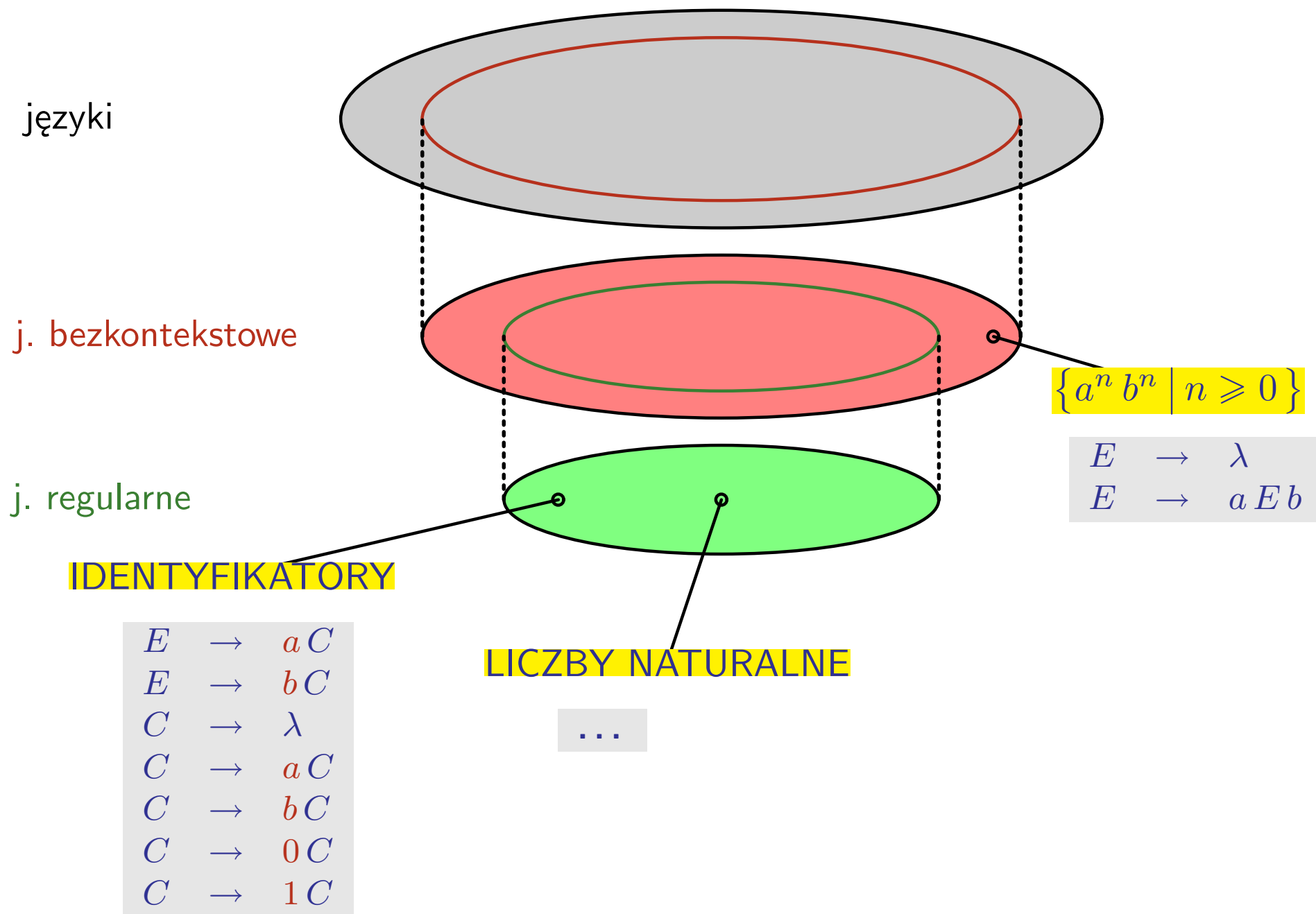


E	$\rightarrow$	aC
E	$\rightarrow$	bC
C	$\rightarrow$	λ
C	$\rightarrow$	aC
C	$\rightarrow$	bC
C	$\rightarrow$	$0C$
C	$\rightarrow$	$1C$

