# *A friendly introduction to Combinatorics on Words*

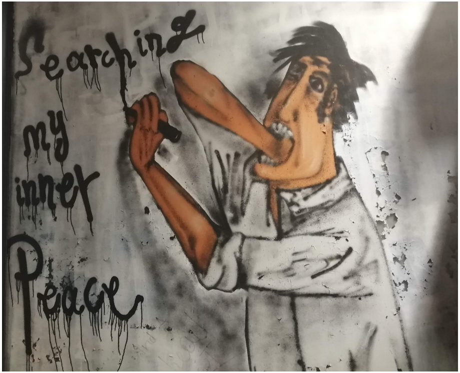
Francesco Dolce

*Janov nad Nisou, 17. května 2024*

*In the beginning was the Word*

# *In the beginning was the Word*

- strč, prst, skrz, krk



# *In the beginning was the Word*

- strč, prst, skrz, krk
- ACGA, TACGGACATTA, CATATACG


As, for instance, on Veronika's talk on Monday.

# *In the beginning was the Word*

- strč, prst, skrz, krk
- ACGA, TACGGACATTA, CATATACG
- 01, 100010, 0100101

# *In the beginning was the Word*

- strč, prst, skrz, krk
- ACGA, TACGGACATTA, CATATACG
- 01, 100010, 0100101
- ♡♡♣ ♢ ♠♡ ♠
- ...

# *Let's start with the ABC*

**Definition**

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

**Example**

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

# *Lenght matters*

**Definition**

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

**Example**

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

# *Lenght 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 $|w|_a = \#\{a\text{'s in } w\}$.

**Example**

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

# *Lenght 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 $|w|_a = \#\{a\text{'s in } w\}$.

**Example**

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

**Proposition**

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

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

# *Lenght 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 $|w|_a = \#\{a\text{'s in } w\}$.

**Example**

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

**Proposition**

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

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

# *The start, the end, 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 $w$,
- $s$ is a *suffix* of $w$,
- if $p, s \in \mathcal{A}^+$, $f$ is an *internal factor* of $w$.

## Example

$w =$ `nejblbější`

# *The start, the end, 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 $w$,
- $s$ is a *suffix* of $w$,
- if $p, s \in \mathcal{A}^+$, $f$ is an *internal factor* of $w$.

**Example**

$w =$ `nejblbější`

**Example**

The factors of `abaa` are :

# *The start, the end, 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 $w$,
- $s$ is a *suffix* of $w$,
- if $p, s \in \mathcal{A}^+$, $f$ is an *internal factor* of $w$.

**Example**

$$w = \texttt{nejblbější}$$

**Example**

The factors of abaa are : $\varepsilon$, a, b, aa, ab, ba, aba, baa, $w =$ abaa.

# Languages

**Definition**

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

**Example**

- $\mathcal{L}_0 = \{\varepsilon\}$,
- $\mathcal{L}_1 = \{\mathtt{a}, \mathtt{b}, \mathtt{aba}, \mathtt{bb}\}$,
- $\mathcal{L}_2 = \{w \in \mathcal{A}^* \;:\; |w| < 10\}$,
- $\mathcal{L}_3 = ab^*a = \{\mathtt{aa}, \mathtt{aba}, \mathtt{abba}, \mathtt{abbba}, \mathtt{abbba}, \ldots\}$,
- $\mathcal{L}_4 = \{w \in \mathcal{A}^* \;:\; |w|_b = 1\}$,

# 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**

- $\mathcal{L}_0 = \{\varepsilon\}$,
- $\mathcal{L}_1 = \{\mathrm{a}, \mathrm{b}, \mathrm{aba}, \mathrm{bb}\}$,
- $\mathcal{L}_2 = \{w \in \mathcal{A}^* \ : \ |w| < 10\}$,
- $\mathcal{L}_3 = ab^*a = \{\mathrm{aa}, \mathrm{aba}, \mathrm{abba}, \mathrm{abbba}, \mathrm{abbba}, \dots\}$,
- $\mathcal{L}_4 = \{w \in \mathcal{A}^* \ : \ |w|_b = 1\}$,
- $\mathcal{L}(\mathrm{abaa}) = \{\varepsilon, \ \mathrm{a}, \ \mathrm{b}, \ \mathrm{aa}, \ \mathrm{ab}, \ \mathrm{ba}, \ \mathrm{aba}, \ \mathrm{baa}, \ w\}$

# 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

- $\mathcal{L}_0 = \{\varepsilon\}$,
- $\mathcal{L}_1 = \{\mathsf{a}, \mathsf{b}, \mathsf{aba}, \mathsf{bb}\}$,
- $\mathcal{L}_2 = \{w \in \mathcal{A}^* \ : \ |w| < 10\}$,
- $\mathcal{L}_3 = ab^*a = \{\mathsf{aa}, \mathsf{aba}, \mathsf{abba}, \mathsf{abbba}, \mathsf{abbba}, \ldots\}$,
- $\mathcal{L}_4 = \{w \in \mathcal{A}^* \ : \ |w|_b = 1\}$,
- $\mathcal{L}(\mathsf{abaa}) = \{\varepsilon, \ \mathsf{a}, \ \mathsf{b}, \ \mathsf{aa}, \ \mathsf{ab}, \ \mathsf{ba}, \ \mathsf{aba}, \ \mathsf{baa}, \ w\}$

# *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**

- $\mathcal{L}_0 = \{\varepsilon\}$, ✓
- $\mathcal{L}_1 = \{\mathrm{a}, \mathrm{b}, \mathrm{aba}, \mathrm{bb}\}$,
- $\mathcal{L}_2 = \{w \in \mathcal{A}^* \; : \; |w| < 10\}$,
- $\mathcal{L}_3 = ab^*a = \{\mathrm{aa}, \mathrm{aba}, \mathrm{abba}, \mathrm{abbba}, \mathrm{abbba}, \ldots\}$,
- $\mathcal{L}_4 = \{w \in \mathcal{A}^* \; : \; |w|_b = 1\}$,
- $\mathcal{L}(\mathrm{abaa}) = \{\varepsilon, \; \mathrm{a}, \; \mathrm{b}, \; \mathrm{aa}, \; \mathrm{ab}, \; \mathrm{ba}, \; \mathrm{aba}, \; \mathrm{baa}, \; w\}$

