# On Union-Free and Deterministic Union-Free Languages

Galina Jirásková *and* Benedek Nagy

Kosice, Slovakia

Debrecen, Hungary

IFIP-TCS 2012, Amsterdam

# Outline – what is already known

- Regular languages
- Union-free languages
- Union-normal form
- 1cfp automata
- Deterministic union-free languages

- Parikh sets

# Outline – what is already known

- Regular languages (Kleene 1950's)
- Union-free languages
- Union-normal form (Nagy 2004)
- 1cfp automata (Nagy 2006)
- Deterministic union-free languages (Jirásková, Masopust, 2010)
- Parikh sets (1960's)

# Outline – what is new

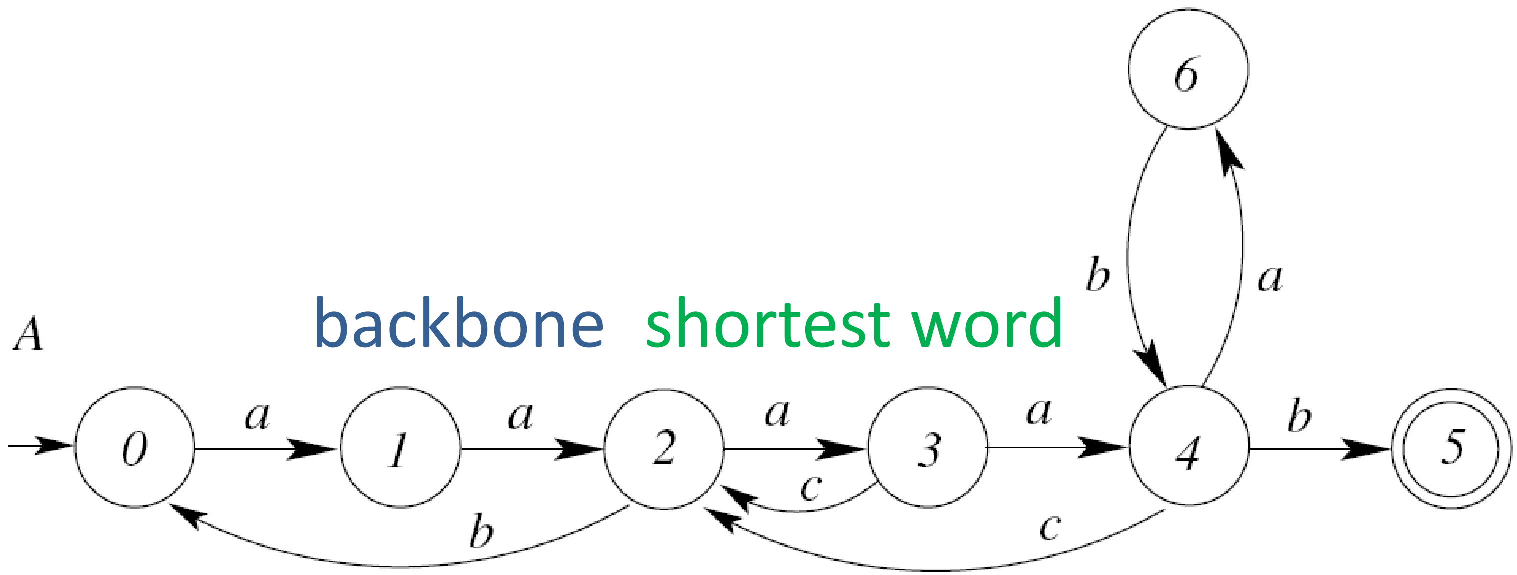
- Representation of Deterministic Union-Free Languages: Balloon automata
- Closure prop. of det. union-free languages
- Normal form with det. union-free lang. ?

# Outline – what is new

- Representation of Deterministic Union-Free Languages: Balloon automata
- Closure prop. of det. union-free languages
- Normal form with det. union-free lang. ? NO !
- New family: finite union of det. Union-free…
- Some properties of this class
- Parikh-sets of union-free languages

