A gentle introduction to Combinatorics on Words

Francesco Dolce

Konference TIGR CoW 2022

Janov nad Nisou, 19. května 2022

Some words about words

Some words about words

- *strč*, *prst*, *skrz*, *krk*

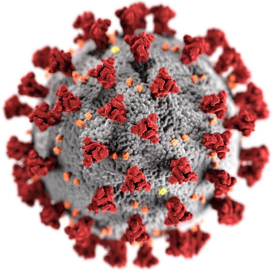
Some words about words

- *strč*, *prst*, *skrz*, *krk*

- 001, 101000, 010101010

Some words about words

- *strč*, *prst*, *skrz*, *krk*
- 001, 101000, 010101010
- ACGA, TACGGACATTA, CATATACG



Some words about words

- *strč*, *prst*, *skrz*, *krk*
- 001, 101000, 010101010
- ACGA, TACGGACATTA, CATATACG
- ♡♣♡♢, ♡♡♠♢
- ...

Let's start with the A,B,C's

Definition

- $\mathcal{A}$ (finite) is an *alphabet*
- $a \in \mathcal{A}$ is a *letter*
- $\mathcal{A}^*$ is the free monoid and $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$ the free semigroup
- $w \in \mathcal{A}^*$ is a *(finite) word*

Example

- $\varepsilon, \mathtt{a}, \mathtt{bba}, \mathtt{abba} \in \{\mathtt{a}, \mathtt{b}\}^*$
- $\mathtt{a} \cdot \mathtt{bb} = \mathtt{abb}$

Length matters

Definition

The *length* $|w|$ of a word $w = a_0 a_1 \cdots a_{n-1}$ is n.

Example

- $|\varepsilon| = 0$, $|\mathtt{a}| = 1$, $|\mathtt{abbba}| = 5$.

Length matters

Definition

The *length* $|w|$ of a word $w = a_0 a_1 \cdots a_{n-1}$ is n.

For each letter a, we have $|w|_a = \#\{\text{number of } a\text{'s in } w\}$.

Example

- $|\varepsilon| = 0$, $|\mathtt{a}| = 1$, $|\mathtt{abbba}| = 5$.
- $|\mathtt{abbba}|_{\mathtt{a}} = 2$, $|\mathtt{abbba}|_{\mathtt{b}} = 3$, $|\mathtt{abbba}|_{\mathtt{c}} = 0$.

Length matters

Definition

The *length* $|w|$ of a word $w = a_0 a_1 \cdots a_{n-1}$ is n.

For each letter a, we have $|w|_a = \#\{\text{number of } a\text{'s in } w\}$.

Example

- $|\varepsilon| = 0, \ \ |\mathtt{a}| = 1, \ \ |\mathtt{abbba}| = 5.$
- $|\mathtt{abbba}|_{\mathtt{a}} = 2, \ \ |\mathtt{abbba}|_{\mathtt{b}} = 3, \ \ |\mathtt{abbba}|_{\mathtt{c}} = 0.$

Proposition

For every word $w \in \mathcal{A}^*$ we have

$$|w| = \sum_{a \in \mathcal{A}} |w|_a.$$

Of beginnings, endings and everything in between

Definition

Let $w = pfs \in \mathcal{A}^*$ with $p, f, s \in \mathcal{A}^*$.

- f is a *factor* of w,
- p is a *prefix* of w (*proper prefix* is $p \neq w$),
- s is a *suffix* of w (*proper suffix* if $s \neq w$),
- if $p, s \in \mathcal{A}^+$, f is an *internal factor* of w.

Example

$w =$ nejblbější

Languages

Definition

A (finite or infinite) subset of $\mathcal{A}^*$ is called a *language*.

Example

- $L_1 = \{\mathtt{a}, \mathtt{b}, \mathtt{aba}, \mathtt{bb}\}$,
- $L_2 = \{w \in \mathcal{A}^* : |w| < 10\}$,
- $L_3 = \mathtt{aba}^* = \{\mathtt{ab}, \mathtt{aba}, \mathtt{abaa}, \mathtt{abaaa}, \mathtt{abaaa}, \ldots\}$,
- $L_4 = \{w \in \mathcal{A}^* : |w|_{\mathtt{b}} = 1\}$.

Languages

Definition

A (finite or infinite) subset of $\mathcal{A}^*$ is called a *language*.

The *language of a word* w is the set $\mathcal{L}(w)$ of all its factors.