# 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**

- $\mathcal{L}_0 = \{\varepsilon\}$, ✓
- $\mathcal{L}_1 = \{\mathrm{a}, \mathrm{b}, \mathrm{aba}, \mathrm{bb}\}$, ✗ $\mathrm{aba} \in \mathcal{L}_1$ but $\mathrm{ab} \notin \mathcal{L}_1$
- $\mathcal{L}_2 = \{w \in \mathcal{A}^* \ : \ |w| < 10\}$,
- $\mathcal{L}_3 = ab^*a = \{\mathrm{aa}, \mathrm{aba}, \mathrm{abba}, \mathrm{abbba}, \mathrm{abbba}, \ldots\}$,
- $\mathcal{L}_4 = \{w \in \mathcal{A}^* \ : \ |w|_b = 1\}$,
- $\mathcal{L}(\mathrm{abaa}) = \{\varepsilon,\ \mathrm{a},\ \mathrm{b},\ \mathrm{aa},\ \mathrm{ab},\ \mathrm{ba},\ \mathrm{aba},\ \mathrm{baa},\ w\}$

# 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**

- $\mathcal{L}_0 = \{\varepsilon\}$, ✓
- $\mathcal{L}_1 = \{\mathrm{a}, \mathrm{b}, \mathrm{aba}, \mathrm{bb}\}$, ✗ $\mathrm{aba} \in \mathcal{L}_1$ but $\mathrm{ab} \notin \mathcal{L}_1$
- $\mathcal{L}_2 = \{w \in \mathcal{A}^* \ : \ |w| < 10\}$, ✓
- $\mathcal{L}_3 = ab^*a = \{\mathrm{aa}, \mathrm{aba}, \mathrm{abba}, \mathrm{abbba}, \mathrm{abbba}, \ldots\}$,
- $\mathcal{L}_4 = \{w \in \mathcal{A}^* \ : \ |w|_b = 1\}$,
- $\mathcal{L}(\mathrm{abaa}) = \{\varepsilon,\ \mathrm{a},\ \mathrm{b},\ \mathrm{aa},\ \mathrm{ab},\ \mathrm{ba},\ \mathrm{aba},\ \mathrm{baa},\ w\}$

# *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**

- $\mathcal{L}_0 = \{\varepsilon\}$, ✓
- $\mathcal{L}_1 = \{\mathrm{a}, \mathrm{b}, \mathrm{aba}, \mathrm{bb}\}$, ✗ $\mathrm{aba} \in \mathcal{L}_1$ but $\mathrm{ab} \notin \mathcal{L}_1$
- $\mathcal{L}_2 = \{w \in \mathcal{A}^* \ : \ |w| < 10\}$, ✓
- $\mathcal{L}_3 = ab^*a = \{\mathrm{aa}, \mathrm{aba}, \mathrm{abba}, \mathrm{abbba}, \mathrm{abbba}, \ldots\}$, ✗ $\mathrm{aba} \in \mathcal{L}_3$ but $\mathrm{b} \notin \mathcal{L}_3$
- $\mathcal{L}_4 = \{w \in \mathcal{A}^* \ : \ |w|_b = 1\}$,
- $\mathcal{L}(\mathrm{abaa}) = \{\varepsilon,\ \mathrm{a},\ \mathrm{b},\ \mathrm{aa},\ \mathrm{ab},\ \mathrm{ba},\ \mathrm{aba},\ \mathrm{baa},\ w\}$

# 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**

- $\mathcal{L}_0 = \{\varepsilon\}$, ✓
- $\mathcal{L}_1 = \{\mathrm{a}, \mathrm{b}, \mathrm{aba}, \mathrm{bb}\}$, ✗ $\mathrm{aba} \in \mathcal{L}_1$ but $\mathrm{ab} \notin \mathcal{L}_1$
- $\mathcal{L}_2 = \{w \in \mathcal{A}^* \ : \ |w| < 10\}$, ✓
- $\mathcal{L}_3 = ab^*a = \{\mathrm{aa}, \mathrm{aba}, \mathrm{abba}, \mathrm{abbba}, \mathrm{abbba}, \ldots\}$, ✗ $\mathrm{aba} \in \mathcal{L}_3$ but $\mathrm{b} \notin \mathcal{L}_3$
- $\mathcal{L}_4 = \{w \in \mathcal{A}^* \ : \ |w|_b = 1\}$, ✗ $\mathrm{ab} \in \mathcal{L}_4$ but $\mathrm{a} \notin \mathcal{L}_4$
- $\mathcal{L}(\mathrm{abaa}) = \{\varepsilon,\ \mathrm{a},\ \mathrm{b},\ \mathrm{aa},\ \mathrm{ab},\ \mathrm{ba},\ \mathrm{aba},\ \mathrm{baa},\ w\}$

# *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**

- $\mathcal{L}_0 = \{\varepsilon\}$, ✓
- $\mathcal{L}_1 = \{\mathrm{a}, \mathrm{b}, \mathrm{aba}, \mathrm{bb}\}$, ✗ $\mathrm{aba} \in \mathcal{L}_1$ but $\mathrm{ab} \notin \mathcal{L}_1$
- $\mathcal{L}_2 = \{w \in \mathcal{A}^* \ : \ |w| < 10\}$, ✓
- $\mathcal{L}_3 = ab^*a = \{\mathrm{aa}, \mathrm{aba}, \mathrm{abba}, \mathrm{abbba}, \mathrm{abbba}, \ldots\}$, ✗ $\mathrm{aba} \in \mathcal{L}_3$ but $\mathrm{b} \notin \mathcal{L}_3$
- $\mathcal{L}_4 = \{w \in \mathcal{A}^* \ : \ |w|_b = 1\}$, ✗ $\mathrm{ab} \in \mathcal{L}_4$ but $\mathrm{a} \notin \mathcal{L}_4$
- $\mathcal{L}(\mathrm{abaa}) = \{\varepsilon, \ \mathrm{a}, \ \mathrm{b}, \ \mathrm{aa}, \ \mathrm{ab}, \ \mathrm{ba}, \ \mathrm{aba}, \ \mathrm{baa}, \ w\}$ ✓ by construction