# Definition of union-free lang.

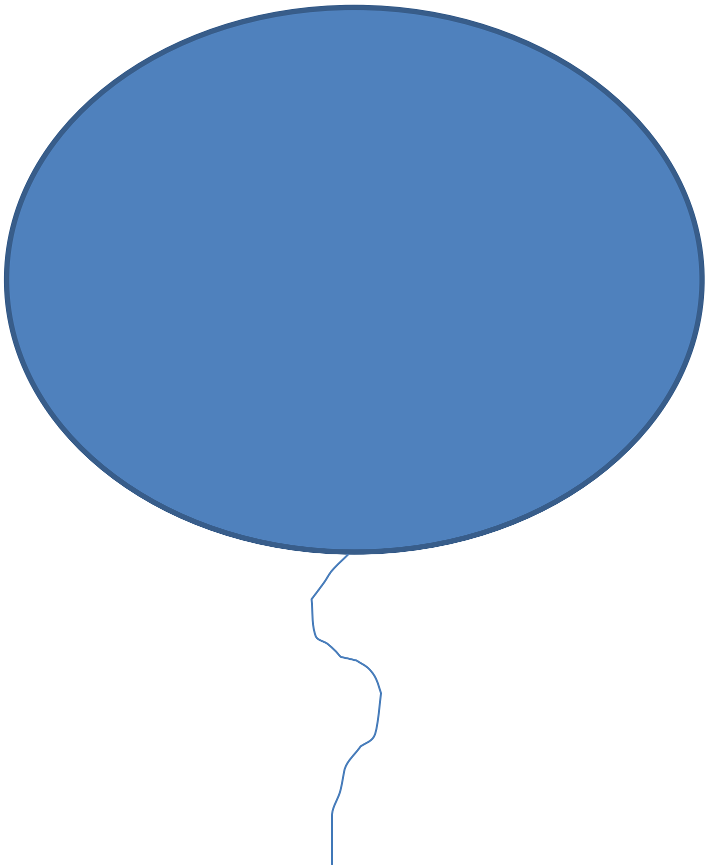
- Regular expression without union
  (only concatenation and Kleene-star can be used)
- This class is accepted by 1CFPA (Nagy, 2006):
  1cycle-free-path automata

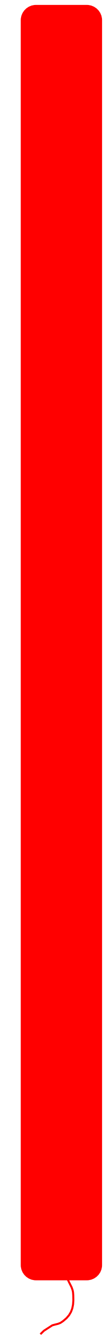


# Determ. union-free languages

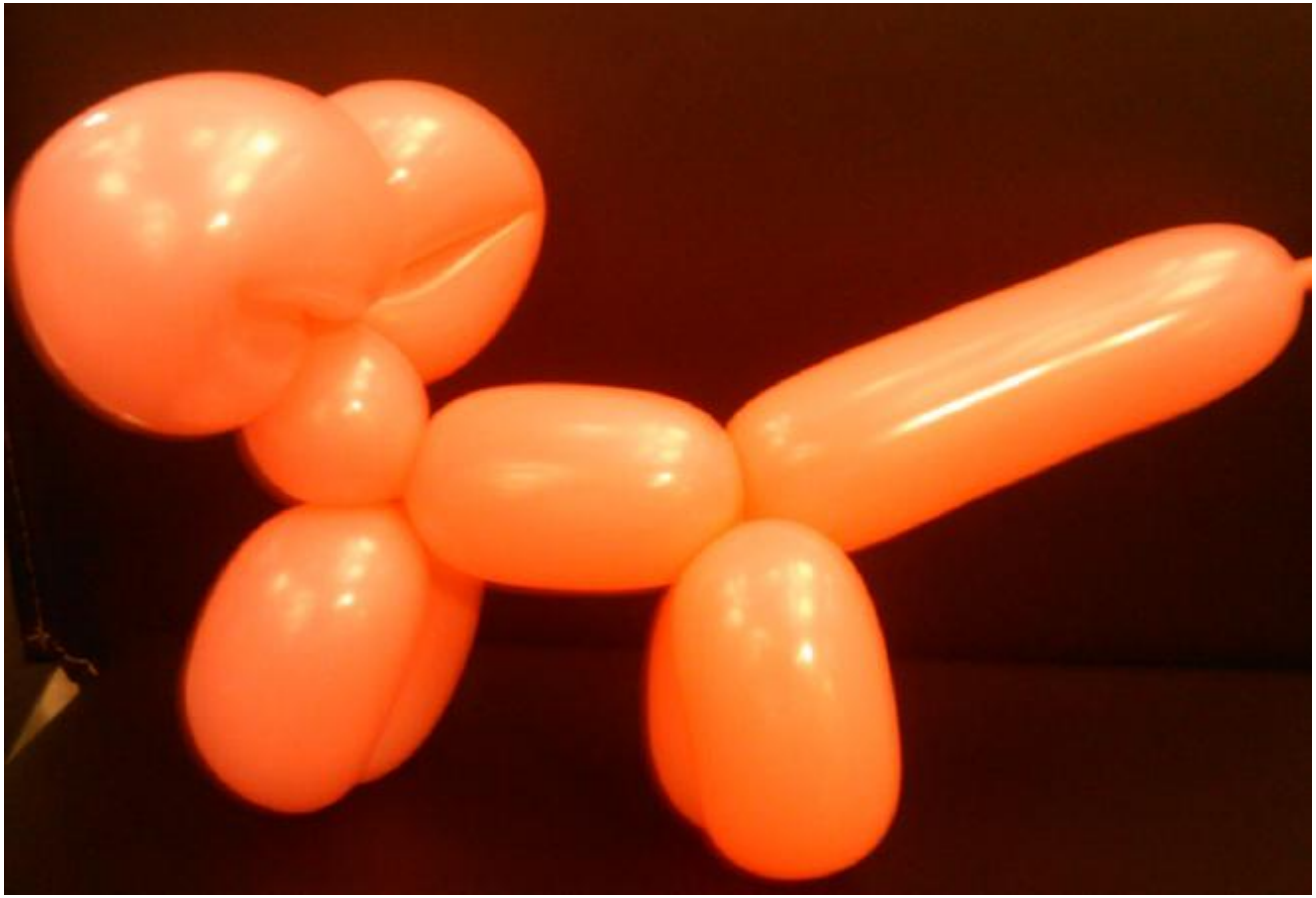
- Defined by determinisitc 1cfp automata
- Strictly smaller group of languages than the class of union-free regular…