Example

- $L_1 = \{\mathrm{a}, \mathrm{b}, \mathrm{aba}, \mathrm{bb}\}$,
- $L_2 = \{w \in \mathcal{A}^* : |w| < 10\}$,
- $L_3 = \mathrm{aba}^* = \{\mathrm{ab}, \mathrm{aba}, \mathrm{abaa}, \mathrm{abaaa}, \mathrm{abaaa}, \ldots\}$,
- $L_4 = \{w \in \mathcal{A}^* : |w|_\mathrm{b} = 1\}$.
- $\mathcal{L}(\mathrm{abba}) = \{\varepsilon, \mathrm{a}, \mathrm{b}, \mathrm{ab}, \mathrm{ba}, \mathrm{bb}, \mathrm{abb}, \mathrm{bba}, \mathrm{abba}\}$.

Languages

Definition

A (finite or infinite) subset of $\mathcal{A}^*$ is called a *language*.

The *language of a word* w is the set $\mathcal{L}(w)$ of all its factors.

A language is *factorial* if it contains the factors of every of its elements.

Example

- $L_1 = \{\mathtt{a}, \mathtt{b}, \mathtt{aba}, \mathtt{bb}\}$,
- $L_2 = \{w \in \mathcal{A}^* : |w| < 10\}$,
- $L_3 = \mathtt{aba}^* = \{\mathtt{ab}, \mathtt{aba}, \mathtt{abaa}, \mathtt{abaaa}, \mathtt{abaaa}, \ldots\}$,
- $L_4 = \{w \in \mathcal{A}^* : |w|_{\mathtt{b}} = 1\}$.
- $\mathcal{L}(\mathtt{abba}) = \{\varepsilon, \mathtt{a}, \mathtt{b}, \mathtt{ab}, \mathtt{ba}, \mathtt{bb}, \mathtt{abb}, \mathtt{bba}, \mathtt{abba}\}$.

Languages

Definition

A (finite or infinite) subset of $\mathcal{A}^*$ is called a *language*.

The *language of a word* w is the set $\mathcal{L}(w)$ of all its factors.

A language is *factorial* if it contains the factors of every of its elements.

Example

- $L_1 = \{\mathrm{a}, \mathrm{b}, \mathrm{aba}, \mathrm{bb}\}$, ✗ $\mathrm{aba} \in L_1$ but $\mathrm{ab} \notin L_1$
- $L_2 = \{w \in \mathcal{A}^* : |w| < 10\}$,
- $L_3 = \mathrm{aba}^* = \{\mathrm{ab}, \mathrm{aba}, \mathrm{abaa}, \mathrm{abaaa}, \mathrm{abaaa}, \ldots\}$,
- $L_4 = \{w \in \mathcal{A}^* : |w|_{\mathrm{b}} = 1\}$.
- $\mathcal{L}(\mathrm{abba}) = \{\varepsilon, \mathrm{a}, \mathrm{b}, \mathrm{ab}, \mathrm{ba}, \mathrm{bb}, \mathrm{abb}, \mathrm{bba}, \mathrm{abba}\}$.

Languages

Definition

A (finite or infinite) subset of $\mathcal{A}^*$ is called a *language*.

The *language of a word* w is the set $\mathcal{L}(w)$ of all its factors.

A language is *factorial* if it contains the factors of every of its elements.

Example

- $L_1 = \{a, b, aba, bb\}$, ✗ $aba \in L_1$ but $ab \notin L_1$
- $L_2 = \{w \in \mathcal{A}^* : |w| < 10\}$, ✓
- $L_3 = aba^* = \{ab, aba, abaa, abaaa, abaaa, \ldots\}$,
- $L_4 = \{w \in \mathcal{A}^* : |w|_b = 1\}$.
- $\mathcal{L}(abba) = \{\varepsilon, a, b, ab, ba, bb, abb, bba, abba\}$.

Languages

Definition

A (finite or infinite) subset of $\mathcal{A}^*$ is called a *language*.

The *language of a word* w is the set $\mathcal{L}(w)$ of all its factors.

A language is *factorial* if it contains the factors of every of its elements.

Example

- $L_1 = \{a, b, aba, bb\}$, ✗ $aba \in L_1$ but $ab \notin L_1$
- $L_2 = \{w \in \mathcal{A}^* : |w| < 10\}$, ✓
- $L_3 = aba^* = \{ab, aba, abaa, abaaa, abaaa, \ldots\}$, ✗ $aba \in L_3$ but $b \notin L_3$
- $L_4 = \{w \in \mathcal{A}^* : |w|_b = 1\}$.
- $\mathcal{L}(abba) = \{\varepsilon, a, b, ab, ba, bb, abb, bba, abba\}$.

Languages

Definition

A (finite or infinite) subset of $\mathcal{A}^*$ is called a *language*.