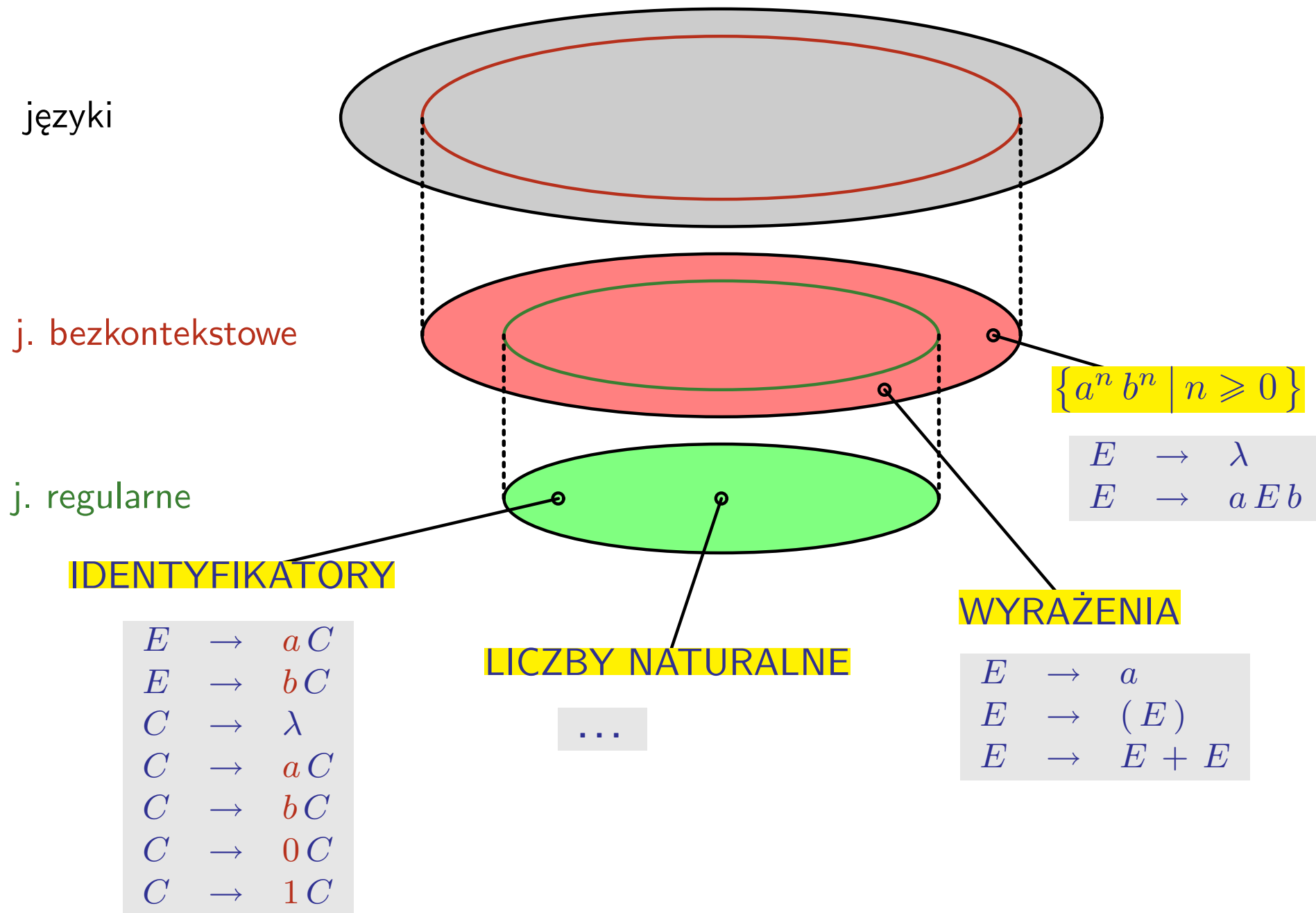
Hierarchia języków



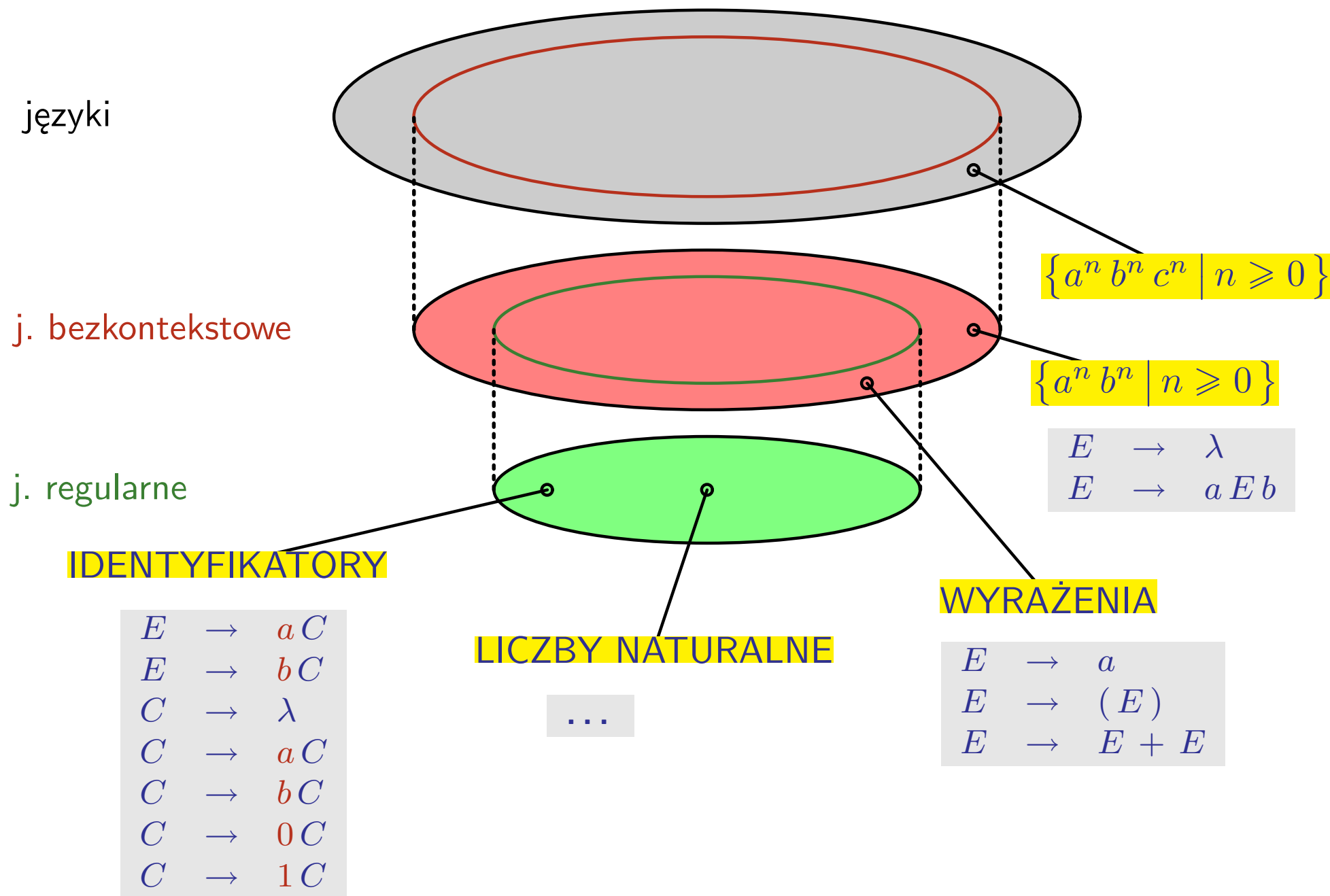
Hierarchia języków



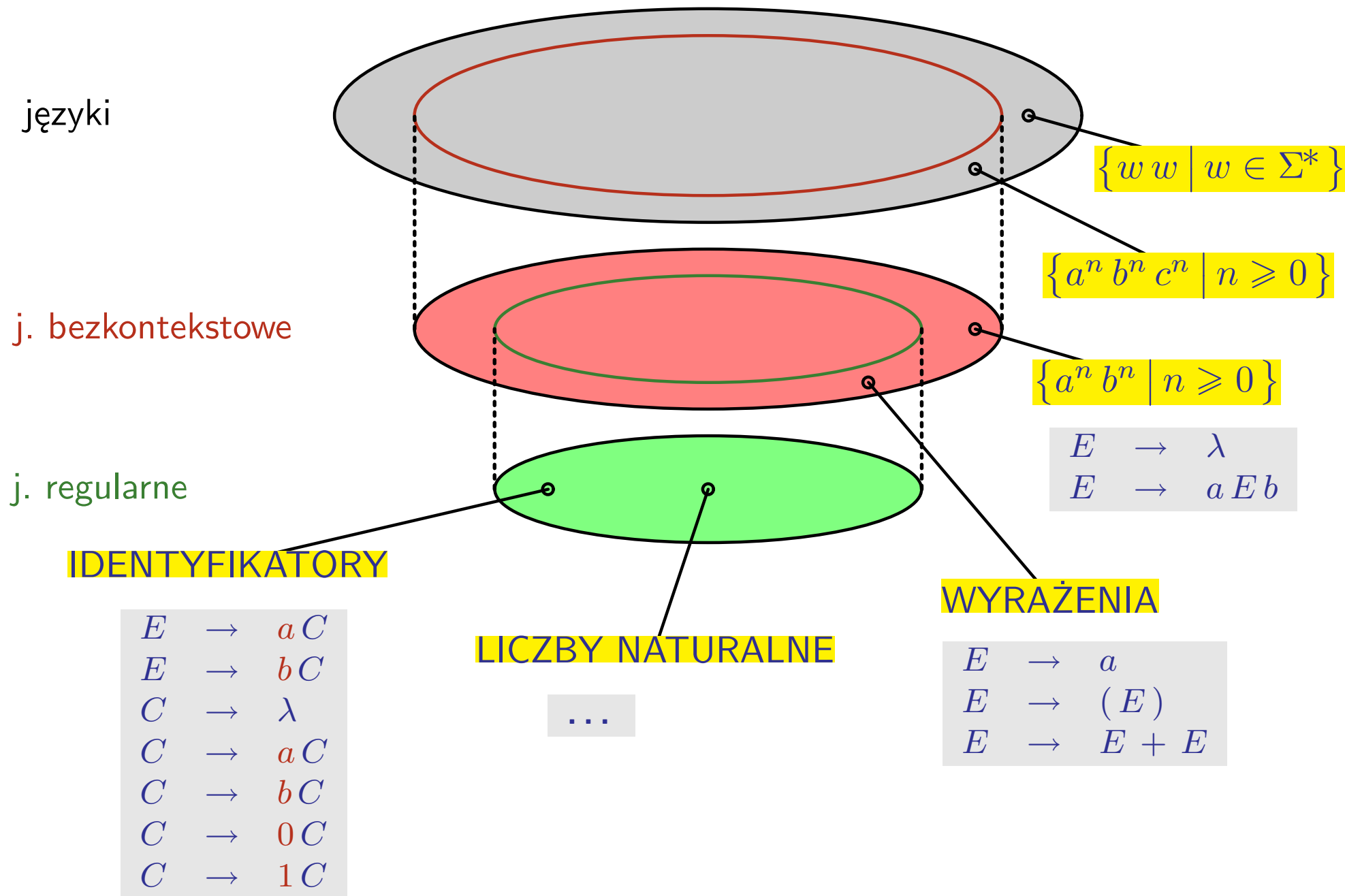
Hierarchia języków



Hierarchia języków



Hierarchia języków



Czy dany język jest bezkontekstowy lub regularny?

Żeby dowieść, że język **jest** bezkontekstowy [ew. regularny], należy znaleźć dla niego gramatykę bezkontekstową [prawoliniową].

Czy dany język jest bezkontekstowy lub regularny?

Żeby dowieść, że język **jest** bezkontekstowy [ew. regularny], należy znaleźć dla niego gramatykę bezkontekstową [prawoliniową].

Żeby dowieść, że język **nie jest** bezkontekstowy [regularny], należy udowodnić, że nie posiada gramatyki bezkontekstowej [prawoliniowej]

Czy dany język jest bezkontekstowy lub regularny?

Żeby dowieść, że język **jest** bezkontekstowy [ew. regularny], należy znaleźć dla niego gramatykę bezkontekstową [prawoliniową].

Żeby dowieść, że język **nie jest** bezkontekstowy [regularny], należy udowodnić, że nie posiada gramatyki bezkontekstowej [prawoliniowej] — to jest **na ogół trudne**.

Czy dany język jest bezkontekstowy lub regularny?

Żeby dowieść, że język **jest** bezkontekstowy [ew. regularny], należy znaleźć dla niego gramatykę bezkontekstową [prawoliniową].