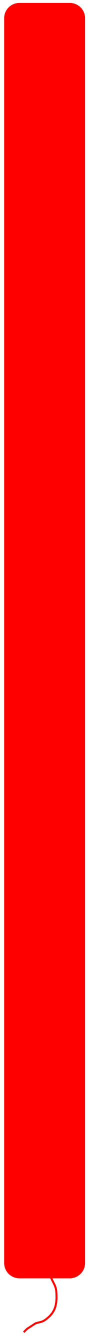
# Balloon



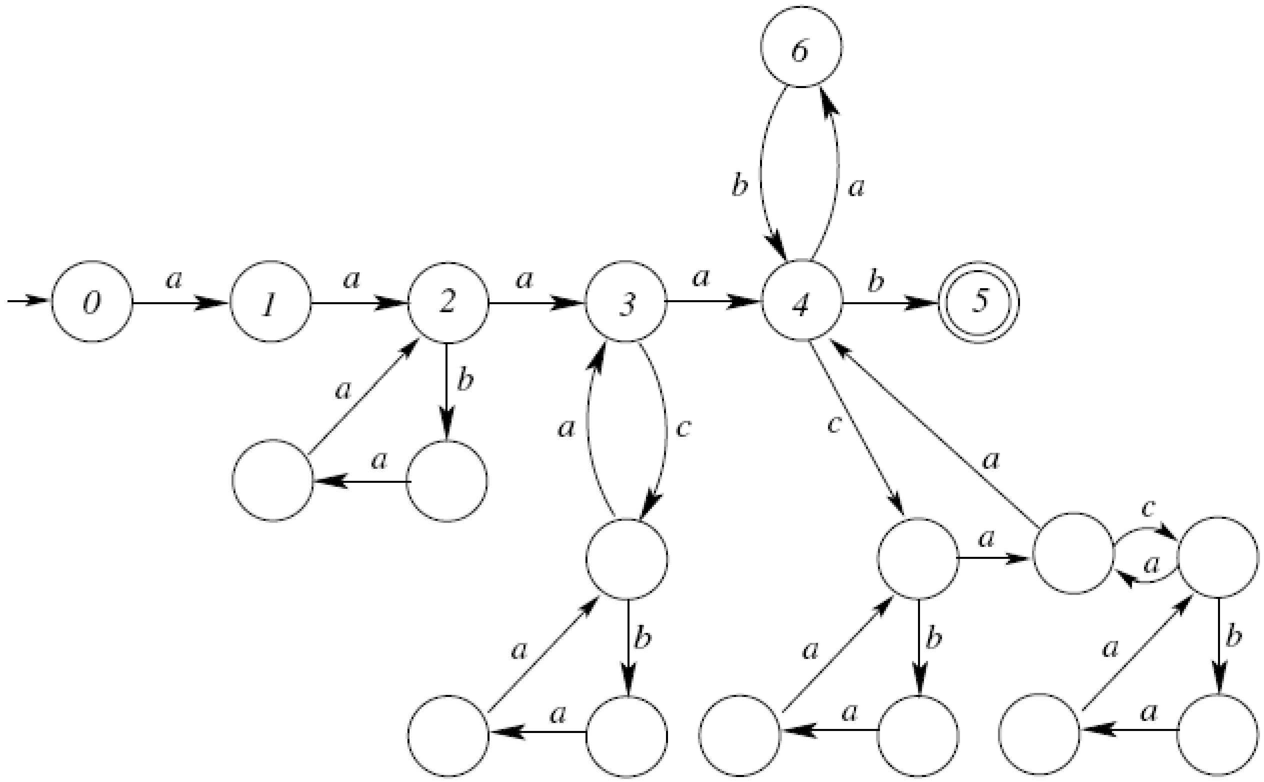


# Balloon





# Balloon automata (DFA)




- The in and out node is the same at a branching
- Every det. u-free language is accepted by a balloon DFA and vice-versa. (constructive proof)

# Cycle depth of balloon dfa

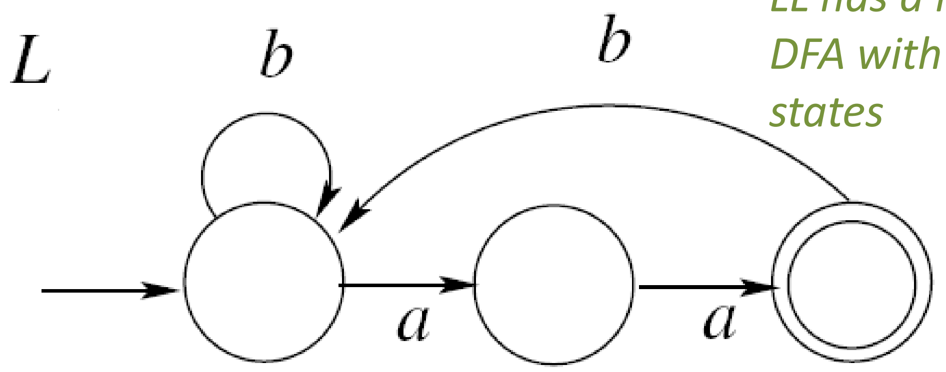
- *The cycle depth of a balloon automaton is the maximal number of its nested cycles.*
- The cycle depth of a balloon automaton is 0 for cycle-free automata.

ANTI

# Closure properties of det. u-free l.

- *The class of deterministic union-free languages is not closed under*
- *complement,* $\{\varepsilon\}^c = \Sigma^+,$
- *union,* $\{a\} \cup \{b\} = \{a, b\},$
- *intersection,* $b^*ab^* \cap a^*ba^* \subseteq \{ab, ba\} \cup \{a, b\}^{\geq 3}$