# *Factor complexity*

**Definition**

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

**Example**

$$w = \texttt{abaccb}$$

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_w(n)$ | | | | | | | | | | | |

$\mathcal{L}(w) = \{ \qquad \}$

# *Factor complexity*

**Definition**

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

**Example**

$$w = \texttt{abaccb}$$

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_w(n)$ | 1 | | | | | | | | | | |

$\mathcal{L}(w) = \{\varepsilon,$ $\}$

# *Factor complexity*

## Definition

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

## Example

$$w = \texttt{abaccb}$$

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_w(n)$ | 1 | 3 | | | | | | | | | |

$$\mathcal{L}(w) = \{\varepsilon, \underbrace{\texttt{a}, \texttt{b}, \texttt{c}}_{3}, \qquad \}$$

# Factor complexity

## Definition

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

## Example

$$w = \texttt{abaccb}$$

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_w(n)$ | 1 | 3 | 5 | | | | | | | | |

$$\mathcal{L}(w) = \{\varepsilon, \underbrace{\texttt{a}, \texttt{b}, \texttt{c}}_{3}, \underbrace{\texttt{ab}, \texttt{ac}, \texttt{ba}, \texttt{cb}, \texttt{cc}}_{5}, \quad \}$$

# *Factor complexity*

**Definition**

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

**Example**

$$w = \mathtt{abaccb}$$

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_w(n)$ | 1 | 3 | 5 | 4 | | | | | | | |

$$\mathcal{L}(w) = \{\varepsilon, \underbrace{\mathtt{a}, \mathtt{b}, \mathtt{c}}_{3}, \underbrace{\mathtt{ab}, \mathtt{ac}, \mathtt{ba}, \mathtt{cb}, \mathtt{cc}}_{5}, \underbrace{\mathtt{aba}, \mathtt{acc}, \mathtt{bac}, \mathtt{ccb}}_{4}, \qquad \}$$

# Factor complexity

**Definition**

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

**Example**

$$w = \mathtt{abaccb}$$

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_w(n)$ | 1 | 3 | 5 | 4 | 3 | | | | | | |

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

# Factor complexity

**Definition**

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

**Example**

$$w = \mathtt{abaccb}$$

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_w(n)$ | 1 | 3 | 5 | 4 | 3 | 2 | | | | | |

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

# *Factor complexity*

## Definition

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

## Example

$$w = \mathtt{abaccb}$$

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_w(n)$ | 1 | 3 | 5 | 4 | 3 | 2 | 1 | | | | |

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

# *Factor complexity*

**Definition**

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

**Example**

$$w = \mathtt{abaccb}$$

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\cdots$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_w(n)$ | 1 | 3 | 5 | 4 | 3 | 2 | 1 | 0 | 0 | 0 | $\cdots$ |

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

# *Special factors*

**Definition**

An element $u$ of a language $\mathcal{L}$ is *right-special* if $ua, ub \in \mathcal{L}$ for two distinct letters $a, b$.

Similar definition for *left-special*.

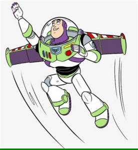
A factor is *bispecial* if it is both left- and right-special.
An *ordinary* factor is a factor that is not bispecial.

**Example**

Let $w = \texttt{abaccb}$.

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

# *Infinite words*



## Definition

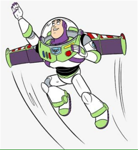
An *infinite word* is a sequence $\mathbf{w} = a_0 a_1 a_2 \cdots$, with $a_i$ letters.

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

## Example

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

# *Infinite words*



**Definition**

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

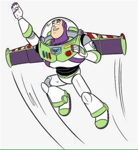
The set of all (right-)infinite words over $\mathcal{A}$ is denoted $\mathcal{A}^{\mathbb{N}}$.

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

**Example**

- $\mathbf{w} = \mathtt{aabaaaaaaaaaaaaaa} \cdots \in \{\mathtt{a}, \mathtt{b}\}^{\mathbb{N}}$;
- aab is a proper prefix,
- $\mathbf{w}$ is a prefix,
- baa is an internal factor,
- $\mathtt{abaaaaaa} \cdots = aba^{\omega}$ is a suffix,

# *Infinite words*



**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 $\mathcal{A}^{\mathbb{N}}$.

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

**Example**

- $\mathbf{w} = \mathtt{aabaaaaaaaaaaaaaa} \cdots \in \{\mathtt{a}, \mathtt{b}\}^{\mathbb{N}}$;
- aab is a proper prefix,
- $\mathbf{w}$ is a prefix,
- baa is an internal factor,
- $\mathtt{abaaaaaa} \cdots = aba^{\omega}$ is a suffix,
- $p_{\mathbf{w}}(n) = 4$ for every $n \geq 3$ (Exercise !)

# *Periodic words*

**Definition**

An infinite word of the form $uv^\omega$, with $u \in \mathcal{A}^*,\ v \in \mathcal{A}^+$ is (*eventually*) *periodic*.
If $u = \varepsilon$, it is *purely periodic*.

We can also extend the *factor complexity* to infinite words.

# Periodic words

**Definition**

An infinite word of the form $uv^\omega$, with $u \in \mathcal{A}^*$, $v \in \mathcal{A}^+$ is (*eventually*) *periodic*.
If $u = \varepsilon$, it is *purely periodic*.

We can also extend the *factor complexity* to infinite words.

**Theorem** [Morse, Hedlund (1938))]

An infinite word $\mathbf{w}$ is eventually periodic iff for some $n$ we have $p_{\mathbf{w}}(n) = p_{\mathbf{w}}(n+1)$.

