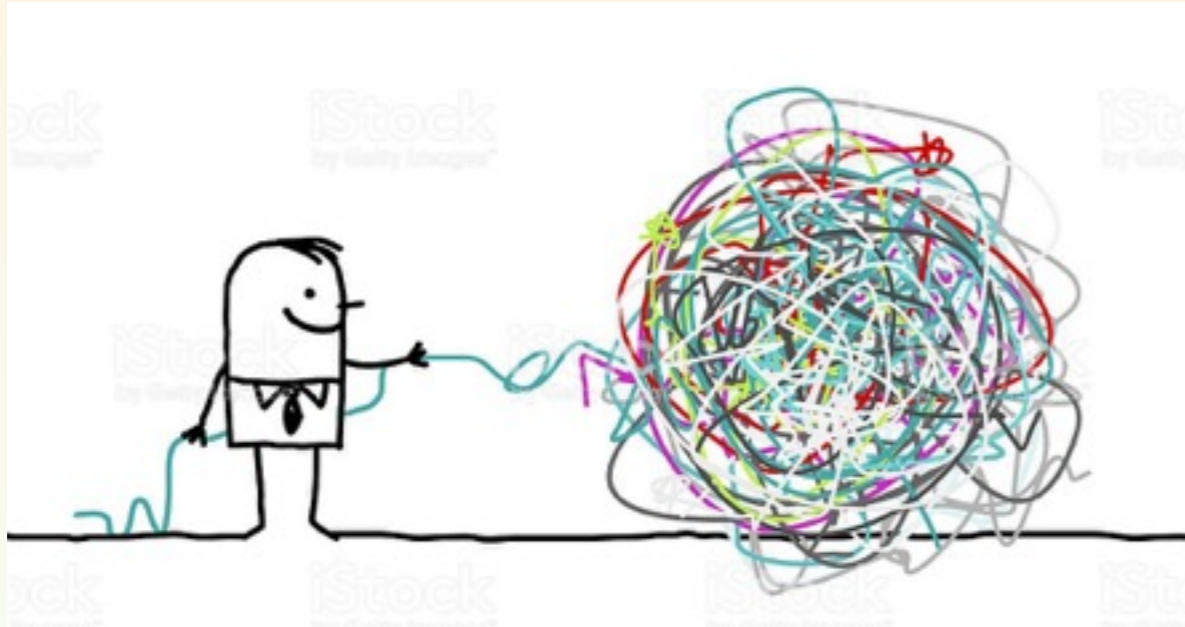


A tour of recent results on word transducers

Anca Muscholl



(based on joint work with F. Baschenis, O. Gauwin, G. Puppis)

Transductions

transform objects - here: **words**

transduction: mapping (or relation) from words to words

metamorphosis \longrightarrow mtmrphss **erase vowels**

metamorphosis \longrightarrow sisohpromatem **mirror**

metamorphosis \longrightarrow metamorphosismetamorphosis **duplicate**

metamorphosis \longrightarrow phosismetamor **permute circularly**

Transductions: some history

Early notion in formal language theory, motivated by coding theory, compilation, linguistics,...:

Moore 1956 “Gedankenexperimente on sequential machines”

Schützenberger 1961, Ginsburg-Rose 1966, Nivat 1968, Aho-Hopcroft-Ullman 1969, Engelfriet 1972, Eilenberg 1976, Choffrut 1977, Berstel 1979.

Extended later to more general objects, in particular to graphs.
Logical transductions are crucial (Courcelle 1994).

Transducers

1DFT, 1NFT: **one-way** (non-)deterministic finite-state transducers

metamorphosis \longrightarrow mtmrphss erase vowels

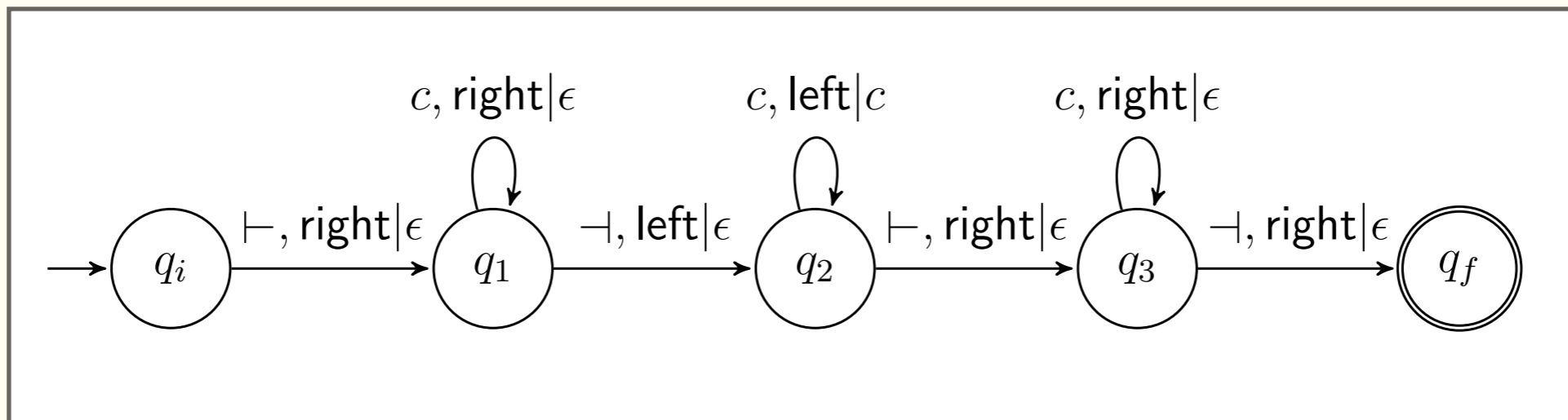
2DFT, 2NFT: **two-way** (non-)deterministic finite-state transducers

