

Bispecial factors in DOL systems

Journées combinatoires de Bordeaux

Herman Goulet-Ouellet, Karel Klouda and Štěpán Starosta 03.02.2025 Part 1



"In spite of the simplicity of the basic definitions, the theory of DOL systems is at present very rich and challenging. Apart from providing applications to formal languages and biology, this theory has also shed new light on the very basic mathematical notion of an endomorphism defined on a free monoid."

A. Salomaa and G. Rozenberg The mathematical theory of L systems, 1980 "In spite of the simplicity of the basic definitions, the theory of DOL systems is at present very rich and challenging. Apart from providing applications to **formal languages** and biology, this theory has also shed new light on the very basic mathematical notion of an endomorphism defined on a free monoid."

A. Salomaa and G. Rozenberg The mathematical theory of L systems, 1980



ALPHABET Finite set A of symbols (letters).



ALPHABET Finite set A of symbols (letters). FREE MONOID Set A* of all finite words over A (strings).



ALPHABET Finite set A of symbols (letters). FREE MONOID Set A* of all finite words over A (strings). LENGTH The length of a word $u \in A^*$ is denoted |u|.



ALPHABET Finite set A of symbols (letters). FREE MONOID Set A* of all finite words over A (strings). LENGTH The length of a word $u \in A^*$ is denoted |u|. EMPTY WORD The unique word $\varepsilon \in A^*$ of length 0.



ALPHABET Finite set A of symbols (letters). FREE MONOID Set A* of all finite words over A (strings). LENGTH The length of a word $u \in A^*$ is denoted |u|. EMPTY WORD The unique word $\varepsilon \in A^*$ of length 0. CONCATENATION Write uv for the concatenation of $u, v \in A^*$.



ALPHABET Finite set A of symbols (letters). FREE MONOID Set A* of all finite words over A (strings). LENGTH The length of a word $u \in A^*$ is denoted |u|. EMPTY WORD The unique word $\varepsilon \in A^*$ of length 0. CONCATENATION Write uv for the concatenation of $u, v \in A^*$. MORPHISM Map $\varphi : A^* \to B^*$ such that $\varphi(uv) = \varphi(u)\varphi(v)$.



D O L Deterministic Context-free Lindenmayer¹

$$S = (A, \varphi : A^* \to A^*, w \in A^*)$$

- A is an alphabet.
- $\varphi: A^* \to A^*$ is a **morphism**.
- $w \in A^*$ is a word (axiom).

¹Aristid Lindenmayer, 1925–1989, Hungarian biologist.

Examples



Example 1 (Fibonacci)

$$(\{a,b\}, \varphi: a \mapsto ab, b \mapsto a, a)$$



Example 1 (Fibonacci)

$$(\{a,b\}, \varphi: a \mapsto ab, b \mapsto a, a)$$

Example 2 (Thue–Morse)

$$(\{a,b\}, \varphi: a \mapsto ab, b \mapsto ba, a)$$



Example 1 (Fibonacci)

$$(\{a,b\}, \varphi: a \mapsto ab, b \mapsto a, a)$$

Example 2 (Thue–Morse)

$$(\{a,b\}, \varphi: a \mapsto ab, b \mapsto ba, a)$$

Example 3 (Quaternary system)

 $(\{a, b, c, d\}, \varphi: a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, c)$



Example 1 (Fibonacci)

$$(\{a,b\}, \varphi: a \mapsto ab, b \mapsto a, a)$$

Example 2 (Thue–Morse)

$$(\{a,b\}, \varphi: a \mapsto ab, b \mapsto ba, a)$$

Example 3 (Quaternary system)

 $(\{a, b, c, d\}, \varphi: a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, c)$

Example 4 (Ternary system)

 $(\{a, b, c\}, \varphi: a \mapsto abaca, b \mapsto aba, c \mapsto aba, a)$



$$S = (A, \varphi: A^* \to A^*, w \in A^*)$$

SEQUENCE $(\varphi^n(w))_{n \in \mathbb{N}}$. Language $\mathbf{L}(S) = \{ u \in A^* \mid \exists n \in \mathbb{N}, \exists w \in W, u \leq_{fac} \varphi^n(w) \}.$