# Periodic words

**Definition**

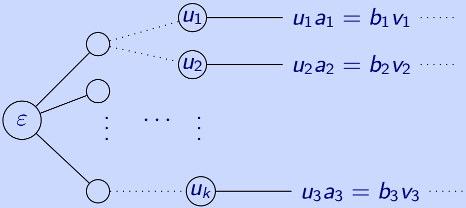
An infinite word of the form $uv^\omega$, with $u \in \mathcal{A}^*$, $v \in \mathcal{A}^+$ is (*eventually*) *periodic*.
If $u = \varepsilon$, it is *purely periodic*.

We can also extend the *factor complexity* to infinite words.

**Theorem [Morse, Hedlund (1938))]**

An infinite word $\mathbf{w}$ is eventually periodic iff for some $n$ we have $p_{\mathbf{w}}(n) = p_{\mathbf{w}}(n+1)$.

Proof.



# *Periodic words*

**Definition**

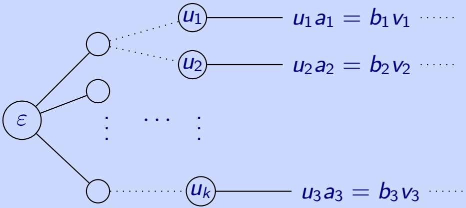
An infinite word of the form $uv^\omega$, with $u \in \mathcal{A}^*$, $v \in \mathcal{A}^+$ is (*eventually*) *periodic*.
If $u = \varepsilon$, it is *purely periodic*.

We can also extend the *factor complexity* to infinite words.

**Theorem** [Morse, Hedlund (1938))]

An infinite word $\mathbf{w}$ is eventually periodic iff for some $n$ we have $p_\mathbf{w}(n) = p_\mathbf{w}(n+1)$.

Proof.





Want to check whether this proof is correct ?
See Štěpán's talk later !

# *Periodic words*

**Definition**

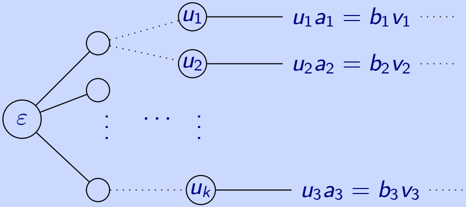
An infinite word of the form $uv^\omega$, with $u \in \mathcal{A}^*$, $v \in \mathcal{A}^+$ is (*eventually*) *periodic*.
If $u = \varepsilon$, it is *purely periodic*.

We can also extend the *factor complexity* to infinite words.

**Theorem** [Morse, Hedlund (1938))]

An infinite word $\mathbf{w}$ is eventually periodic iff for some $n$ we have $p_\mathbf{w}(n) = p_\mathbf{w}(n+1)$.

Proof.





Want to check whether this proof is correct ?
See Štěpán's talk later !


Can a (not necessarily periodic) word have repeated factors ?
More on Daniela's talk tomorrow !

# 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**

$$\psi_1 : \begin{cases} \mathtt{a} \mapsto \mathtt{010} \\ \mathtt{b} \mapsto \mathtt{1} \end{cases} , \qquad \psi_2 : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{b} \end{cases} \qquad \psi_3 : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{a} \end{cases}$$

# 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 $\psi : \mathcal{A}^* \to \mathcal{A}^*$ s.t there exists a letter $a \in \mathcal{A}$ with

- $\psi(a) = as$ and
- $\lim_{n\to\infty} |\psi^n(a)| = \infty$.

The word $\lim_{n\to\infty} \psi^n(a)$ is a *fixed point* of the substitution.

**Example**

$$\psi_1 : \begin{cases} \mathtt{a} \mapsto \mathtt{010} \\ \mathtt{b} \mapsto \mathtt{1} \end{cases}, \qquad \psi_2 : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{b} \end{cases} \qquad \psi_3 : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{a} \end{cases}$$

# *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 $\psi : \mathcal{A}^* \to \mathcal{A}^*$ s.t there exists a letter $a \in \mathcal{A}$ with

- $\psi(a) = as$ and
- $\lim_{n\to\infty} |\psi^n(a)| = \infty$.

The word $\lim_{n\to\infty} \psi^n(a)$ is a *fixed point* of the substitution.

**Example**

$$\psi_1 : \begin{cases} \mathtt{a} \mapsto \mathtt{010} \\ \mathtt{b} \mapsto \mathtt{1} \end{cases}, \qquad \psi_2 : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{b} \end{cases} \qquad \psi_3 : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{a} \end{cases}$$

A substitution $\psi$ is *primitive* if there is a $k$ such that $b \in \mathcal{L}(\psi^k(a))$ for every $a, b \in \mathcal{A}$.

# *Thue-Morse (and many others)*

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

$$\mathbf{x} = \mathtt{abbabaabbaababbabaababbaabbabaabbaababbaabbabaa}\cdots$$

of the morphism

$$\tau : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{ba} \end{cases} .$$

# *Thue-Morse (and many others)*

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

$$\mathbf{x} = \texttt{abbabaabbaababbabaababbaabbabaabbaababbaabbabaa}\cdots$$

of the morphism

$$\tau : \begin{cases} \texttt{a} \mapsto \texttt{ab} \\ \texttt{b} \mapsto \texttt{ba} \end{cases} .$$

**Definition**

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

# *Thue-Morse (and many others)*

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

$$\mathbf{x} = \mathtt{abbabaabbaababbabaababbaabbabaabbaababbaabbabaa}\cdots$$

of the morphism

$$\tau : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{ba} \end{cases} .$$

**Definition**

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


Do you want to know more about this sequence ?
Maaany occasions to do so : see Herman's talk later, Samuel's one tomorrow, Martina's one on Monday, and probably on others too !

# *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 $|\varphi^n(\mathtt{a})|$, i.e., $1, 2, 3, 5, 8, \ldots$ are the *Fibonacci numbers*.

# $S$-adic words

**Definition**

An infinite word $\mathbf{w}$ is said to be $S$-*adic* if there is a sequence of morphisms $\mathbf{s} = (\sigma_n : \mathcal{A}^* \to \mathcal{A}^*)_n$ and a sequence of letters $\mathbf{a} = (a_n \in \mathcal{A})_n$ such that

$$\mathbf{w} = \lim_{n \to \infty} \sigma_0 \sigma_1 \cdots \sigma_n (a_{n+1}).$$

The pair $(\mathbf{s}, \mathbf{a})$ is called an $S$-*adic representation* of $\mathbf{w}$.