Żeby dowieść, że język **nie jest** bezkontekstowy [regularny], należy udowodnić, że nie posiada gramatyki bezkontekstowej [prawoliniowej] — to jest **na ogół trudne**.

Bardzo nieformalne kryteria negatywne:

1. Rozpoznanie słowa języka **regularnego** nie wymaga porównywania liczby liter.
Np. język $\{a^n b^n \mid n \geq 0\}$ nie jest regularny.

Czy dany język jest bezkontekstowy lub regularny?

Żeby dowieść, że język **jest** bezkontekstowy [ew. regularny], należy znaleźć dla niego gramatykę bezkontekstową [prawoliniową].

Żeby dowieść, że język **nie jest** bezkontekstowy [regularny], należy udowodnić, że nie posiada gramatyki bezkontekstowej [prawoliniowej] — to jest **na ogół trudne**.

Bardzo nieformalne kryteria negatywne:

1. Rozpoznanie słowa języka **regularnego** nie wymaga porównywania liczby liter.
Np. język $\{a^n b^n \mid n \geq 0\}$ nie jest regularny.
2. Rozpoznanie słowa języka **bezkontekstowego** może wymagać porównania liczby liter ale tylko w dwóch blokach.
Np. $\{a^n b^n \mid n \geq 0\}$ jest a $\{a^n b^n a^n \mid n \geq 0\}$ nie jest bezkontekstowy.

Czy dany język jest bezkontekstowy lub regularny?

Żeby dowieść, że język **jest** bezkontekstowy [ew. regularny], należy znaleźć dla niego gramatykę bezkontekstową [prawoliniową].

