Two Results on Discontinuous Input Processing

Vojtěch Vorel

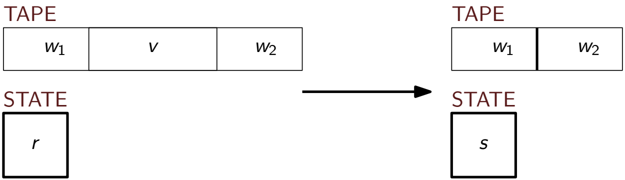
Charles University
Prague, Czech Republic

DCFS 2016

General Jumping Finite Automaton

- *GJFA* is a triple $M = (Q, \Sigma, R, q_0, F)$:
  - Q ... finite set of *states*
  - Σ ... finite *alphabet*
  - R ... finite set of *rules* from $Q \times \Sigma^* \times Q$
  - q_0 ... the *initial state*
  - F ... the set of *final states*

- Step of computation:



whenever $(r, v, s) \in R$.

General Jumping Finite Automaton

- *GJFA* is a triple $M = (Q, \Sigma, R, q_0, F)$:
  - Q ... finite set of *states*
  - Σ ... finite *alphabet*
  - R ... finite set of *rules* from $Q \times \Sigma^* \times Q$
  - q_0 ... the *initial state*
  - F ... the set of *final states*

- Initial configuration:

  TAPE

  w

  STATE

  q_0

Accepting configuration:

TAPE

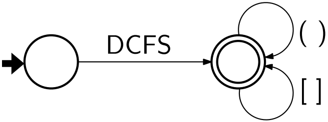
empty

STATE

$f \in F$

Example GJFA

$\Sigma = \{D, C, F, S, (,), [,]\}$

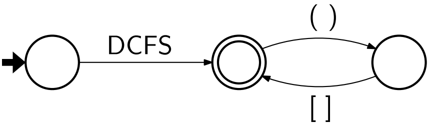


Accepts e.g.:

- DCFS ( ) [ ]
- (DCFS) [( )]
- [DCFS ( )] ( )
- ...

Example GJFA

$\Sigma = \{D, C, F, S, (,), [,]\}$



Accepts e.g.:

- DCFS ( ) [ ]
- (DCFS) [( )]
- [DCFS ( )] ( )

But not:

- (DCFS [ ])

Models Related To GJFA

		states	context	axioms	end marks	lang. class notation
M.Z. (2012)	general jumping finite automata	YES	no	no	no	**GJFA**
A.K.R.V. (2010)	graph-controlled insertion systems	YES	YES	YES	no	$\mathrm{LStP}_*(\mathrm{ins}_*^{k,k})$
Păun (1990)	regular control semicontextual grammars without app. checking					$= \mathcal{C}_k$
Păun (1998)	insertion systems	no	YES	YES	no	INS_*^k
Păun (1985)	semi-contextual grammars					$= \mathcal{J}_k$
Č.M. (2010)	clearing restarting automata	no	YES	no	YES	$\mathcal{L}(k\text{-cl-RA})$

M.Z.: Meduna, Zemek
A.K.R.V.: Alhazov, Krassovitski, Rogozhin, Verlan
Č.M.: Černo, Mráz

k...max. context size

Models Related To GJFA

		states	context	axioms	end marks	lang. class notation
M.Z. (2012)	general jumping finite automata	YES	no	no	no	**GJFA**
A.K.R.V. (2010)	graph-controlled insertion systems	YES	YES	YES	no	$\mathrm{LStP}_*(\mathrm{ins}_*^{k,k})$
Păun (1990)	regular control semicontextual grammars without app. checking					$= \mathcal{C}_k$