$S = (\{a, b\}, \varphi: a \mapsto ab, b \mapsto a, a)$



$S = (\{a, b, c, d\}, \varphi: a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, c)$

SEQUENCE a, baa, adcbaabaa, baaadcdcadcbaabaaadcbaabaa,... LANGUAGE $L(S) = \{\varepsilon, a, b, aa, ab, ad, ba, ca, cb, cd, dc, aaa, aab, aad, aba, adc, baa, cad, cba, cdc, ... \}$



$S = (\{a, b, c, d\}, \varphi: a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, c)$

SEQUENCE a, baa, adcbaabaa, baaadcdcadcbaabaaadcbaabaa,... LANGUAGE $L(S) = \{\varepsilon, a, b, aa, ab, ad, ba, ca, cb, cd, dc,$

aaa, aab, aad, aba, adc, baa, cad, cba, cdc, ... }

• **L**(*S*) is computable (proof by SageMath).



$$S = (A, \varphi: A^* \to A^*, w \in A^*)$$

• What kinds of infinite words are generated by S (if any)?



$$S = (A, \varphi: A^* \to A^*, w \in A^*)$$

- What kinds of infinite words are generated by S (if any)?
- What is the structure of **L**(*S*)?



$$S = (A, \varphi: A^* \to A^*, w \in A^*)$$

- What kinds of infinite words are generated by S (if any)?
- What is the structure of **L**(*S*)?
- How many factors of length *n* are in **L**(*S*)?



$$S = (A, \varphi: A^* \to A^*, w \in A^*)$$

- What kinds of infinite words are generated by S (if any)?
- What is the structure of **L**(*S*)?
- How many factors of length n are in L(S)?
- Where and how do words occur in **L**(S)?



$$S = (A, \varphi: A^* \to A^*, w \in A^*)$$

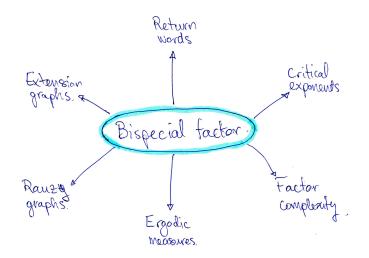
- What kinds of infinite words are generated by S (if any)?
- What is the structure of **L**(*S*)?
- How many factors of length n are in L(S)?
- Where and how do words occur in **L**(S)?
- Can words be extended in L(S) in a predictable way?



$$S = (A, \varphi: A^* \to A^*, w \in A^*)$$

- What kinds of infinite words are generated by S (if any)?
- What is the structure of **L**(*S*)?
- How many factors of length n are in L(S)?
- Where and how do words occur in L(S)?
- Can words be extended in L(S) in a predictable way?





7

Part 2





$$S = (A, \varphi: A^* \to A^*, w \in A^*)$$

• $u \in L(S)$ is left special if $\#\{a \in A \mid au \in L(S)\} \ge 2$.



$$S = (A, \varphi: A^* \to A^*, w \in A^*)$$

- $u \in \mathbf{L}(S)$ is left special if $\#\{a \in A \mid au \in \mathbf{L}(S)\} \ge 2$.
- $u \in \mathbf{L}(S)$ is right special if $\#\{a \in A \mid ua \in \mathbf{L}(S)\} \ge 2$.



$$S = (A, \varphi: A^* \to A^*, w \in A^*)$$

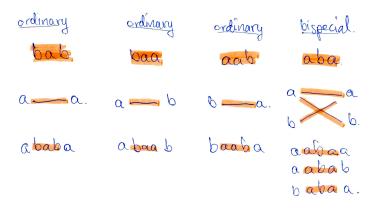
- $u \in \mathbf{L}(S)$ is left special if $\#\{a \in A \mid au \in \mathbf{L}(S)\} \ge 2$.
- $u \in L(S)$ is right special if $\#\{a \in A \mid ua \in L(S)\} \ge 2$.
- $u \in L(S)$ is **bispecial** if it is both left and right special.