Żeby dowieść, że język **nie jest** bezkontekstowy [regularny], należy udowodnić, że nie posiada gramatyki bezkontekstowej [prawoliniowej] — to jest **na ogół trudne**.

Bardzo nieformalne kryteria negatywne:

1. Rozpoznanie słowa języka **regularnego** nie wymaga porównywania liczby liter.
Np. język $\{a^n b^n \mid n \geq 0\}$ nie jest regularny.
2. Rozpoznanie słowa języka **bezkontekstowego** może wymagać porównania liczby liter ale tylko w dwóch blokach.
Np. $\{a^n b^n \mid n \geq 0\}$ jest a $\{a^n b^n a^n \mid n \geq 0\}$ nie jest bezkontekstowy.
3. Rozpoznanie słowa języka **bezkontekstowego** może wymagać **porównania** liczby liter j.w., ale nie może wymagać **obliczeń** arytmetycznych.
Np. $\{a^n b^{n^2} \mid n \geq 0\}$ nie jest bezkontekstowy.

Czy dany język jest bezkontekstowy lub regularny?

Żeby dowieść, że język **jest** bezkontekstowy [ew. regularny], należy znaleźć dla niego gramatykę bezkontekstową [prawoliniową].

Żeby dowieść, że język **nie jest** bezkontekstowy [regularny], należy udowodnić, że nie posiada gramatyki bezkontekstowej [prawoliniowej] — to jest **na ogół trudne**.

Bardzo nieformalne kryteria negatywne:

1. Rozpoznanie słowa języka **regularnego** nie wymaga porównywania liczby liter.
Np. język $\{a^n b^n \mid n \geq 0\}$ nie jest regularny.
2. Rozpoznanie słowa języka **bezkontekstowego** może wymagać porównania liczby liter ale tylko w dwóch blokach.
Np. $\{a^n b^n \mid n \geq 0\}$ jest a $\{a^n b^n a^n \mid n \geq 0\}$ nie jest bezkontekstowy.
3. Rozpoznanie słowa języka **bezkontekstowego** może wymagać **porównania** liczby liter j.w., ale nie może wymagać **obliczeń** arytmetycznych.
Np. $\{a^n b^{n^2} \mid n \geq 0\}$ nie jest bezkontekstowy.
4. Rozpoznanie słowa języka **bezkontekstowego** nie może wymagać pamiętania słów bez ograniczenia długości.
Np. $\{ww^R \mid w \in \Sigma^*\}$ jest a $\{ww \mid w \in \Sigma^*\}$ nie jest bezkontekstowy.

Definiowanie języka przez wyrażenie regularne (**WR**)

Języki **regularne** (ale nie dowolne bezkontekstowe) można definiować bez użycia nieterminali.

Definiowanie języka przez wyrażenie regularne (**WR**)

Języki **regularne** (ale nie dowolne bezkontekstowe) można definiować bez użycia nieterminali. Często jest to prostsze i czytelniejsze niż podawanie gramatyki.

Definiowanie języka przez wyrażenie regularne (**WR**)

Języki **regularne** (ale nie dowolne bezkontekstowe) można definiować bez użycia nieterminali. Często jest to prostsze i czytelniejsze niż podawanie gramatyki.

Wyrażenia regularne spotyka się w różnych postaciach

Definiowanie języka przez wyrażenie regularne (**WR**)

Języki **regularne** (ale nie dowolne bezkontekstowe) można definiować bez użycia nieterminali. Często jest to prostsze i czytelniejsze niż podawanie gramatyki.

Wyrażenia regularne spotyka się w różnych postaciach; na przykład

- w **komendach uniksowej konsoli** tekstowej

Definiowanie języka przez wyrażenie regularne (**WR**)

Języki **regularne** (ale nie dowolne bezkontekstowe) można definiować bez użycia nieterminali. Często jest to prostsze i czytelniejsze niż podawanie gramatyki.

Wyrażenia regularne spotyka się w różnych postaciach; na przykład

- w komendach uniksowej konsoli tekstowej, np.:

usuń z akt. katalogu wszystkie pliki o nazwach zawierających myślnik:

```
rm *-*
```

Definiowanie języka przez wyrażenie regularne (**WR**)

Języki **regularne** (ale nie dowolne bezkontekstowe) można definiować bez użycia nieterminali. Często jest to prostsze i czytelniejsze niż podawanie gramatyki.

Wyrażenia regularne spotyka się w różnych postaciach; na przykład

- w komendach uniksowej konsoli tekstowej, np.:

usuń z akt. katalogu wszystkie pliki o nazwach zawierających myślnik:

```
rm *-*
```

- w językach programowania; np. w Javie

Definiowanie języka przez wyrażenie regularne (**WR**)

Języki **regularne** (ale nie dowolne bezkontekstowe) można definiować bez użycia nieterminali. Często jest to prostsze i czytelniejsze niż podawanie gramatyki.

Wyrażenia regularne spotyka się w różnych postaciach; na przykład

- w komendach uniksowej konsoli tekstowej, np.:

usuń z akt. katalogu wszystkie pliki o nazwach zawierających myślnik:

```
rm *-*
```

- w językach programowania; np. w Javie:

sprawdź, czy dany napis zawiera myślnik:

```
iis.ans-elblag.pl/~stefan/Dydaktyka/JezForm/Wyklady/02-pom/Wzorce.java
```

Definiowanie języka przez wyrażenie regularne (**WR**)

Języki **regularne** (ale nie dowolne bezkontekstowe) można definiować bez użycia nieterminali. Często jest to prostsze i czytelniejsze niż podawanie gramatyki.

Wyrażenia regularne spotyka się w różnych postaciach; na przykład

- w komendach uniksowej konsoli tekstowej, np.:

usuń z akt. katalogu wszystkie pliki o nazwach zawierających myślnik:

```
rm *-*
```

- w językach programowania; np. w Javie:

sprawdź, czy dany napis zawiera myślnik:

```
iis.ans-elblag.pl/~stefan/Dydaktyka/JezForm/Wyklady/02-pom/Wzorce.java
```

- w edytorach przy szukaniu i zastępowaniu

Definiowanie języka przez wyrażenie regularne (**WR**)

Języki **regularne** (ale nie dowolne bezkontekstowe) można definiować bez użycia nieterminali. Często jest to prostsze i czytelniejsze niż podawanie gramatyki.