The *language of a word* w is the set $\mathcal{L}(w)$ of all its factors.

A language is *factorial* if it contains the factors of every of its elements.

Example

- $L_1 = \{\mathrm{a}, \mathrm{b}, \mathrm{aba}, \mathrm{bb}\}$, ✗ $\mathrm{aba} \in L_1$ but $\mathrm{ab} \notin L_1$
- $L_2 = \{w \in \mathcal{A}^* : |w| < 10\}$, ✓
- $L_3 = \mathrm{aba}^* = \{\mathrm{ab}, \mathrm{aba}, \mathrm{abaa}, \mathrm{abaaa}, \mathrm{abaaa}, \ldots\}$, ✗ $\mathrm{aba} \in L_3$ but $\mathrm{b} \notin L_3$
- $L_4 = \{w \in \mathcal{A}^* : |w|_{\mathrm{b}} = 1\}$. ✗ $\mathrm{ab} \in L_4$ but $\mathrm{a} \notin L_4$
- $\mathcal{L}(\mathrm{abba}) = \{\varepsilon, \mathrm{a}, \mathrm{b}, \mathrm{ab}, \mathrm{ba}, \mathrm{bb}, \mathrm{abb}, \mathrm{bba}, \mathrm{abba}\}$.

Languages

Definition

A (finite or infinite) subset of $\mathcal{A}^*$ is called a *language*.

The *language of a word* w is the set $\mathcal{L}(w)$ of all its factors.

A language is *factorial* if it contains the factors of every of its elements.

Example

- $L_1 = \{a, b, aba, bb\}$, ✗ $aba \in L_1$ but $ab \notin L_1$
- $L_2 = \{w \in \mathcal{A}^* : |w| < 10\}$, ✓
- $L_3 = aba^* = \{ab, aba, abaa, abaaa, abaaa, \ldots\}$, ✗ $aba \in L_3$ but $b \notin L_3$
- $L_4 = \{w \in \mathcal{A}^* : |w|_b = 1\}$. ✗ $ab \in L_4$ but $a \notin L_4$
- $\mathcal{L}(abba) = \{\varepsilon, a, b, ab, ba, bb, abb, bba, abba\}$. ✓ by construction

Factor complexity

The *factor complexity* function of a word w is the map $\mathcal{C}_w(n) : \mathcal{L}(w) \to \mathbb{N}$ counting the distinct factors of any length.

Example

$$w = \mathsf{abaccb}$$

n	0	1	2	3	4	5	6	7	8	9	10	...
$\mathcal{C}_w(n)$	1	3	5	4	3	2	1	0	0	0	0	...

$$\mathcal{L}(w) = \{\varepsilon, \underbrace{\mathsf{a}, \mathsf{b}, \mathsf{c}}_{3}, \underbrace{\mathsf{ab}, \mathsf{ac}, \mathsf{ba}, \mathsf{cb}, \mathsf{cc}}_{5}, \underbrace{\mathsf{aba}, \mathsf{acc}, \mathsf{bac}, \mathsf{ccb}}_{4}, \underbrace{\mathsf{abac}, \mathsf{accb}, \mathsf{bacc}}_{3}, \underbrace{\mathsf{abacc}, \mathsf{baccb}}_{2}, w\}$$

Special factors

Definition

An element u of a language L (resp. a factor u of a word w) is *right-special* if $ua, ub \in L$ (resp. $ua, ub \in \mathcal{L}(w)$) for two different letters a, b.
Similar definition for *left-special*.
A factor *bispecial* if it is both left- and right-special.

Example

Let $w = \mathtt{abaccb}$

- $\mathtt{a}$ is right-special, since $\mathtt{ab}, \mathtt{ac} \in \mathcal{L}(w)$;
- $\mathtt{b}$ is left-special, since $\mathtt{ab}, \mathtt{cb} \in \mathcal{L}(w)$;
- $\mathtt{c}$ is bispecial since it is both left-special and right-special ($\mathtt{ac}, \mathtt{cc}, \mathtt{cb} \in \mathcal{L}(w)$).

Powers

Definition

The n^{th}-*power* of a word w is defined recursively as

$$w^0 = \varepsilon, \qquad w^n = w^{n-1}w \text{ for every integer}^a\ n > 0.$$

When $n = 2$ (resp. $n = 3$) we call it a *square* (resp. a *cube*).

[a] What if n is not an integer? Wait for Lubka's and Daniela's talks later.