Example



Fibonacci

$$S = (\{a, b\}, \varphi: a \mapsto ab, b \mapsto a, a)$$

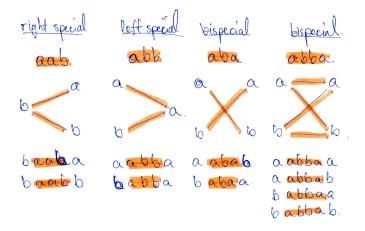






Thue-Morse

$$S = (\{a, b\}, \varphi: a \mapsto ab, b \mapsto ba, a)$$

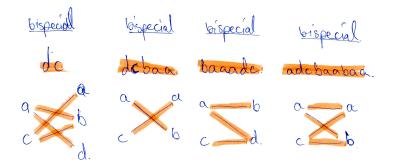


Example



QUATERNARY SYSTEM

 $\mathsf{S} = (\{a, b, c, d\}, \quad \varphi \colon a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, \quad c)$





RECURSIVE EXPRESSIONS FOR FACTOR COMPLEXITY

$$p_S(n) = \#\{u \in \mathbf{L}(S) : |u| = n\}$$
$$m_S(u) = \#(\mathbf{L}(S) \cap AuA) - \#(\mathbf{L}(S) \cap uA) - \#(\mathbf{L}(S) \cap uA) + 1$$
(bilateral order)



RECURSIVE EXPRESSIONS FOR FACTOR COMPLEXITY

$$p_{S}(n) = \#\{u \in L(S) : |u| = n\}$$
$$m_{S}(u) = \#(L(S) \cap AuA) - \#(L(S) \cap uA) - \#(L(S) \cap uA) + 1$$
(bilateral order)

Theorem (Cassaigne, 1997) $p_S(n+2) - 2p_S(n+1) + p_S(n) = \sum_{|u|=n} m_S(u).$



RECURSIVE EXPRESSIONS FOR FACTOR COMPLEXITY

$$p_{S}(n) = \#\{u \in L(S) : |u| = n\}$$
$$m_{S}(u) = \#(L(S) \cap AuA) - \#(L(S) \cap uA) - \#(L(S) \cap uA) + 1$$
(bilateral order)

Theorem (Cassaigne, 1997) $p_S(n+2) - 2p_S(n+1) + p_S(n) = \sum_{|u|=n} m_S(u).$

• $m_S(u) = 0$ when u is not bispecial.



$$S = (\{a, b\}, \varphi: a \mapsto ab, b \mapsto ba, a)$$

Theorem (Luca & Mione, 1994)

Every bispecial factor $u \in \mathbf{L}(S)$ of length ≥ 2 is of the form

$$\varphi^n(ab), \quad \varphi^n(ba), \quad \varphi^n(aba), \quad \varphi^n(bab),$$



$$S = (\{a, b\}, \varphi: a \mapsto ab, b \mapsto ba, a)$$

Theorem (Luca & Mione, 1994)

Every bispecial factor $u \in \mathbf{L}(S)$ of length ≥ 2 is of the form

$$\varphi^n(ab), \quad \varphi^n(ba), \quad \varphi^n(aba), \quad \varphi^n(bab),$$

Corollary (Brlek, 1989; Cassaigne, 1997)

In the Thue–Morse system the values of $p_S(n)$ for $n \ge 3$ are

$$p_{5}(n) = \begin{cases} 4n - 2 \cdot 2^{m} - 4 & \text{if } 2 \cdot 2^{n} < n \le 3 \cdot 2^{m} \\ 2n + 4 \cdot 2^{m} - 2 & \text{if } 3 \cdot 2^{n} < n \le 4 \cdot 2^{m} \end{cases}$$