**Example**

$$(\mathbf{s}, \mathbf{a}) = ((\varphi, \tau, \varphi, \tau, \ldots), (\mathtt{a}, \mathtt{a}, \mathtt{a}, \ldots)) \quad \text{where} \quad \varphi : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{a} \end{cases}, \quad \tau : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{ba} \end{cases}.$$

# $S$-adic words

**Definition**

An infinite word $\mathbf{w}$ is said to be $S$-*adic* if there is a sequence of morphisms $\mathbf{s} = (\sigma_n : \mathcal{A}^* \to \mathcal{A}^*)_n$ and a sequence of letters $\mathbf{a} = (a_n \in \mathcal{A})_n$ such that

$$\mathbf{w} = \lim_{n\to\infty} \sigma_0\sigma_1\cdots\sigma_n(a_{n+1}).$$

The pair $(\mathbf{s}, \mathbf{a})$ is called an $S$-*adic representation* of $\mathbf{w}$.

**Example**

$$(\mathbf{s}, \mathbf{a}) = ((\varphi, \tau, \varphi, \tau, \ldots), (\mathtt{a}, \mathtt{a}, \mathtt{a}, \ldots)) \quad \text{where} \quad \varphi : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{a} \end{cases}, \quad \tau : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{ba} \end{cases}.$$

$\varphi(\mathtt{a}) = \mathtt{ab}$

# $S$-adic words

**Definition**

An infinite word $\mathbf{w}$ is said to be *$S$-adic* if there is a sequence of morphisms $\mathbf{s} = (\sigma_n : \mathcal{A}^* \to \mathcal{A}^*)_n$ and a sequence of letters $\mathbf{a} = (a_n \in \mathcal{A})_n$ such that

$$\mathbf{w} = \lim_{n\to\infty} \sigma_0\sigma_1\cdots\sigma_n(a_{n+1}).$$

The pair $(\mathbf{s}, \mathbf{a})$ is called an *$S$-adic representation* of $\mathbf{w}$.

**Example**

$$(\mathbf{s}, \mathbf{a}) = ((\varphi, \tau, \varphi, \tau, \ldots), (\mathrm{a}, \mathrm{a}, \mathrm{a}, \ldots)) \quad \text{where} \quad \varphi : \begin{cases} \mathrm{a} \mapsto \mathrm{ab} \\ \mathrm{b} \mapsto \mathrm{a} \end{cases}, \quad \tau : \begin{cases} \mathrm{a} \mapsto \mathrm{ab} \\ \mathrm{b} \mapsto \mathrm{ba} \end{cases}.$$

$\varphi(\mathrm{a}) = \mathrm{ab}$

$\varphi \circ \tau(\mathrm{a}) = \varphi(\mathrm{ab}) = \mathrm{aba}$

# $S$-adic words

**Definition**

An infinite word $\mathbf{w}$ is said to be *$S$-adic* if there is a sequence of morphisms $\mathbf{s} = (\sigma_n : \mathcal{A}^* \to \mathcal{A}^*)_n$ and a sequence of letters $\mathbf{a} = (a_n \in \mathcal{A})_n$ such that

$$\mathbf{w} = \lim_{n \to \infty} \sigma_0 \sigma_1 \cdots \sigma_n(a_{n+1}).$$

The pair $(\mathbf{s}, \mathbf{a})$ is called an *$S$-adic representation* of $\mathbf{w}$.

**Example**

$$(\mathbf{s}, \mathbf{a}) = ((\varphi, \tau, \varphi, \tau, \ldots), (\mathrm{a}, \mathrm{a}, \mathrm{a}, \ldots)) \quad \text{where} \quad \varphi : \begin{cases} \mathrm{a} \mapsto \mathrm{ab} \\ \mathrm{b} \mapsto \mathrm{a} \end{cases}, \quad \tau : \begin{cases} \mathrm{a} \mapsto \mathrm{ab} \\ \mathrm{b} \mapsto \mathrm{ba} \end{cases}.$$

$\varphi(\mathrm{a}) = \mathrm{ab}$
$\varphi \circ \tau(\mathrm{a}) = \varphi(\mathrm{ab}) = \mathrm{aba}$
$\varphi \circ \tau \circ \varphi(\mathrm{a}) = \varphi(\tau(\mathrm{ab})) = \varphi(\mathrm{abba}) = \mathrm{abaaab}$

# $S$-adic words

**Definition**

An infinite word $\mathbf{w}$ is said to be $S$-*adic* if there is a sequence of morphisms $\mathbf{s} = (\sigma_n : \mathcal{A}^* \to \mathcal{A}^*)_n$ and a sequence of letters $\mathbf{a} = (a_n \in \mathcal{A})_n$ such that

$$\mathbf{w} = \lim_{n\to\infty} \sigma_0\sigma_1\cdots\sigma_n(a_{n+1}).$$

The pair $(\mathbf{s}, \mathbf{a})$ is called an $S$-*adic representation* of $\mathbf{w}$.

**Example**

$$(\mathbf{s},\mathbf{a}) = ((\varphi,\tau,\varphi,\tau,\ldots),(\mathrm{a},\mathrm{a},\mathrm{a},\ldots)) \quad \text{where} \quad \varphi : \begin{cases} \mathrm{a} \mapsto \mathrm{ab} \\ \mathrm{b} \mapsto \mathrm{a} \end{cases}, \quad \tau : \begin{cases} \mathrm{a} \mapsto \mathrm{ab} \\ \mathrm{b} \mapsto \mathrm{ba} \end{cases}.$$

$\varphi(\mathrm{a}) = \mathrm{ab}$
$\varphi\circ\tau(\mathrm{a}) = \varphi(\mathrm{ab}) = \mathrm{aba}$
$\varphi\circ\tau\circ\varphi(\mathrm{a}) = \varphi(\tau(\mathrm{ab})) = \varphi(\mathrm{abba}) = \mathrm{abaaab}$
$\varphi\circ\tau\circ\varphi\circ\tau(\mathrm{a}) = \varphi(\tau(\varphi(\mathrm{ab}))) = \varphi(\tau(\mathrm{aba})) = \varphi(\mathrm{abbaab}) = \mathrm{abaaababa}$