Wyrażenia regularne spotyka się w różnych postaciach; na przykład

- w komendach uniksowej konsoli tekstowej, np.:

usuń z akt. katalogu wszystkie pliki o nazwach zawierających myślnik:

```
rm *-*
```

- w językach programowania; np. w Javie:

sprawdź, czy dany napis zawiera myślnik:

```
iis.ans-elblag.pl/~stefan/Dydaktyka/JezForm/Wyklady/02-pom/Wzorce.java
```

- w edytorach przy szukaniu i zastępowaniu, np.

wszystkie myślniki wydzielone spacjami zastąp przez podkreślone odstępy:

```
Alt-X replace-regexp
```

```
" - "
```

```
" _ "
```

Definiowanie języka przez wyrażenie regularne (**WR**)

Definiowanie języka przez wyrażenie regularne (**WR**)

DEFINICJA: (wyrażenia regularne)

(specjalne znaki nie należące do alfabetu: {, }, |, *)

Definiowanie języka przez wyrażenie regularne (**WR**)

DEFINICJA: (wyrażenia regularne)

(specjalne znaki nie należące do alfabetu: {, }, |, *)

- **litery** alfabetu są **WR**

Definiowanie języka przez wyrażenie regularne (**WR**)

DEFINICJA: (wyrażenia regularne)

(specjalne znaki nie należące do alfabetu: $\{$, $\}$, $|$, $*$)

- **litery** alfabetu są **WR**
- **złożenie (konkatenacja)** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1W_2\}$ jest **WR**

Definiowanie języka przez wyrażenie regularne (**WR**)

DEFINICJA: (wyrażenia regularne)

(specjalne znaki nie należące do alfabetu: $\{$, $\}$, $|$, $*$)

- **litery** alfabetu są **WR**
- **złożenie (konkatenacja)** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1W_2\}$ jest **WR**
- **suma** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1 \mid W_2\}$ jest **WR**

Definiowanie języka przez wyrażenie regularne (**WR**)

DEFINICJA: (wyrażenia regularne)

(specjalne znaki nie należące do alfabetu: $\{$, $\}$, $|$, $*$)

- **littery** alfabetu są **WR**
- **złożenie (konkatenacja)** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1W_2\}$ jest **WR**
- **suma** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1 \mid W_2\}$ jest **WR**
- **domk. Kleenego (iteracja)** **WR** jest **WR**:
jeśli W jest **WR**, to $\{W\}^*$ jest **WR**

Definiowanie języka przez wyrażenie regularne (**WR**)

DEFINICJA: (wyrażenia regularne)

(specjalne znaki nie należące do alfabetu: $\{$, $\}$, $|$, $*$)

- **littery** alfabetu są **WR**
- **złożenie (konkatenacja)** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1W_2\}$ jest **WR**
- **suma** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1 | W_2\}$ jest **WR**
- **domk. Kleenego (iteracja)** **WR** jest **WR**:
jeśli W jest **WR**, to $\{W\}^*$ jest **WR**

Umowa notacyjna: opuszczamy zbędne klamry, przyjmując, że

- najsilniej łączy $*$ a najslabiej $|$,
- złożenie i suma łączą do lewej.

Definiowanie języka przez wyrażenie regularne (**WR**)

DEFINICJA: (wyrażenia regularne)

(specjalne znaki nie należące do alfabetu: $\{$, $\}$, $|$, $*$)

- **litery** alfabetu są **WR**
- **złożenie (konkatenacja)** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1W_2\}$ jest **WR**
- **suma** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1 | W_2\}$ jest **WR**
- **domk. Kleenego (iteracja)** **WR** jest **WR**:
jeśli W jest **WR**, to $\{W\}^*$ jest **WR**

Umowa notacyjna: opuszczamy zbędne klamry, przyjmując, że

- najsilniej łączy $*$ a najslabiej $|$,
- złożenie i suma łączą do lewej.

Np. zamiast

$$\left\{ \left\{ \left\{ \{ab\} a \right\} | \left\{ a \{c\}^* \right\} \right\} | \{cb\} \right\}$$

Definiowanie języka przez wyrażenie regularne (**WR**)

DEFINICJA: (wyrażenia regularne)

(specjalne znaki nie należące do alfabetu: $\{$, $\}$, $|$, $*$)

- **littery** alfabetu są **WR**
- **złożenie (konkatenacja)** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1W_2\}$ jest **WR**
- **suma** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1 | W_2\}$ jest **WR**
- **domk. Kleenego (iteracja)** **WR** jest **WR**:
jeśli W jest **WR**, to $\{W\}^*$ jest **WR**

Umowa notacyjna: opuszczamy zbędne klamry, przyjmując, że

- najsilniej łączy $*$ a najslabiej $|$,
- złożenie i suma łączą do lewej.

Np. zamiast

$$\left\{ \left\{ \left\{ \{ab\} a \right\} | \left\{ a \{c\}^* \right\} \right\} | \{cb\} \right\}$$

można napisać

$$aba | ac^* | cb$$

Definiowanie języka przez wyrażenie regularne (**WR**)

DEFINICJA: (wyrażenia regularne)

(specjalne znaki nie należące do alfabetu: $\{$, $\}$, $|$, $*$)

- **littery** alfabetu są **WR**
- **złożenie (konkatenacja)** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1W_2\}$ jest **WR**
- **suma** **WR** jest **WR**:
jeśli W_1 i W_2 są **WR**, to $\{W_1 | W_2\}$ jest **WR**
- **domk. Kleenego (iteracja)** **WR** jest **WR**:
jeśli W jest **WR**, to $\{W\}^*$ jest **WR**

Umowa notacyjna: opuszczamy zbędne klamry, przyjmując, że

- najsilniej łączy $*$ a najslabiej $|$,
- złożenie i suma łączą do lewej.