Example

- If $w = \mathtt{aba}$ then

$$w^0 = \varepsilon, \quad w^1 = \mathtt{aba}, \quad w^2 = \mathtt{abaaba}, \quad w^3 = \mathtt{abaabaaba}, \quad \ldots$$

Powers

Definition

The n^{th}-*power* of a word w is defined recursively as

$$w^0 = \varepsilon, \qquad w^n = w^{n-1}w \text{ for every integer}^a\ n > 0.$$

When $n = 2$ (resp. $n = 3$) we call it a *square* (resp. a *cube*).

A word w is *primitive* if, whenever $w = u^k$ then $k = 1$ and $w = u$.

---

[a] What if n is not an integer? Wait for Lubka's and Daniela's talks later.

Example

- If $w = \mathsf{aba}$ then

$$w^0 = \varepsilon, \quad w^1 = \mathsf{aba}, \quad w^2 = \mathsf{abaaba}, \quad w^3 = \mathsf{abaabaaba}, \quad \ldots$$

- $\mathsf{a}, \mathsf{b}, \mathsf{ab}, \mathsf{babba}$ are primitive, while $\mathsf{aaa}, \mathsf{abab}$ are NOT.

Conjugated words

Definition

Two words $w, w' \in \mathcal{A}^+$ are *conjugated*, denoted $w \equiv w'$, if there exist $x, y \in \mathcal{A}^+$ s.t. $w = xy$ and $w' = yx$.

The *class of conjugacy* of w is $[w] = \{w' \mid w' \equiv w\}$

Example

- $\mathsf{aba} \equiv \mathsf{aab}$, $\mathsf{abab} \equiv \mathsf{baba}$.
- $[\mathsf{aba}] = \{\mathsf{aab}, \mathsf{aba}, \mathsf{baa}\}$, $[\mathsf{abab}] = \{\mathsf{abab}, \mathsf{baba}\}$.

Oredered alphabets

Definition

Let us consider a total order $<$ on $\mathcal{A}$.
This order can be extended to $\mathcal{A}^*$, and it is called *lexicographical order*, by setting

$$u < v \iff \begin{cases} v = us & s \in \mathcal{A}^* \\ \text{or} & \\ u = pas,\ v = pbt & p, s, t \in \mathcal{A}^*,\ a, b \in \mathcal{A},\ a < b \end{cases}$$

Example

If $\mathcal{A} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ and $\mathtt{a} < \mathtt{b} < \mathtt{c}$, then

$$\mathtt{a} < \mathtt{aab} < \mathtt{ab} < \mathtt{aba} < \mathtt{b} < \mathtt{bac} < \mathtt{bb}.$$

Lyndon words

Definition [R. Lyndon (1954), А. И. Ширшов (1953)]

A word $w \in \mathcal{A}^+$ is a *Lyndon word* (or *правильное слово*) if for all $p, s \in \mathcal{A}^+$ s.t. $w = ps$ one has one of the three following equivalent conditions:

1. $w < sp$,
2. $w < s$,
3. $p < s$.

Example

a, b, ab, aab, ababb are Lyndon words, while abab and ba are NOT .

Lyndon words

Definition [R. Lyndon (1954), А. И. Ширшов (1953)]

A word $w \in \mathcal{A}^+$ is a *Lyndon word* (or *правильное слово*) if for all $p, s \in \mathcal{A}^+$ s.t. $w = ps$ one has one of the three following equivalent conditions:

1. $w < sp$,
2. $w < s$,
3. $p < s$.

Example

a, b, ab, aab, ababb are Lyndon words, while abab and ba are NOT .

Proposition

A word w is Lyndon word iff w is primitive and smaller than all its conjugates.

Lyndon factorization

Theorem [Lyndon (1954)]

Each word $w \in \mathcal{A}^+$ can be factorized in a unique way as $w = \ell_1 \ell_2 \cdots \ell_n$, with ℓ_i Lyndon word for every i and $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_n$.

Example

- aacab
- bc.bc.a
- b.abb.ab.a
- ab.a.a
- b.aaac.a

Do nekonečna a ještě dál

(see also Viola's talk just next)

Definition

An *infinite word* is a sequence $\mathbf{w} = a_0 a_1 a_2 \cdots$, with a_i letters.

The set of all (right-)infinite words over $\mathcal{A}$ is denoted by $\mathcal{A}^{\mathbb{N}}$.

Example

- $\mathbf{w} = \mathtt{abbaaaaaaaaaaaaaa} \cdots \in \{\mathtt{a}, \mathtt{b}\}^{\mathbb{N}}$;

Do nekonečna a ještě dál

(see also Viola's talk just next)

Definition

An *infinite word* is a sequence $\mathbf{w} = a_0 a_1 a_2 \cdots$, with a_i letters.