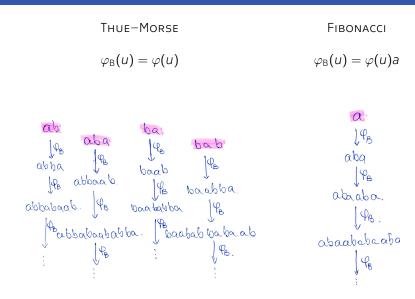


$$S = (\{a, b\}, \varphi: a \mapsto ab, b \mapsto ba, a)$$

- Every bispecial factor is of the form $\varphi_{B}^{n}(a)$ where $\varphi_{B}(u) = \varphi(u)a$.
- The complexity is $p_S(n) = n + 1$ (Sturmian system).

Bispecial factors



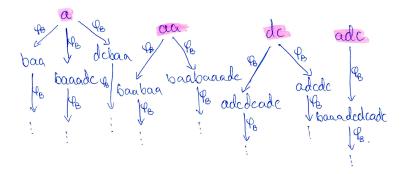




QUATERNARY SYSTEM

$$S = (\{a, b, c, d\}, \varphi: a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, c)$$

 $\varphi_{B}(u) = x_{u}\varphi(u)y_{u}$ where x_{u} and y_{u} depend on the extensions of u.





Klouda, 2012

Under certain **key conditions**, there exists a finite set of initial bispecial words and a finitely determined rule from which all other bispecial factors are obtained.



Klouda, 2012

Under certain **key conditions**, there exists a finite set of initial bispecial words and a finitely determined rule from which all other bispecial factors are obtained.

 These conditions also imply that the set of initial bispecial words and the rule can both be calculated algorithmically.



PROPAGATING For all $a \in A$, $\varphi(a) \neq \varepsilon$ (non-erasing).



PROPAGATING For all $a \in A$, $\varphi(a) \neq \varepsilon$ (non-erasing).

Non-pushy There are only finitely many $u \in \mathbf{L}(S)$ such that $\lim_{n \to \infty} |\varphi^n(u)| < \infty$.



PROPAGATING For all $a \in A$, $\varphi(a) \neq \varepsilon$ (non-erasing).

Non-pushy There are only finitely many $u \in \mathbf{L}(S)$ such that $\lim_{n\to\infty} |\varphi^n(u)| < \infty$.

EXTENDABLE For all $u \in L(S)$, there are letters $a, b \in A$ such that $aub \in L(S)$.



PROPAGATING For all $a \in A$, $\varphi(a) \neq \varepsilon$ (non-erasing).

Non-pushy There are only finitely many $u \in \mathbf{L}(S)$ such that $\lim_{n\to\infty} |\varphi^n(u)| < \infty$.

EXTENDABLE For all $u \in L(S)$, there are letters $a, b \in A$ such that $aub \in L(S)$.

INJECTIVE φ is injective on **L**(*S*).



PROPAGATING For all $a \in A$, $\varphi(a) \neq \varepsilon$ (non-erasing).

Non-pushy There are only finitely many $u \in \mathbf{L}(S)$ such that $\lim_{n\to\infty} |\varphi^n(u)| < \infty$.

EXTENDABLE For all $u \in L(S)$, there are letters $a, b \in A$ such that $aub \in L(S)$.

INJECTIVE φ is injective on **L**(*S*). (can be weakened!)



PROPAGATING For all $a \in A$, $\varphi(a) \neq \varepsilon$ (non-erasing).

Non-pushy There are only finitely many
$$u \in \mathbf{L}(S)$$
 such that $\lim_{n\to\infty} |\varphi^n(u)| < \infty$.

EXTENDABLE For all $u \in L(S)$, there are letters $a, b \in A$ such that $aub \in L(S)$.

INJECTIVE φ is injective on **L**(*S*). (can be weakened!)

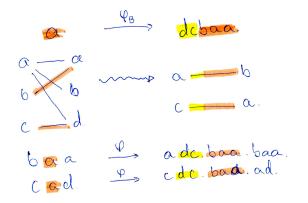
CIRCULAR More on this later...