Np. zamiast

$$\left\{ \left\{ \left\{ \{ab\} a \right\} | \left\{ a \{c\}^* \right\} \right\} | \{cb\} \right\}$$

można napisać

aba	ac^*	cb
=====	=====	=====

Definiowanie języka przez wyrażenie regularne (**WR**)

Każdemu wyrażeniu regularnemu **odpowiada** język

Definiowanie języka przez wyrażenie regularne (**WR**)

Każdemu wyrażeniu regularnemu **odpowiada** język:

- pojedynczej literze σ **odpowiada** jednowyrazowy język $\{\sigma\}$

Definiowanie języka przez wyrażenie regularne (**WR**)

Każdemu wyrażeniu regularnemu **odpowiada** język:

- pojedynczej literze σ **odpowiada** jednowyrazowy język $\{\sigma\}$
- jeśli wyrażeniu W_1 **odpowiada** język L_1
a wyrażeniu W_2 **odpowiada** język L_2 ,
to wyrażeniu $\{W_1W_2\}$ **odpowiada** język złożenia (konkatenacji)
$$\{w_1w_2 \mid w_1 \in L_1 \wedge w_2 \in L_2\}$$

Definiowanie języka przez wyrażenie regularne (**WR**)

Każdemu wyrażeniu regularnemu **odpowiada** język:

- pojedynczej literze σ **odpowiada** jednowyrazowy język $\{\sigma\}$
- jeśli wyrażeniu W_1 **odpowiada** język L_1
a wyrażeniu W_2 **odpowiada** język L_2 ,
to wyrażeniu $\{W_1 W_2\}$ **odpowiada** język złożenia (konkatenacji)
$$\{w_1 w_2 \mid w_1 \in L_1 \wedge w_2 \in L_2\}$$
- jeśli wyrażeniu W_1 **odpowiada** język L_1
a wyrażeniu W_2 **odpowiada** język L_2 ,
to wyrażeniu $\{W_1 \mid W_2\}$ **odpowiada** suma języków
$$L_1 \cup L_2 = \{w \mid w \in L_1 \vee w \in L_2\}$$

Definiowanie języka przez wyrażenie regularne (**WR**)

Każdemu wyrażeniu regularnemu **odpowiada** język:

- pojedynczej literze σ **odpowiada** jednowyrazowy język $\{\sigma\}$
- jeśli wyrażeniu W_1 **odpowiada** język L_1
 a wyrażeniu W_2 **odpowiada** język L_2 ,
 to wyrażeniu $\{W_1 W_2\}$ **odpowiada** język złożenia (konkatenacji)

$$\{w_1 w_2 \mid w_1 \in L_1 \wedge w_2 \in L_2\}$$
- jeśli wyrażeniu W_1 **odpowiada** język L_1
 a wyrażeniu W_2 **odpowiada** język L_2 ,
 to wyrażeniu $\{W_1 \mid W_2\}$ **odpowiada** suma języków

$$L_1 \cup L_2 = \{w \mid w \in L_1 \vee w \in L_2\}$$
- jeśli wyrażeniu W **odpowiada** język L ,
 to wyrażeniu $\{W\}^*$ **odpowiada** nieskończona suma języków

$$\{\lambda\} \cup L \cup LL \cup LLL \cup LLLL \cup \dots$$

Wyrażenia regularne a języki regularne

TWIERDZENIE:

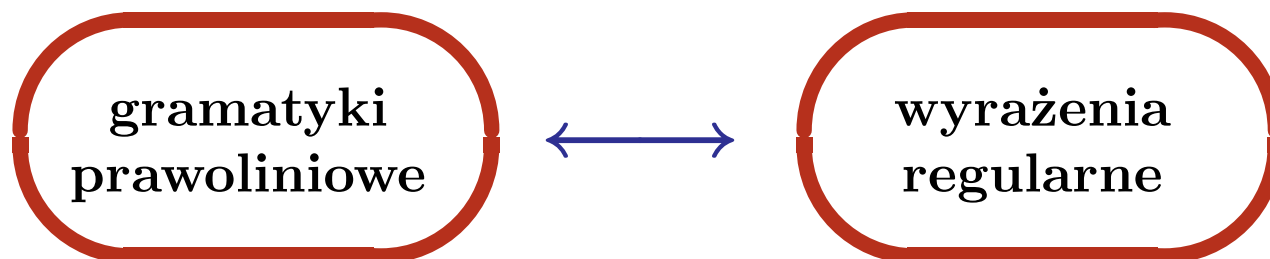
*Każdy **język regularny** odpowiada jakiemuś **wyrażeniu regularnemu**.*

Wyrażenia regularne a języki regularne

TWIERDZENIE:

Każdy *język regularny* odpowiada jakiemuś *wyrażeniu regularnemu*.

Każdy język, który *odpowiada* jakiemuś *wyrażeniu regularnemu*, jest *regularny*.

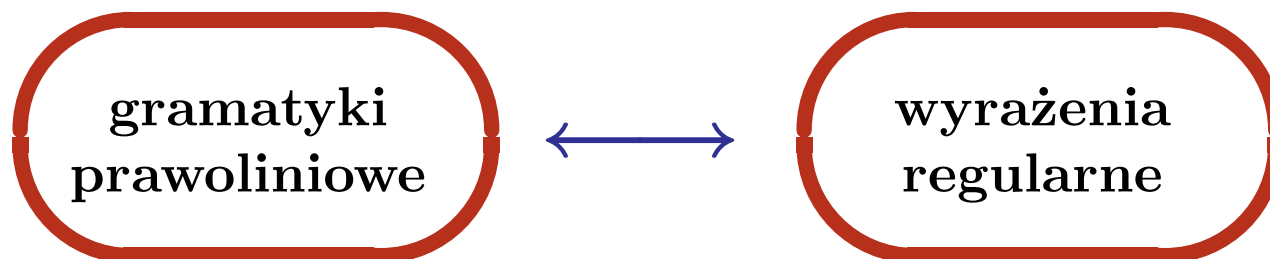


Wyrażenia regularne a języki regularne

TWIERDZENIE:

Każdy *język regularny* odpowiada jakiemuś *wyrażeniu regularnemu*.

Każdy język, który odpowiada jakiemuś *wyrażeniu regularnemu*, jest *regularny*.