# $S$-*adic words*

**Definition**

An infinite word $\mathbf{w}$ is said to be $S$-*adic* if there is a sequence of alphabets $(\mathcal{A}_n)_n$, a sequence of morphisms $\mathbf{s} = (\sigma_n : {\mathcal{A}_{n+1}}^* \to {\mathcal{A}_n}^*)_n$ and a sequence of letters $\mathbf{a} = (a_n \in \mathcal{A}_n)_n$ such that

$$\mathbf{w} = \lim_{n\to\infty} \sigma_0\sigma_1 \cdots \sigma_n(a_{n+1}).$$

The pair $(\mathbf{s}, \mathbf{a})$ is called an $S$-*adic representation* of $\mathbf{w}$.

**Example**

$$(\mathbf{s}, \mathbf{a}) = ((\varphi, \tau, \varphi, \tau, \ldots), (\mathtt{a}, \mathtt{a}, \mathtt{a}, \ldots)) \quad \text{where} \quad \varphi : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{a} \end{cases}, \quad \tau : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{ba} \end{cases}.$$

# $S$-adic words

**Definition**

An infinite word $\mathbf{w}$ is said to be *$S$-adic* if there is a sequence of alphabets $(\mathcal{A}_n)_n$, a sequence of morphisms $\mathbf{s} = (\sigma_n : {\mathcal{A}_{n+1}}^* \to {\mathcal{A}_n}^*)_n$ and a sequence of letters $\mathbf{a} = (a_n \in \mathcal{A}_n)_n$ such that

$$\mathbf{w} = \lim_{n\to\infty} \sigma_0\sigma_1\cdots\sigma_n(a_{n+1}).$$

The pair $(\mathbf{s}, \mathbf{a})$ is called an *$S$-adic representation* of $\mathbf{w}$.

**Example**

$$(\mathbf{s}, \mathbf{a}) = ((\varphi, \tau, \varphi, \tau, \ldots), (\mathtt{a}, \mathtt{a}, \mathtt{a}, \ldots)) \quad \text{where} \quad \varphi : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{a} \end{cases}, \quad \tau : \begin{cases} \mathtt{a} \mapsto \mathtt{ab} \\ \mathtt{b} \mapsto \mathtt{ba} \end{cases}.$$

The pair $(\mathbf{s}, \mathbf{a})$ is (*purely*) *periodic* if $(\sigma_{m+n}, a_{m+n}) = (\sigma_m, a_m)$ for all $m$.

It is *primitive* if for all $r \geq 0$ there is $r' > r$ s.t. all letters of $\mathcal{A}_r$ occur in $\sigma_r\sigma_{r+1}\cdots\sigma_{r'}(a)$ for all $a \in \mathcal{A}_{r'+1}$.

# *Sturmian words*

**Definition**

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

**Example**

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

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

# *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 length 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 length only one factor is right special ; and
3. its set of factors is closed under reversal.

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

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

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

# *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 length only one factor is right special ; and
3. its set of factors is closed under reversal.

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

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

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

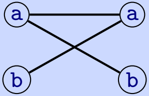
# *Dendric words*

The *extension graph* of a word $w \in \mathcal{L}$ is the undirected bipartite graph $\mathcal{E}(w)$ with vertices $L(w) \sqcup R(w)$ and edges $B(w)$, where

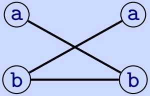
$$
\begin{aligned}
L(w) &= \{u \in \mathcal{A} \mid uw \in \mathcal{L}\} \\
R(w) &= \{v \in \mathcal{A} \mid wv \in \mathcal{L}\} \\
B(w) &= \{(u, v) \in \mathcal{A} \times \mathcal{A} \mid uwv \in \mathcal{L}\}
\end{aligned}
$$

**Example (Fibonacci, $\mathcal{L} = \{\varepsilon, \text{a}, \text{b}, \text{aa}, \text{ab}, \text{ba}, \text{aab}, \text{aba}, \text{baa}, \text{bab}, \ldots\}$)**

$\mathcal{E}(\varepsilon)$




$\mathcal{E}(a)$



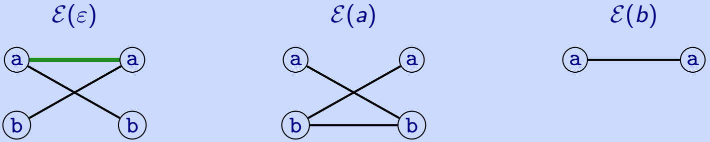

$\mathcal{E}(b)$

# *Dendric words*

The *extension graph* of a word $w \in \mathcal{L}$ is the undirected bipartite graph $\mathcal{E}(w)$ with vertices $L(w) \sqcup R(w)$ and edges $B(w)$, where

$$\begin{aligned} L(w) &= \{u \in \mathcal{A} \mid uw \in \mathcal{L}\} \\ R(w) &= \{v \in \mathcal{A} \mid wv \in \mathcal{L}\} \\ B(w) &= \{(u, v) \in \mathcal{A} \times \mathcal{A} \mid uwv \in \mathcal{L}\} \end{aligned}$$

**Example (Fibonacci, $\mathcal{L} = \{\varepsilon, \text{a}, \text{b}, \text{aa}, \text{ab}, \text{ba}, \text{aab}, \text{aba}, \text{baa}, \text{bab}, \ldots\}$)**