The set of all (right-)infinite words over $\mathcal{A}$ is denoted by $\mathcal{A}^{\mathbb{N}}$.

We can naturally extend the notions of *prefix*, *suffix*, *factor*, etc.

Example

- $\mathbf{w} = \mathtt{abbaaaaaaaaaaaaaa} \cdots \in \{\mathtt{a}, \mathtt{b}\}^{\mathbb{N}}$;
- ab is a proper prefix,
- $\mathbf{w}$ is a prefix;
- baa is an internal factor,
- baaaaaaa $\cdots$ is a suffix;
- $\mathtt{a}^{\omega} = \mathtt{aaaaa} \cdots$;
- $\mathcal{C}_{\mathbf{w}}(n) = 4$ for every $n \geq 2$ (Prove it!).

Recurrence and uniformly recurrence

Definition

A language $\mathcal{L}$ is *recurrent* if for every $u,v \in \mathcal{L}$ there is a $w \in \mathcal{L}$ such that uwv is in $\mathcal{L}$.

Example (Fibonacci, see later)

$$\mathbf{f} = \mathtt{abaababaabaababaababaabaababaaba} \cdots$$

Recurrence and uniformly recurrence

Definition

A language $\mathcal{L}$ is *recurrent* if for every $u, v \in \mathcal{L}$ there is a $w \in \mathcal{L}$ such that uwv is in $\mathcal{L}$.

$\mathcal{L}$ is *uniformly recurrent* if for every $u \in \mathcal{L}$ there exists an $n \in \mathbb{N}$ such that u is a factor of every word of length n in $\mathcal{L}$.

Example (Fibonacci, see later)

$$\mathbf{f} = \underbrace{\mathtt{abaa}}_{4}\,\mathtt{ba}\,\underbrace{\mathtt{baab}}_{4}\,\underbrace{\mathtt{aaba}}_{4}\,\mathtt{baababaaba}\,\underbrace{\mathtt{abab}}_{4}\,\mathtt{a}\cdots$$

Recurrence and uniformly recurrence

Definition

A language $\mathcal{L}$ is *recurrent* if for every $u,v \in \mathcal{L}$ there is a $w \in \mathcal{L}$ such that uwv is in $\mathcal{L}$.

$\mathcal{L}$ is *uniformly recurrent* if for every $u \in \mathcal{L}$ there exists an $n \in \mathbb{N}$ such that u is a factor of every word of length n in $\mathcal{L}$.

Proposition

Uniformly recurrence $\Longrightarrow$ Recurrence.

Recurrence and uniformly recurrence

Definition

A language $\mathcal{L}$ is *recurrent* if for every $u,v \in \mathcal{L}$ there is a $w \in \mathcal{L}$ such that uwv is in $\mathcal{L}$.

$\mathcal{L}$ is *uniformly recurrent* if for every $u \in \mathcal{L}$ there exists an $n \in \mathbb{N}$ such that u is a factor of every word of length n in $\mathcal{L}$.

Proposition

Uniformly recurrence $\Longrightarrow$ Recurrence.

Example (counter-example)

$$\mathbf{x} = \mathtt{a.b.aa.ab.ba.bb.aaa.aab.aba.abb.baa} \cdots$$

is recurrent, but $\mathtt{a}^n$ is never a factor of $\mathtt{b}^m$.

Morphisms

Definition

A *morphism* is a map $\psi : \mathcal{A}^* \to \mathcal{B}^*$ such that $\psi(uv) = \psi(u)\psi(v)$ for every $u, v \in \mathcal{A}^*$.

Example

$$\psi_1 : \begin{cases} \mathtt{a} \to \mathtt{010} \\ \mathtt{b} \to \mathtt{1} \end{cases} , \qquad \psi_2 : \begin{cases} \mathtt{a} \to \mathtt{ab} \\ \mathtt{b} \to \mathtt{b} \end{cases}$$

Morphisms

Definition

A *morphism* is a map $\psi : \mathcal{A}^* \to \mathcal{B}^*$ such that $\psi(uv) = \psi(u)\psi(v)$ for every $u, v \in \mathcal{A}^*$.

A *substitution* is a morphism $\psi : \mathcal{A} \to \mathcal{A}$ such that there exists a letter $a \in \mathcal{A}$ with $\psi(a) = as$ and $\lim_{n\to\infty} |\psi^n(a)| = \infty$. The word $\psi^\omega(a)$ is a *fixed point* of the substitution.