M.Z.: Meduna, Zemek
A.K.R.V.: Alhazov, Krassovitski, Rogozhin, Verlan
Č.M.: Černo, Mráz

k...max. context size

Models Related To GJFA

		states	context	axioms	end marks	lang. class notation
M.Z. (2012)	general jumping finite automata	YES	no	(YES)	no	**GJFA**
A.K.R.V. (2010)	graph-controlled insertion systems	YES	YES	YES	no	$\mathrm{LStP}_*(\mathrm{ins}_*^{k,k})$
Păun (1990)	regular control semicontextual grammars without app. checking					$= \mathcal{C}_k$

M.Z.: Meduna, Zemek
A.K.R.V.: Alhazov, Krassovitski, Rogozhin, Verlan
Č.M.: Černo, Mráz

k...max. context size

Case of k=0

		states	context	axioms	end marks	lang. class notation
M.Z. (2012)	general jumping finite automata	YES		(YES)	no	**GJFA**
A.K.R.V. (2010)	graph-controlled insertion systems	YES		YES	no	$\mathrm{LStP}_*(\mathrm{ins}_*^{k,k})$
Păun (1990)	regular control semicontextual grammars without app. checking					$= \mathcal{C}_k$

M.Z.: Meduna, Zemek
A.K.R.V.: Alhazov, Krassovitski, Rogozhin, Verlan
Č.M.: Černo, Mráz

k...max. context size

~~General~~ Jumping Finite Automaton

JFA $\subseteq$ **GJFA**

- R … finite set of *rules* ~~from $Q \times \Sigma^* \times Q$~~ from $Q \times \Sigma \times Q$

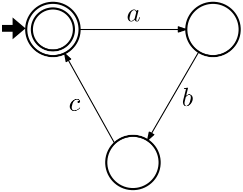
The following are equivalent:

- $L \in$ **JFA**
- $L = \mathrm{perm}(K)$ for $K \in$ **REG**
- L can be expressed from Σ using:
  - union (binary)
  - shuffle (binary)
  - iterated shuffle (unary)

"alphabetic shuffle expressions"

Chomsky Hierarchy

- **GJFA $\subseteq$ CSL**
  - Proof: $L(M)$ is easily decidable in linear space
- **JFA $\not\subseteq$ CFL**
  - Proof:



Chomsky Hierarchy

- **REG** $\nsubseteq$ **GJFA**
  - Example No. 1: $L = a^*b^*$
  - Example No. 2: $L = (ab)^*$
  - Lemma: If $L \in$ **GJFA**, the first deleted (or last inserted) factor of $w \in L$ can be moved **anywhere**!
- **FIN** $\subseteq$ **GJFA**
- **FIN** $\nsubseteq$ **JFA**
  - Proof: Each $L \in$ **JFA** is permutation-closed

Chomsky Hierarchy

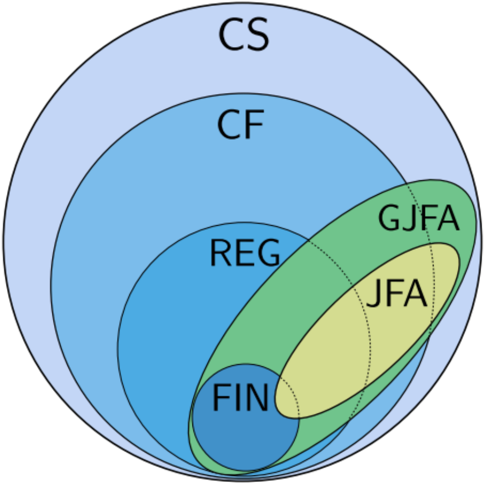


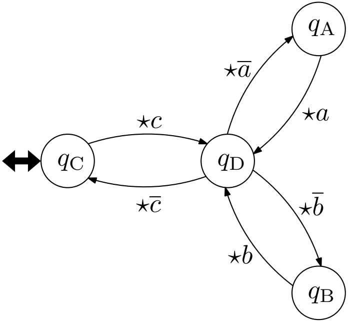
Figure by Meduna, Vrábel, Zemek

Closure Properties

	GJFA	JFA
Endmarking	−	−
Concatenation	−	−
Shuffle	−	+
Union	+	+
Complement	−	+
Intersection	−	+
Int. with regular languages	−	−
Kleene star	−	−
Reversal	+	+
Homomorphism	−	−
Inverse homomorphism	−	+

Complexity of GJFA Languages

- **GJFA** contains an NP-complete language
  - Example:



$\Sigma = \{a, \overline{a}, b, \overline{b}, c, \overline{c}, \star\}$

+ reduction from the EXACT COVER PROBLEM.