Part 3

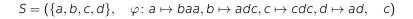


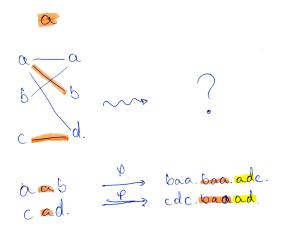


$$\mathsf{S} = \big(\{a, b, c, d\}, \quad \varphi \colon a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, \quad c\big)$$

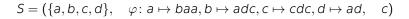


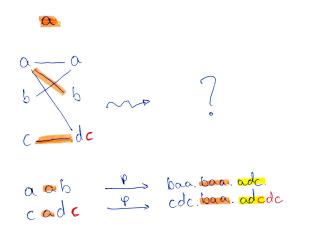






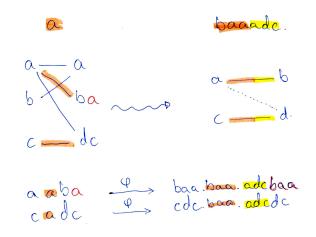








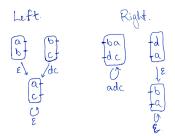
 $\mathsf{S} = \big(\{a, b, c, d\}, \quad \varphi \colon a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, \quad c\big)$





$$S = (\{a, b, c, d\}, \quad \varphi \colon a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, \quad c)$$

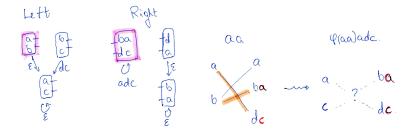
 $\varphi_{B}(u)$ depends on admissible left and right extensions of u in S.





$$S = (\{a, b, c, d\}, \varphi: a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, c)$$

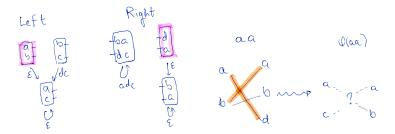
 $\varphi_{B}(u)$ depends on admissible left and right extensions of u in S.





$S = (\{a, b, c, d\}, \varphi: a \mapsto baa, b \mapsto adc, c \mapsto cdc, d \mapsto ad, c)$

 $\varphi_{B}(u)$ depends on admissible left and right extensions of u in S.





Theorem (Klouda, 2012) Under the **key conditions**²:

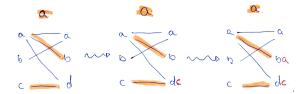
- 1. *S* has forky graphs which can be effectively computed.
- 2. There is a computable constant N such that any bispecial factor longer than N is the image by $\varphi_{\rm B}$ of a shorter one.

²Propagating, non-pushy, extendable, injective, circular



Theorem (Klouda, 2012) Under the **key conditions**²:

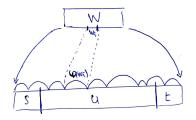
- 1. *S* has forky graphs which can be effectively computed.
- 2. There is a computable constant N such that any bispecial factor longer than N is the image by $\varphi_{\rm B}$ of a shorter one.
 - Main difficulty: make sure the extension process terminates.



²Propagating, non-pushy, extendable, injective, circular

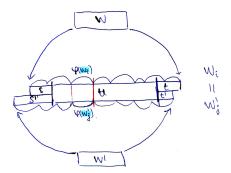






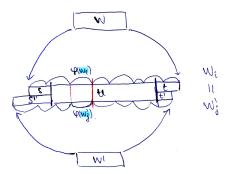


SYNCHRONIZED INTERPRETATIONS





SYNCHRONIZED INTERPRETATIONS



CIRCULAR Interpretations synchronize after a bounded delay³.

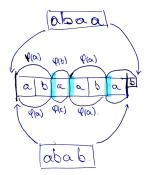
³using definition of Klouda & Starosta, 2019

Examples



Fibonacci

 $S = (\{a, b\}, \varphi: a \mapsto ab, b \mapsto a, a)$



Synchronization delay is 0.