Example

$$\psi_1 : \begin{cases} \mathtt{a} \to \mathtt{010} \\ \mathtt{b} \to \mathtt{1} \end{cases}, \qquad \psi_2 : \begin{cases} \mathtt{a} \to \mathtt{ab} \\ \mathtt{b} \to \mathtt{b} \end{cases}$$

$$\psi_2^\omega(\mathtt{a}) = \mathtt{abbbbbbbbbbbbbbbbbb} \cdots = \mathtt{ab}^\omega.$$

Two important fixed points 1/2

Thue-Morse

The *Thue-Morse word* is defined as a fixed point

$$\mathbf{x} = 01101001100101101001011001101001100101100110100\cdots$$

of the morphism

$$\theta : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10 \end{cases}$$

Two important fixed points 1/2

Thue-Morse

The *Thue-Morse word* is defined as a fixed point

$$\mathbf{x} = 01101001100101101001011001101001100101100110100\cdots$$

of the morphism

$$\theta : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10 \end{cases}$$

Definition

A morphism ψ is a k-*uniform* if $|\psi(a)| = k$ for every letter a.

Two important fixed points 2/2

Fibonacci

The *Fibonacci word* is defined as the fixed point

$$\mathbf{f} = \varphi^{\omega}(\mathtt{a}) = \mathtt{abaababaabaababa}\cdots$$

of the morphism

$$\varphi : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{a} \end{cases}$$

The lengths of the prefixes $|\varphi^n(\mathtt{a})|_n = 0, 1, 2, 3, 5, 8, \ldots$ are the *Fibonacci numbers*[1].

[1] See Jan's talk on Saturday.

Sturmian words

(More about Sturmian words and Sturmian morphisms in Edita's talk on Sunday)

Definition

An infinite word $\mathbf{w}$ is *Sturmian* if it has $n+1$ distinct factors of length n for every $n \geq 0$.

Example (Fibonacci)

$$\mathbf{f} = \mathsf{abaababaabaababa} \cdots$$

$$\mathcal{L}(\mathbf{f}) = \{\underbrace{\varepsilon}_{1}, \underbrace{\mathsf{a}, \mathsf{b}}_{2}, \underbrace{\mathsf{aa}, \mathsf{ab}, \mathsf{ba}}_{3}, \underbrace{\mathsf{aab}, \mathsf{aba}, \mathsf{baa}, \mathsf{bab}}_{4}, \underbrace{\mathsf{aaba}, \mathsf{abaa}, \mathsf{abab}, \mathsf{baab}, \mathsf{baba}}_{5}, \ldots\}$$

Arnoux-Rauzy words

Definition

An infinite word $\mathbf{w}$ over an alphabet of k letters is an *Arnoux-Rauzy word* if

1. it has $(k-1)n+1$ distinct factors of length n for every $n \geq 0$;
2. for each lenght only one factor is right special; and
3. its set of factors is closed under reversal.

Example (Tribonacci: $\psi : \mathtt{a} \mapsto \mathtt{ab},\ \mathtt{b} \mapsto \mathtt{ac},\ \mathtt{c} \mapsto \mathtt{a}$)

$$\mathbf{t} = \mathtt{abacabaabacababacabaabaca}\cdots$$

$$\mathcal{L}(\mathbf{t}) = \{\underbrace{\varepsilon}_{1}, \underbrace{\mathtt{a}, \mathtt{b}, \mathtt{c}}_{3}, \underbrace{\mathtt{aa}, \mathtt{ab}, \mathtt{ac}, \mathtt{ba}, \mathtt{ca}}_{5}, \underbrace{\mathtt{aab}, \mathtt{aba}, \mathtt{aca}, \mathtt{baa}, \mathtt{bab}, \mathtt{bac}, \mathtt{cab}}_{7}, \ldots\}$$

Arnoux-Rauzy words

Definition

An infinite word $\mathbf{w}$ over an alphabet of k letters is an *Arnoux-Rauzy word* if

1. it has $(k-1)n+1$ distinct factors of length n for every $n \geq 0$;
2. for each lenght only one factor is right special; and
3. its set of factors is closed under reversal.

Example (Tribonacci: ψ : a $\mapsto$ ab, b $\mapsto$ ac, c $\mapsto$ a)

$$\mathbf{t} = \mathtt{abacabaabacababacabaabaca}\cdots$$

$$\mathcal{L}(\mathbf{t}) = \{\underbrace{\varepsilon}_{1}, \underbrace{\mathtt{a}, \mathtt{b}, \mathtt{c}}_{3}, \underbrace{\mathtt{aa}, \mathtt{ab}, \mathtt{ac}, \mathtt{ba}, \mathtt{ca}}_{5}, \underbrace{\mathtt{aab}, \mathtt{aba}, \mathtt{aca}, \mathtt{baa}, \mathtt{bab}, \mathtt{bac}, \mathtt{cab}}_{7}, \ldots\}$$

Arnoux-Rauzy words

Definition

An infinite word $\mathbf{w}$ over an alphabet of k letters is an *Arnoux-Rauzy word* if

1. it has $(k-1)n+1$ distinct factors of length n for every $n \geq 0$;
2. for each lenght only one factor is right special; and
3. its set of factors is closed under reversal.