- *symmetric difference,* $\{a\} \oplus \{b\} = \{a, b\},$
- *concatenation,*
- *square,*
- *…*

*The shortest word is not unique*

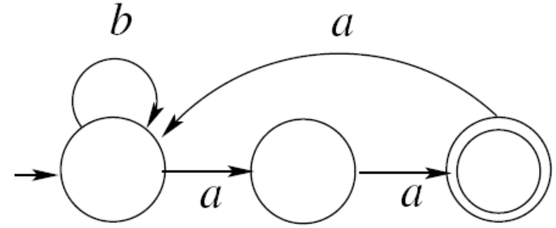



*LL has a minimal DFA with 2 final states*

ANTI

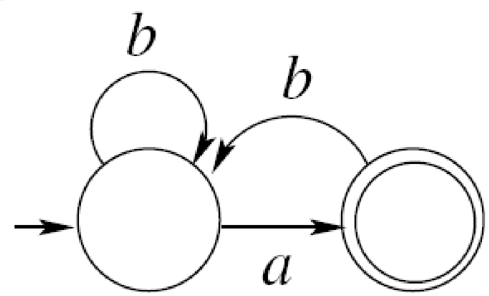
# Closure properties of det. u-free l.

- *The class of deterministic union-free languages is not closed under*




- *star,*
- *cyclic shift,*
- *shuffle,*
- *reversal,*
- *homomorphism,*
- *inverse morphism.*

$$\text{Shift}(\{ab\}) = \{ab, ba\},$$
$$\{a\} \sqcup\!\sqcup \{b\} = \{ab, ba\},$$




h(a) = a, h(b) = ab,
h(c) = c
h(ab*ac*) =
a(ab)*ac*

$$h^{-1}(\{aa\}) = \{aa, ab, ba, bb\} \quad h(a) = h(b) = a$$

# Union-normal form

- Every regular language is a finite union of union-free languages (Nagy, 2004)
- What about det. u-free?