Examples



THUE-MORSE $S = (\{a, b\}, \varphi: a \mapsto ab, b \mapsto ba, a)$

Synchronization delay is 1.

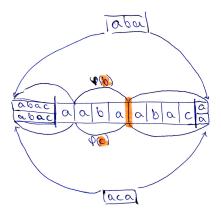
66





Ternary system

$$S = (\{a, b, c\}, \varphi: a \mapsto abaca, b \mapsto aba, c \mapsto aba, a)$$

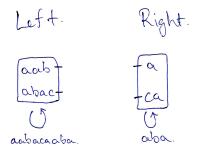


Synchronization delay is 9.



$$S = (\{a, b, c\}, \varphi: a \mapsto abaca, b \mapsto aba, c \mapsto aba, a)$$

This is a non-injective systems which has computable forky graphs if we ignore the bispecial factor *a* (of length 1).





$$S = (\{a, b, c\}, \varphi: a \mapsto acabaca, b \mapsto aba, c \mapsto aba, a)$$

This is a non-injective systems which **does not have computable forky graphs**. It satisfies all key conditions except for injectivity.





1.1 Propagating: trivial.



- 1.1 **Propagating**: trivial.
- 1.2 Non-pushy: Ehrenfeucht & Rozenberg, 1983.



- 1.1 Propagating: trivial.
- 1.2 Non-pushy: Ehrenfeucht & Rozenberg, 1983.
- 1.3 Extendable: simple.



- 1.1 Propagating: trivial.
- 1.2 Non-pushy: Ehrenfeucht & Rozenberg, 1983.
- 1.3 Extendable: simple.
- 1.4 Injective: unclear, use Sardinas–Paterson as proxy.



- 1.1 **Propagating**: trivial.
- 1.2 Non-pushy: Ehrenfeucht & Rozenberg, 1983.
- 1.3 Extendable: simple.
- 1.4 Injective: unclear, use Sardinas–Paterson as proxy.
- 1.5 Circular: Klouda & Starosta, 2019.



- 1.1 **Propagating**: trivial.
- 1.2 Non-pushy: Ehrenfeucht & Rozenberg, 1983.
- 1.3 **Extendable**: simple.
- 1.4 Injective: unclear, use Sardinas–Paterson as proxy.
- 1.5 **Circular**: Klouda & Starosta, 2019.
- 2. Calculate the minimal delay for circularity.



- 1.1 Propagating: trivial.
- 1.2 Non-pushy: Ehrenfeucht & Rozenberg, 1983.
- 1.3 **Extendable**: simple.
- 1.4 Injective: unclear, use Sardinas–Paterson as proxy.
- 1.5 Circular: Klouda & Starosta, 2019.
- 2. Calculate the minimal delay for circularity.
- 3. Calculate reasonable forky graphs.



- 1.1 **Propagating**: trivial.
- 1.2 Non-pushy: Ehrenfeucht & Rozenberg, 1983.
- 1.3 **Extendable**: simple.
- 1.4 Injective: unclear, use Sardinas–Paterson as proxy.
- 1.5 Circular: Klouda & Starosta, 2019.
- 2. Calculate the minimal delay for circularity.
- 3. Calculate reasonable forky graphs.
- 4. Calculate an initial set of bispecial factors.



1. Package in a new SageMath class **DFOL**.



- 1. Package in a new SageMath class **DFOL**.
- 2. Web server for easy access.



- 1. Package in a new SageMath class DFOL.
- 2. Web server for easy access.
- 3. Decide (eventual) injectivity on the language.



- 1. Package in a new SageMath class DFOL.
- 2. Web server for easy access.
- 3. Decide (eventual) injectivity on the language.
- 4. Beyond injective, extendable, non-erasing.



- 1. Package in a new SageMath class DFOL.
- 2. Web server for easy access.
- 3. Decide (eventual) injectivity on the language.
- 4. Beyond injective, extendable, non-erasing.
- 5. Finer structure of bispecial factors.