Example (Tribonacci: ψ : $\mathtt{a} \mapsto \mathtt{ab},\ \mathtt{b} \mapsto \mathtt{ac},\ \mathtt{c} \mapsto \mathtt{a}$)

$$\mathbf{t} = \mathtt{abacabaabacababacabaabaca}\cdots$$

$$\mathcal{L}(\mathbf{t}) = \{\underbrace{\varepsilon}_{1}, \underbrace{\mathtt{a}, \mathtt{b}, \mathtt{c}}_{3}, \underbrace{\mathtt{aa}, \mathtt{ab}, \mathtt{ac}, \mathtt{ba}, \mathtt{ca}}_{5}, \underbrace{\mathtt{aab}, \mathtt{aba}, \mathtt{aca}, \mathtt{baa}, \mathtt{bab}, \mathtt{bac}, \mathtt{cab}}_{7}, \ldots\}$$

Palindromes
(see Viola's talk for that too)

Definition

A *palindrome* is a finite word w that is equal to its reversal $\widetilde{w}$.

Example

- kayak
- blb, krk, oko
- nepochopen

Palindromes

(see Viola's talk for that too)

Definition

A *palindrome* is a finite word w that is equal to its reversal $\widetilde{w}$.

Example

- kayak
- blb, krk, oko
- nepochopen
- Taco Cat

Palindromes

(see Viola's talk for that too)

Definition

A *palindrome* is a finite word w that is equal to its reversal $\widetilde{w}$.

Example

- kayak
- blb, krk, oko
- nepochopen
- Taco Cat
- V elipse spí lev

Palindromes

(see Viola's talk for that too)

Definition

A *palindrome* is a finite word w that is equal to its reversal $\widetilde{w}$.

Example

- kayak
- blb, krk, oko
- nepochopen
- Taco Cat
- V elipse spí lev
- Jelenovi pivo nelej

Palindromes

(see Viola's talk for that too)

Definition

A *palindrome* is a finite word w that is equal to its reversal $\widetilde{w}$.

Example

- kayak
- blb, krk, oko
- nepochopen
- Taco Cat
- V elipse spí lev
- Jelenovi pivo nelej
- Ital platí
- ...

Rich words

Theorem [Droubay, Justin, Pirillo (2001)]

A word of length n has at most $n + 1$ palindrome factors

A word with maximal number of palindromes is *rich*.

Rich words

Theorem [Droubay, Justin, Pirillo (2001)]

A word of length n has at most $n+1$ palindrome factors

A word with maximal number of palindromes is *rich*.

- $\mathcal{P}\{\texttt{pizza}\} = \{\varepsilon, \texttt{a}, \texttt{i}, \texttt{p}, \texttt{z}, \texttt{zz}\}$

  $\#\mathcal{P}\{\texttt{w}\} = 6 = |\texttt{w}| + 1$ ✓

Rich words

Theorem [Droubay, Justin, Pirillo (2001)]

A word of length n has at most $n+1$ palindrome factors

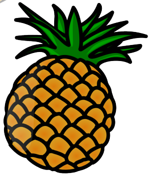
A word with maximal number of palindromes is *rich*.

- $\mathcal{P}\{\texttt{pizza}\} = \{\varepsilon, \texttt{a}, \texttt{i}, \texttt{p}, \texttt{z}, \texttt{zz}\}$

  $\#\mathcal{P}\{\texttt{w}\} = 6 = |\texttt{w}| + 1$ ✓

- $\mathcal{P}\{\texttt{ananas}\} = \{\varepsilon, \texttt{a}, \texttt{n}, \texttt{s}, \texttt{ana}, \texttt{nan}, \texttt{anana}\}$

  $\#\mathcal{P}\{\texttt{w}\} = 7 = |\texttt{w}| + 1$ ✓



Rich words

Theorem [Droubay, Justin, Pirillo (2001)]

A word of length n has at most $n+1$ palindrome factors

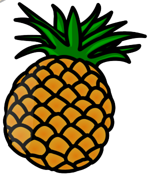
A word with maximal number of palindromes is *rich*.