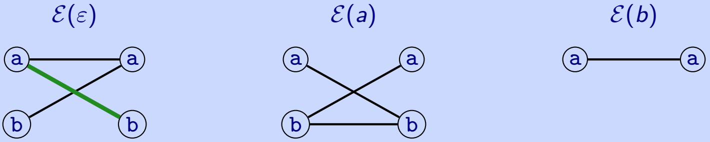


# Dendric words

The *extension graph* of a word $w \in \mathcal{L}$ is the undirected bipartite graph $\mathcal{E}(w)$ with vertices $L(w) \sqcup R(w)$ and edges $B(w)$, where

$$\begin{aligned} L(w) &= \{u \in \mathcal{A} \,|\, uw \in \mathcal{L}\} \\ R(w) &= \{v \in \mathcal{A} \,|\, wv \in \mathcal{L}\} \\ B(w) &= \{(u,v) \in \mathcal{A} \times \mathcal{A} \,|\, uwv \in \mathcal{L}\} \end{aligned}$$

**Example (Fibonacci, $\mathcal{L} = \{\varepsilon, \mathrm{a}, \mathrm{b}, \mathrm{aa}, \mathrm{ab}, \mathrm{ba}, \mathrm{aab}, \mathrm{aba}, \mathrm{baa}, \mathrm{bab}, \ldots\}$)**

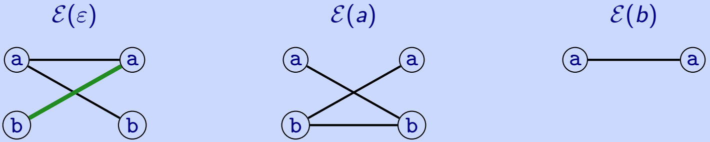


# Dendric words

The *extension graph* of a word $w \in \mathcal{L}$ is the undirected bipartite graph $\mathcal{E}(w)$ with vertices $L(w) \sqcup R(w)$ and edges $B(w)$, where

$$\begin{aligned} L(w) &= \{u \in \mathcal{A} \mid uw \in \mathcal{L}\} \\ R(w) &= \{v \in \mathcal{A} \mid wv \in \mathcal{L}\} \\ B(w) &= \{(u, v) \in \mathcal{A} \times \mathcal{A} \mid uwv \in \mathcal{L}\} \end{aligned}$$

**Example (Fibonacci, $\mathcal{L} = \{\varepsilon, \mathrm{a}, \mathrm{b}, \mathrm{aa}, \mathrm{ab}, \mathrm{ba}, \mathrm{aab}, \mathrm{aba}, \mathrm{baa}, \mathrm{bab}, \ldots\}$)**



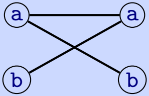
# *Dendric words*

The *extension graph* of a word $w \in \mathcal{L}$ is the undirected bipartite graph $\mathcal{E}(w)$ with vertices $L(w) \sqcup R(w)$ and edges $B(w)$, where

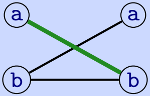
$$\begin{aligned} L(w) &= \{u \in \mathcal{A} \,|\, uw \in \mathcal{L}\} \\ R(w) &= \{v \in \mathcal{A} \,|\, wv \in \mathcal{L}\} \\ B(w) &= \{(u, v) \in \mathcal{A} \times \mathcal{A} \,|\, uwv \in \mathcal{L}\} \end{aligned}$$

**Example (Fibonacci, $\mathcal{L} = \{\varepsilon, \mathrm{a}, \mathrm{b}, \mathrm{aa}, \mathrm{ab}, \mathrm{ba}, \mathrm{aab}, \mathrm{aba}, \mathrm{baa}, \mathrm{bab}, \ldots\}$)**

$\mathcal{E}(\varepsilon)$




$\mathcal{E}(a)$




$\mathcal{E}(b)$

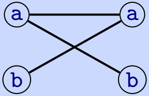
# *Dendric words*

The *extension graph* of a word $w \in \mathcal{L}$ is the undirected bipartite graph $\mathcal{E}(w)$ with vertices $L(w) \sqcup R(w)$ and edges $B(w)$, where

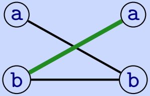
$$\begin{aligned} L(w) &= \{u \in \mathcal{A} \mid uw \in \mathcal{L}\} \\ R(w) &= \{v \in \mathcal{A} \mid wv \in \mathcal{L}\} \\ B(w) &= \{(u,v) \in \mathcal{A} \times \mathcal{A} \mid uwv \in \mathcal{L}\} \end{aligned}$$

**Example (Fibonacci, $\mathcal{L} = \{\varepsilon, \mathrm{a}, \mathrm{b}, \mathrm{aa}, \mathrm{ab}, \mathrm{ba}, \mathrm{aab}, \mathrm{aba}, \mathrm{baa}, \mathrm{bab}, \ldots\}$)**

$\mathcal{E}(\varepsilon)$




$\mathcal{E}(a)$



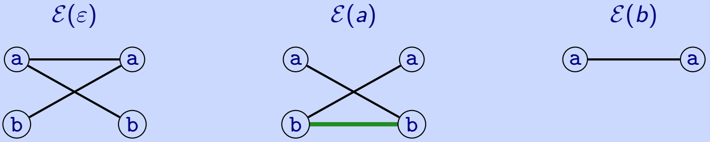

$\mathcal{E}(b)$

# Dendric words

The *extension graph* of a word $w \in \mathcal{L}$ is the undirected bipartite graph $\mathcal{E}(w)$ with vertices $L(w) \sqcup R(w)$ and edges $B(w)$, where

$$\begin{aligned} L(w) &= \{u \in \mathcal{A} \mid uw \in \mathcal{L}\} \\ R(w) &= \{v \in \mathcal{A} \mid wv \in \mathcal{L}\} \\ B(w) &= \{(u, v) \in \mathcal{A} \times \mathcal{A} \mid uwv \in \mathcal{L}\} \end{aligned}$$

**Example (Fibonacci, $\mathcal{L} = \{\varepsilon, \text{a}, \text{b}, \text{aa}, \text{ab}, \text{ba}, \text{aab}, \text{aba}, \text{baa}, \text{bab}, \ldots\}$)**



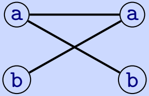
# *Dendric words*

The *extension graph* of a word $w \in \mathcal{L}$ is the undirected bipartite graph $\mathcal{E}(w)$ with vertices $L(w) \sqcup R(w)$ and edges $B(w)$, where