Czyli:

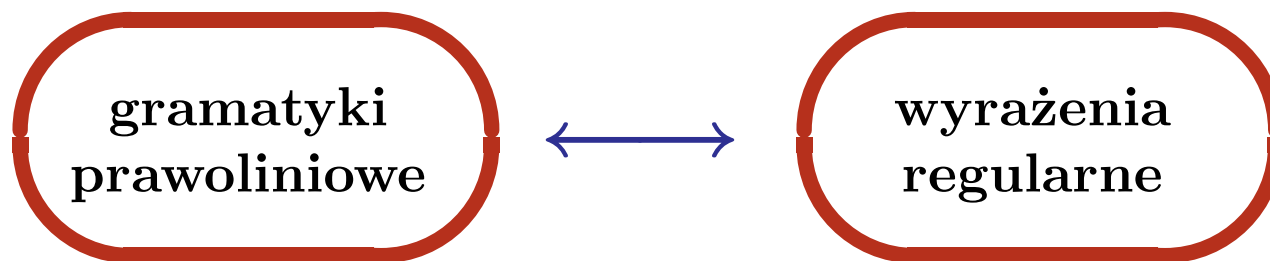
Języki regularne można definiować

Wyrażenia regularne a języki regularne

TWIERDZENIE:

Każdy *język regularny* odpowiada jakiemuś *wyrażeniu regularnemu*.

Każdy język, który odpowiada jakiemuś *wyrażeniu regularnemu*, jest *regularny*.



Czyli:

Języki regularne można definiować

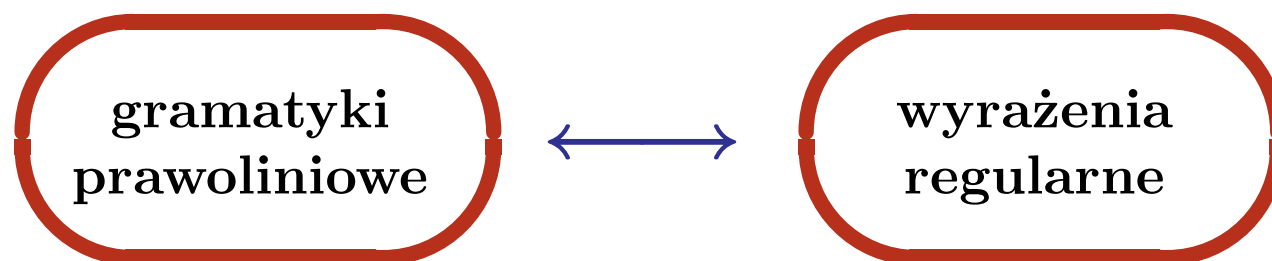
- albo przy pomocy **gramatyk prawoliniowych**

Wyrażenia regularne a języki regularne

TWIERDZENIE:

Każdy **język regularny** odpowiada jakiemuś **wyrażeniu regularnemu**.

Każdy język, który **odpowiada** jakiemuś **wyrażeniu regularnemu**, jest **regularny**.



Czyli:

Języki regularne można definiować

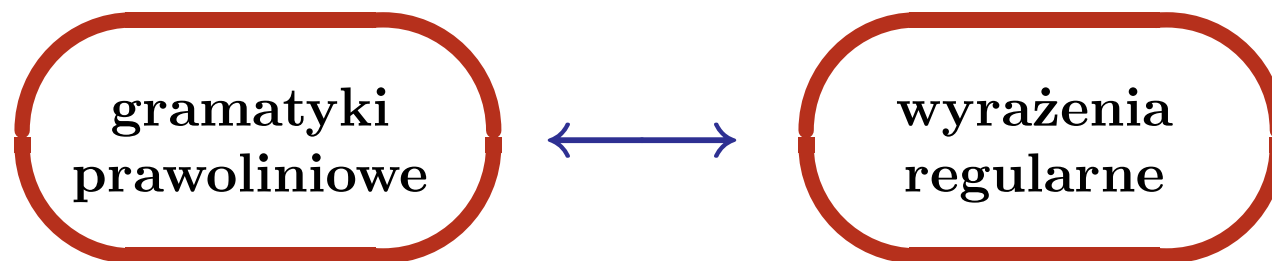
- albo przy pomocy **gramatyk prawoliniowych**,
- albo przy pomocy **wyrażeń regularnych**

Wyrażenia regularne a języki regularne

TWIERDZENIE:

Każdy **język regularny** odpowiada jakiemuś **wyrażeniu regularnemu**.

Każdy język, który **odpowiada** jakiemuś **wyrażeniu regularnemu**, jest **regularny**.



Czyli:

Języki regularne można definiować

- albo przy pomocy **gramatyk prawoliniowych**,
- albo przy pomocy **wyrażeń regularnych**

— to wychodzi na jedno.

Wyrażenia regularne a języki regularne

PRZYKŁAD:

Gramatyka prawoliniowa:

$$S \rightarrow \lambda$$

$$S \rightarrow abS$$

$$S \rightarrow cS$$

Wyrażenia regularne a języki regularne

PRZYKŁAD:

Gramatyka prawoliniowa:

$$\begin{aligned} S &\rightarrow \lambda \\ S &\rightarrow abS \\ S &\rightarrow cS \end{aligned}$$



Wyrażenie regularne:

$$\{ab \mid c\}^*$$

Wyrażenia regularne a języki regularne

PRZYKŁAD:

Gramatyka prawoliniowa:

$$\begin{aligned} S &\rightarrow \lambda \\ S &\rightarrow abS \\ S &\rightarrow cS \end{aligned}$$



Wyrażenie regularne:

$$\{ab \mid c\}^*$$

PRZYKŁAD:

Wyrażenie regularne:

$$a^*c^* \mid ba$$

Wyrażenia regularne a języki regularne

PRZYKŁAD:

Gramatyka prawoliniowa:

$$\begin{aligned} S &\rightarrow \lambda \\ S &\rightarrow abS \\ S &\rightarrow cS \end{aligned}$$



Wyrażenie regularne:

$$\{ab \mid c\}^*$$

PRZYKŁAD:

Wyrażenie regularne:

$$a^*c^* \mid ba$$



Gramatyka prawoliniowa:

$$\begin{aligned} S &\rightarrow A \\ S &\rightarrow ba \\ A &\rightarrow C \\ A &\rightarrow aA \\ C &\rightarrow \lambda \\ C &\rightarrow cC \end{aligned}$$

Wyrażenia regularne a języki regularne

PRZYKŁAD:

Gramatyka prawoliniowa:

$$\begin{aligned} S &\rightarrow \lambda \\ S &\rightarrow abS \\ S &\rightarrow cS \end{aligned}$$



Wyrażenie regularne:

$$\{ab \mid c\}^*$$

PRZYKŁAD:

Wyrażenie regularne:

$$a^*c^* \mid ba$$



Gramatyka prawoliniowa:

$$\begin{aligned} S &\rightarrow A \\ S &\rightarrow ba \\ A &\rightarrow C \\ A &\rightarrow aA \\ C &\rightarrow \lambda \\ C &\rightarrow cC \end{aligned}$$

Proste przejście między gramatyką regularną a wyrażeniem regularnym.