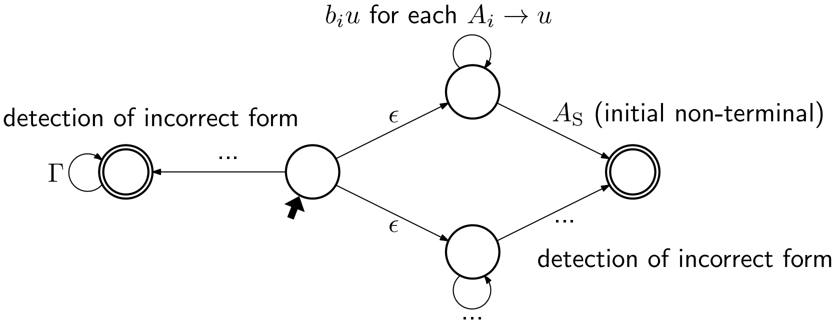
- **JFA** $\subseteq$ **P**
  - Proof: Actually, **JFA** $\subseteq$ **NL**

Complexity of GJFA Properties

- **Universal word problem**
  - NP-complete for **JFA**
  - NP-complete for **GJFA**
- **Disjointness**
  - NP-complete for **JFA**
  - Undecidable for **GJFA**
- **Universality**
  - Undecidable for **GJFA** – New result

Proof Idea

- Take any CFG G on Σ in Greibach normal form
- Produce a 5-state *GJFA* on $\Gamma \supset \Sigma$ that:
  - accepts any $w' \in \Gamma$ in *incorrect form*
  - accepts $w' \in \Gamma$ in *correct form* iff it *encodes* $w \in L(G)$



Consequences of "*Universality is Undecidable for GJFA*"

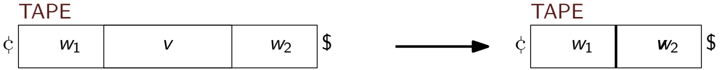
- Undecidability of more general properties:
  - Inclusion
  - Equivalence

- Holds in equivalent models:
  - **Graph-controlled insertion systems**
    with zero contexts
  - **Regular control semicontextual grammars without appearance checking**
    with zero contexts

Clearing Restarting Automaton

- *k-clearing restarting automaton*
  is a pair $M = (\Sigma, I)$:
  - Σ ... finite *alphabet*
  - I ... finite set of *rules* of the form (u_L, v, u_R)*:*
    – $u_L \in \Sigma^k \cup \text{¢}\Sigma^{k-1}$ ... *left context*
    – $v \in \Sigma^+$
    – $u_R \in \Sigma^k \cup \Sigma^{k-1}\$$ ... *right context*

- Step of computation:



  whenever $(u_L, v, u_R) \in I$ such that
  - u_L is a suffix of $\text{¢}w_1$,
  - u_R is a prefix of $w_2\$$.

Models Related To Cl-RA

		states	context	axioms	end marks	lang. class notation
M.Z. (2012)	general jumping finite automata	YES	no	no	no	**GJFA**
A.K.R.V. (2010)	graph-controlled insertion systems	YES	YES	YES	no	$\mathrm{LStP}_*(\mathrm{ins}_*^{k,k})$
Păun (1990)	regular control semicontextual grammars without app. checking					$= \mathcal{C}_k$
Păun (1998)	insertion systems	no	YES	YES	no	INS_*^k
Păun (1985)	semi-contextual grammars					$= \mathcal{J}_k$
Č.M. (2010)	clearing restarting automata	no	YES	no	YES	$\mathcal{L}(k\text{-cl-RA})$