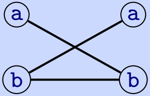
$$
\begin{aligned}
L(w) &= \{u \in \mathcal{A} \mid uw \in \mathcal{L}\} \\
R(w) &= \{v \in \mathcal{A} \mid wv \in \mathcal{L}\} \\
B(w) &= \{(u, v) \in \mathcal{A} \times \mathcal{A} \mid uwv \in \mathcal{L}\}
\end{aligned}
$$

**Example (Fibonacci, $\mathcal{L} = \{\varepsilon, \mathrm{a}, \mathrm{b}, \mathrm{aa}, \mathrm{ab}, \mathrm{ba}, \mathrm{aab}, \mathrm{aba}, \mathrm{baa}, \mathrm{bab}, \dots\}$)**

$\mathcal{E}(\varepsilon)$




$\mathcal{E}(a)$




$\mathcal{E}(b)$

# Dendric words

The *extension graph* of a word $w \in \mathcal{L}$ is the undirected bipartite graph $\mathcal{E}(w)$ with vertices $L(w) \sqcup R(w)$ and edges $B(w)$, where

$$\begin{aligned} L(w) &= \{u \in \mathcal{A} \mid uw \in \mathcal{L}\} \\ R(w) &= \{v \in \mathcal{A} \mid wv \in \mathcal{L}\} \\ B(w) &= \{(u,v) \in \mathcal{A} \times \mathcal{A} \mid uwv \in \mathcal{L}\} \end{aligned}$$

**Definition**

A language $\mathcal{L}$ is (purely) *dendric* if the graph $\mathcal{E}(w)$ is a tree for any $w \in \mathcal{L}$.

Sturmian words (and Arnoux-Rauzy) are dendric.

# *Recurrence and uniforme 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 $uwv \in \mathcal{L}$.

**Example (Fibonacci)**

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

# *Recurrence and uniforme 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 $uwv \in \mathcal{L}$.

$\mathcal{L}$ 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)**

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

# *Recurrence and uniforme 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 $uwv \in \mathcal{L}$.

$\mathcal{L}$ 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 uniforme 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 $uwv \in \mathcal{L}$.

$\mathcal{L}$ 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 $\mathtt{b}^m$.

# *Recurrence and uniforme 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 $uwv \in \mathcal{L}$.

$\mathcal{L}$ 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 $\mathtt{b}^m$.


What if we want a word "starting" and "ending" with $u$?
(see Herman's talk just after that!)

# *Palindromes*

**Definition**

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

**Example**

- kayak
- blb, krk, oko
- nepochopen

# *Palindromes*

**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*

**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*

**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

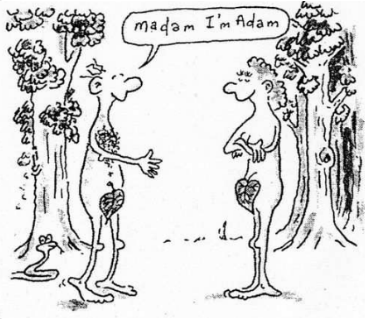
# *Palindromes*

**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
- Madam I'm Adam



# *Palindromes*

**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
- Madam I'm Adam
- 135797531
- Signate, signate, mere me tangis et angis
- ...

# *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*.

- $\mathcal{P}(\mathtt{ananas}) = \{\varepsilon, \mathtt{a}, \mathtt{n}, \mathtt{s}, \mathtt{ana}, \mathtt{nan}, \mathtt{anana}\}$
  $\#\mathcal{P}(w) = 7 = |w| + 1$ ✓

# 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*.

- $\mathcal{P}(\texttt{ananas}) = \{\varepsilon, \texttt{a}, \texttt{n}, \texttt{s}, \texttt{ana}, \texttt{nan}, \texttt{anana}\}$
  $\#\mathcal{P}(w) = 7 = |w| + 1 \qquad \checkmark$

- $\mathcal{P}(\texttt{pizza}) = \{\varepsilon, \texttt{a}, \texttt{i}, \texttt{p}, \texttt{z}, \texttt{zz}\}$
  $\#\mathcal{P}(w) = 6 = |w| + 1 \qquad \checkmark$

# *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*.

- $\mathcal{P}(\texttt{ananas}) = \{\varepsilon, \texttt{a}, \texttt{n}, \texttt{s}, \texttt{ana}, \texttt{nan}, \texttt{anana}\}$
  $\#\mathcal{P}(w) = 7 = |w| + 1$ ✓
- $\mathcal{P}(\texttt{pizza}) = \{\varepsilon, \texttt{a}, \texttt{i}, \texttt{p}, \texttt{z}, \texttt{zz}\}$
  $\#\mathcal{P}(w) = 6 = |w| + 1$ ✓
- $\mathcal{P}(\texttt{hawaiipizza}) = \{\varepsilon, \texttt{a}, \texttt{h}, \texttt{i}, \texttt{p}, \texttt{z}, \texttt{ii}, \texttt{zz}, \texttt{awa}\}$
  $\#\mathcal{P}(w) = 9 < 12 = |w| + 1$ ✗

# *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*.

- $\mathcal{P}(\texttt{ananas}) = \{\varepsilon, \texttt{a}, \texttt{n}, \texttt{s}, \texttt{ana}, \texttt{nan}, \texttt{anana}\}$
  $\#\mathcal{P}(w) = 7 = |w| + 1$ ✓

- $\mathcal{P}(\texttt{pizza}) = \{\varepsilon, \texttt{a}, \texttt{i}, \texttt{p}, \texttt{z}, \texttt{zz}\}$
  $\#\mathcal{P}(w) = 6 = |w| + 1$ ✓


- $\mathcal{P}(\texttt{hawaiipizza}) = \{\varepsilon, \texttt{a}, \texttt{h}, \texttt{i}, \texttt{p}, \texttt{z}, \texttt{ii}, \texttt{zz}, \texttt{awa}\}$
  $\#\mathcal{P}(w) = 9 < 12 = |w| + 1$ ✗

An infinite word (resp. factorial set) is *rich* if all its prefixes (resp. elements) are rich.


More on that on Lubka's talk tomorrow.

# Děkuji za pozornost !