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- | oznacza „lub”,
- * oznacza powtarzanie (0 lub więcej razy)

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- $|$ oznacza „lub”,
 - $*$ oznacza powtarzanie (0 lub więcej razy)
-

PRZYKŁAD:

$\{0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9\}$

do tego wzorca pasuje dowolna cyfra

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- $|$ oznacza „lub”,
- $*$ oznacza powtarzanie (0 lub więcej razy)

PRZYKŁAD:

$\{0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9\}$

do tego wzorca pasuje dowolna cyfra

$\{a | b\}c$

jedno ze słów ac i bc

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- $|$ oznacza „lub”,
- $*$ oznacza powtarzanie (0 lub więcej razy)

PRZYKŁAD:

$\{0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9\}$

do tego wzorca pasuje dowolna cyfra

$\{a b\}c$		$a bc$
jedno ze słów ac i bc		jedno ze słów a i bc

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- $|$ oznacza „lub”,
- $*$ oznacza powtarzanie (0 lub więcej razy)

PRZYKŁAD:

$\{0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9\}$

do tego wzorca pasuje dowolna cyfra

$\{a \mid b\}c$		$a \mid bc$
jedno ze słów ac i bc		jedno ze słów a i bc

$\{a \mid b\}^*$

dowolne słowo napisane literami

a i b (łącznie z λ)

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- $|$ oznacza „lub”,
- $*$ oznacza powtarzanie (0 lub więcej razy)

PRZYKŁAD:

$\{0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9\}$

do tego wzorca pasuje dowolna cyfra

$\{a \mid b\}c$	$ $	$a \mid bc$
jedno ze słów ac i bc		jedno ze słów a i bc

$\{a \mid b\}^*$	$ $	$a^* \mid b^*$
dowolne słowo napisane literami a i b (łącznie z λ)		słowo z samych liter a lub słowo z samych liter b

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- $|$ oznacza „lub”,
 - $*$ oznacza powtarzanie (0 lub więcej razy)
-

PRZYKŁAD:

$\{ab\}^*$

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- $|$ oznacza „lub”,
 - $*$ oznacza powtarzanie (0 lub więcej razy)
-

PRZYKŁAD:

	$\{ab\}^*$		a^*b^*
słowa postaci	$ab\ ab\ \dots\ ab$		

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- $|$ oznacza „lub”,
- $*$ oznacza powtarzanie (0 lub więcej razy)

PRZYKŁAD:

 $\{ab\}^*$

słowa postaci $ab\ ab\ \dots\ ab$

 a^*b^*

słowa postaci $aa\ \dots\ a\ bb\ \dots\ b$

 $\{a^*b\}^*$

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- $|$ oznacza „lub”,
- $*$ oznacza powtarzanie (0 lub więcej razy)

PRZYKŁAD:

 $\{ab\}^*$

słowa postaci $ab\ ab\ \dots\ ab$

 a^*b^*

słowa postaci $aa\ \dots\ a\ bb\ \dots\ b$

 $\{a^*b\}^*$

słowa postaci $aa\ \dots\ b\ aa\ \dots\ b\ \dots\ aa\ \dots\ b$

 $\{0\ |\ \dots\ |\ 9\} \{0\ |\ \dots\ |\ 9\} - \{0\ |\ \dots\ |\ 9\} \{0\ |\ \dots\ |\ 9\} \{0\ |\ \dots\ |\ 9\}$

Wyrażenia regularne

Wyrażenie regularne jest **wzorcem** słowa z języka:

- $|$ oznacza „lub”,
- $*$ oznacza powtarzanie (0 lub więcej razy)

PRZYKŁAD:

 $\{ab\}^*$

słowa postaci $ab\ ab\ \dots\ ab$

 a^*b^*

słowa postaci $aa\ \dots\ a\ bb\ \dots\ b$

 $\{a^*b\}^*$

słowa postaci $aa\ \dots\ b\ aa\ \dots\ b\ \dots\ aa\ \dots\ b$

 $\{0\ |\ \dots\ |\ 9\} \{0\ |\ \dots\ |\ 9\} - \{0\ |\ \dots\ |\ 9\} \{0\ |\ \dots\ |\ 9\} \{0\ |\ \dots\ |\ 9\}$

dowolny kod pocztowy (dwie cyfry, myślnik, trzy cyfry)

Wyrażenia regularne

W „prawdziwym” świecie postać wyrażenia regularnego jest na ogół znacznie bardziej skomplikowana,

Wyrażenia regularne

W „prawdziwym” świecie postać wyrażenia regularnego jest na ogół znacznie bardziej skomplikowana,

ale

w zasadzie daje się sprowadzić do tych omówionych

Wyrażenia regularne

W „prawdziwym” świecie postać wyrażenia regularnego jest na ogół znacznie bardziej skomplikowana,

ale

w zasadzie daje się sprowadzić do tych omówionych:

- pojedyncza litera alfabetu

Wyrażenia regularne

W „prawdziwym” świecie postać wyrażenia regularnego jest na ogół znacznie bardziej skomplikowana,

ale

w zasadzie daje się sprowadzić do tych omówionych:

- pojedyncza litera alfabetu
- złożenie (konkatenacja)

Wyrażenia regularne

W „prawdziwym” świecie postać wyrażenia regularnego jest na ogół znacznie bardziej skomplikowana,

ale

w zasadzie daje się sprowadzić do tych omówionych:

- pojedyncza litera alfabetu
- złożenie (konkatenacja)
- suma

Wyrażenia regularne

W „prawdziwym” świecie postać wyrażenia regularnego jest na ogół znacznie bardziej skomplikowana,

ale

w zasadzie daje się sprowadzić do tych omówionych:

- pojedyncza litera alfabetu
- złożenie (konkatenacja)
- suma
- domk. Kleenego (iteracja)