M.Z.: Meduna, Zemek
A.K.R.V.: Alhazov, Krassovitski, Rogozhin, Verlan
Č.M.: Černo, Mráz

k...max. context size

Models Related To Cl-RA

		states	context	axioms	end marks	lang. class notation
Păun (1998)	insertion systems	no	YES	YES	no	INS_*^k
Păun (1985)	semi-contextual grammars					$= \mathcal{J}_k$
Č.M. (2010)	clearing restarting automata	no	YES	no	YES	$\mathcal{L}(k\text{-cl-RA})$

M.Z.: Meduna, Zemek
A.K.R.V.: Alhazov, Krassovitski, Rogozhin, Verlan
Č.M.: Černo, Mráz

k...max. context size

Models Related to Cl-RA

		states	context	axioms	end marks	lang. class notation
Păun (1998)	insertion systems	no	YES	YES	no	INS_*^k
Păun (1985)	semi-contextual grammars					$= \mathcal{J}_k$
Č.M. (2010)	clearing restarting automata	no	YES	(YES)	YES	$\mathcal{L}(k\text{-cl-RA})$

M.Z.: Meduna, Zemek
A.K.R.V.: Alhazov, Krassovitski, Rogozhin, Verlan
Č.M.: Černo, Mráz

k…max. context size

Chomsky Hierarchy

- **REG** $\subseteq \mathcal{L}(\text{cl-RA})$ up to ε
  - Proof: Use pumping lemma to clear close to the beginning
- **CFL** $\not\subseteq \mathcal{L}(\text{cl-RA})$ even with ignoring ε
  - Example: $L = \{a^n bc^n \mid n \geq 0\}$
- $\mathcal{L}(1\text{-cl-RA}) \subseteq$ **CFL**
- $\mathcal{L}(2\text{-cl-RA}) \not\subseteq$ **CFL**
  - Černo, Mráz: example with $|\Sigma| = 6$
  - New result: example with $|\Sigma| = 2$

Proof Idea

$\Sigma = \{0, 1\}$

- A *defect* in $w \in \Sigma^*$ is a letter with equal neighbours: $0\mathbf{0}0, 1\mathbf{0}1, 0\mathbf{1}0, 1\mathbf{1}1$.
- So, words without defects are of the form $\ldots 001100110011 \ldots$

Define a Cl-RA on Σ generatively:

- Start with a short without defects (e.g., 00)
- Introduce defect at the beginning: $(¢, 10, \mathbf{0}0)$
- Move defect to the right but add two letters:
  - $(0\mathbf{1}, \underline{10}, 00)$ $\quad(01, 10, \mathbf{0}0)$
  - $(0\mathbf{0}, \underline{11}, 01)$ $\quad(00, 11, \mathbf{0}1)$
  - $(1\mathbf{1}, \underline{00}, 10)$ $\quad(11, 00, \mathbf{1}0)$
  - $(1\mathbf{0}, \underline{01}, 11)$ $\quad(10, 01, \mathbf{1}1)$
- Make the defect disappear at the end: $(0\mathbf{1}, \underline{10}, 0\$)$

Proof Idea

Together, transforming a word without defect to another **triples** the length.
But:

- CFL are closed under intersection with $L = \{w \mid w \text{ has no defect}\}$
- Lengths of words in a CFL cannot grow exponentially

Open Questions

Closure properties of Cl-RA:

	Cl-RA
Endmarking	+
Concatenation	−
Shuffle	☮
Union	−
Complement	☮
Intersection	−
Int. with regular languages	−
Kleene star	☮
Reversal	+
Homomorphism	−
Inverse homomorphism	☮

GJFA:

- Expressiveness w.r.t. restricted lengths of labels
- NP-hardness w.r.t. restricted alphabets and lengths of labels
- Complexity of universality for **JFA**

Thank you for your attention!