# Inner cycle of balloon automata

- Inner cycle: there is no further branching: one must follow the only 1 way of the cycle back to its origin…

- LEMMA: *If the cycle depth of a balloon automaton is at least one, and the length of the/an inner cycle is at least two, then there is a state where only one of the symbols defines a transition.*

# MAIN Result

- $((a+b)(a+b))^*$ cannot be written as a finite union of det. u-free languages.

- Proof: the technique is analogous to diagonalisation …
  - Only even lenghts words
  - Every word from $(a+b)^*$ is a prefix of some words
  - Indirect technique

# MAIN Result

- $((a+b)(a+b))^*$
  - Indirect technique:
  - Assume that we have a finite set of det. u-free languages
  - Only even length words: inner cycles have length more than 1 , in each det. 1cfpa (if cycle depth is non zero… i.e., accepting more than 1 word)
  - Let us order the det. 1cfpa’s.
    a word can be constructed that fails on each of them, it is a contradicicition (every word is a prefix..)

# Semi-linear languages

- Ordered alphabet. {a,b,c}
- word $\rightarrow$ vector (Parikh) abacccac $\rightarrow$ (3,1,4)
- language $\rightarrow$ set of vectors (Parikh set)
- linear sets:

$$\{\alpha_0 + n_1\alpha_1 + \cdots + n_m\alpha_m \mid n_j \geq 0 \text{ for } j = 1, 2, \ldots, m\}$$

- semi-linear set: finite union of linear sets
- semi-linear language:

  its Parikh set is semi-linear

# Semi-linear languages

- Parikh theorem…
- Every context-free language is semi-linear.
- For every semi-linear set *S* there is a regular language that's image is *S*
- There is no difference between the commutative closure of regular and context-free languages.

# Parikh sets of union-free lang.

- Do union-free languages have the same expressing power as regular and context-free languages regarding their Parikh sets?

# Parikh sets of u-free l. - RESULTS

- Do union-free languages have the same expressing power as regular and context-free languages regarding their Parikh sets?  NO!
- There is a semi-linear set W that is not a Parikh set of any union-free language.
  Example:  (0,4) + (2,2)
  indicates two shortest paths !

# Parikh sets of u-free l. - RESULTS

- Do union-free languages have the same expressing power as regular and context-free languages regarding their Parikh sets? NO! BUT…
- Every linear set is a Parikh set of a union-free language. Construction of spec. 1cfpa (with cycle depth at most 1.)
- There is a union-free language that has non-linear Parikh set. *a(bb(aab)*ba)*a*

# Parikh sets of u-free - RESULT

- conditional linear sets

$$\alpha = \alpha_0 + \delta_1 n_1 \alpha_1 + \delta_2 n_2 \alpha_i + \cdots + \delta_m n_m \alpha_m$$

- conditional coefficients:
- $\delta_1 = 1$, without condition.
- the condition uses a smaller index

$$\delta_i = 1, \qquad \textit{if there is no condition for } \alpha_i,$$

$$\delta_i = \begin{cases} 1, & \textit{if } \delta_j n_j > 0, \\ 0, & \textit{if } \delta_j n_j = 0, \end{cases} \qquad \textit{if } \alpha_i \textit{ depends on the coefficient of } \alpha_j.$$

- between linear and semi-linear

The conditional linear sets are exactly those sets that are Parikh-sets of union-free languages.

# Summary - conclusion

- Subregular classes …
- Proper infinite hierarchy and
- a new language class

$$dU_* = \bigcup_{i=1}^{\infty} dU_i .$$

this class is closed under union, but not closed under intersection, complement, star

- Further: Balloon automata are maximal representations ?

# Related literature


- Benedek Nagy: A normal form for regular expressions. In: Supplemental Papers for DLT’04 (eds: C. S. Calude, E. Calude, M. J. Dinnen), (CDMTCS-252), Auckland, 2004.
- Benedek Nagy: Union-free regular languages and 1-cycle-free-path-automata. Publ. Math. Debrecen 68 (2006) 183–197.
- Sergey Afonin, Denis Golomazov: Minimal Union-Free Decom-positions of Regular Languages. LATA, LNCS 5457 (2009), 83–92.
- Galina Jirásková and Tomás Masopust: Complexity in union-free regular languages. International Journal of Foundations of Computer Science 22 (2011) 1639–1653. earlier version: DLT 2010, LNCS 6224 (2010), 255–266.

THANK You for your attention !

nbenedek@inf.unideb.hu