- $\mathcal{P}\{\mathtt{pizza}\} = \{\varepsilon, \mathtt{a}, \mathtt{i}, \mathtt{p}, \mathtt{z}, \mathtt{zz}\}$

  $\#\mathcal{P}\{\mathtt{w}\} = 6 = |\mathtt{w}| + 1$ ✓

- $\mathcal{P}\{\mathtt{ananas}\} = \{\varepsilon, \mathtt{a}, \mathtt{n}, \mathtt{s}, \mathtt{ana}, \mathtt{nan}, \mathtt{anana}\}$

  $\#\mathcal{P}\{\mathtt{w}\} = 7 = |\mathtt{w}| + 1$ ✓



- $\mathcal{P}\{\mathtt{hawaiianpizza}\} = \{\varepsilon, \mathtt{a}, \mathtt{h}, \mathtt{i}, \mathtt{n}, \mathtt{p}, \mathtt{w}, \mathtt{z}, \mathtt{ii}, \mathtt{zz}, \mathtt{awa}, \mathtt{aiia}\}$

  $\#\mathcal{P}\{\mathtt{w}\} = 12 < 13 = |\mathtt{w}| + 1$ ✗

Rich words and rich sets

An infinite word **w** is *rich* if all its finite prefixes are rich.
A factorial set is *rich* if all its elements are rich.

Rich words and rich sets

An infinite word **w** is *rich* if all its finite prefixes are rich.
A factorial set is *rich* if all its elements are rich.

- Arnoux-Rauzy words

  [Droubay, Justin, Pirillo (2001)]

$$\mathbf{f} = \varphi^{\omega}(\mathtt{a}) = \mathtt{abaababaabaababaababaabaababaabaababaababaab}\cdots$$

$$\text{where} \quad \varphi = \begin{cases} \mathtt{a} \to \mathtt{ab} \\ \mathtt{b} \to \mathtt{a} \end{cases}$$

Rich words and rich sets

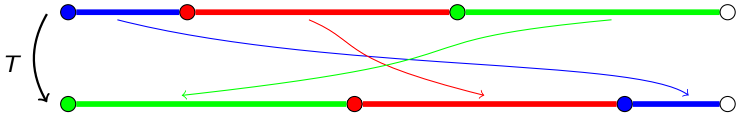
An infinite word **w** is *rich* if all its finite prefixes are rich.
A factorial set is *rich* if all its elements are rich.

- Arnoux-Rauzy words

  [Droubay, Justin, Pirillo (2001)]

- Symmetric regular interval exchange sets

  [Balázi, Masáková, Pelantová (2007)]



Rich words and rich sets

An infinite word **w** is *rich* if all its finite prefixes are rich.
A factorial set is *rich* if all its elements are rich.

- Arnoux-Rauzy words

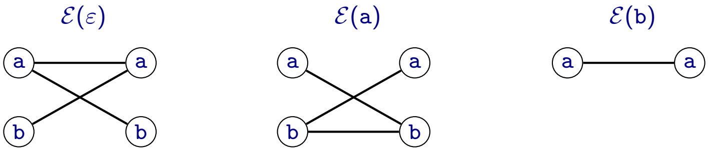
  [Droubay, Justin, Pirillo (2001)]

- Symmetric regular interval exchange sets

  [Baláži, Masáková, Pelantová (2007)]

- Recurrent dendric sets closed under reversal

  [Berthé, De Felice, Delecroix, D., Leroy, Perrin, Reutenauer, Rindone (2016)]



$$\mathcal{L}(\mathbf{f}) = \{\varepsilon, \mathtt{a}, \mathtt{b}, \mathtt{aa}, \mathtt{ab}, \mathtt{ba}, \mathtt{aab}, \mathtt{aba}, \mathtt{baa}, \mathtt{bab}, \ldots\}$$

Rich words and rich sets

An infinite word **w** is *rich* if all its finite prefixes are rich.
A factorial set is *rich* if all its elements are rich.

- Arnoux-Rauzy words

  [Droubay, Justin, Pirillo (2001)]

- Symmetric regular interval exchange sets

  [Baláži, Masáková, Pelantová (2007)]

- Recurrent dendric sets closed under reversal

  [Berthé, De Felice, Delecroix, D., Leroy, Perrin, Reutenauer, Rindone (2016)]

- Complementary-symmetric Rote words

  [Blondin-Massé, Brlek, Labbé, Vuillon (2011)]

- Languages closed under reversal with factor complexity $\mathcal{C}(n) = 2n + 1$

  [Balková, Pelantová, Starosta (2009)]

- etc.

Děkuji za pozornost!