metamorphosis \longrightarrow sisohpromatem mirror

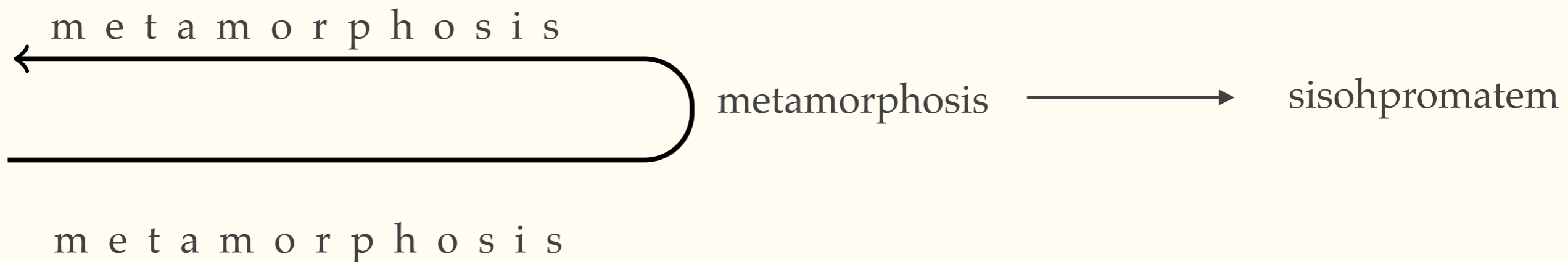
metamorphosis \longrightarrow metamorphosismetamorphosis duplicate

Transduction: binary **relation** over words

Above: functions



2DFT (= deterministic, 2-way) computing the mirror



Logic

MSOT: **monadic second-order** transductions [Courcelle, Engelfriet]

maps structures into structures

- ❖ fixed number of **copies** of input positions
- ❖ **domain** formula: unary MSO formula “c-th copy of input position belongs to the output and is labeled by a ”
- ❖ **order** formula: binary MSO formula “c-th copy of position x precedes the d-th copy of position y in the output”

Logic

MSOT: **monadic second-order** transductions [Courcelle, Engelfriet]

Ex: mirror

- ❖ **domain** formula: $\text{dom}_a(x) \equiv a(x)$
- ❖ **order** formula: $\text{Before}(x, y) = (x > y)$

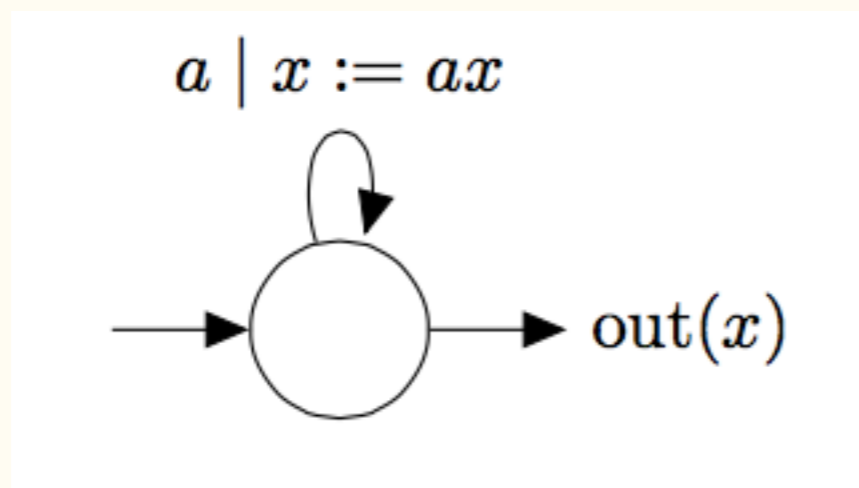
[Engelfriet-Hoogeboom 2001]: MSOT = 2DFT

Streaming transducers SST = MSOT

SST: streaming string transducers [Alur-Cerny 2010]

- ❖ **one-way** automata +
- ❖ finite number of (copyless) **registers**: output can be appended left or right, registers can be concatenated

mirror



metamorphosis



sisohpromatem

Relational transductions

1DFT

$$aW \mapsto Wa$$

2DFT = DSST = MSOT

$$W \mapsto WW$$

1NFT

$$W \mapsto \Sigma^{|W|}$$

decidable equivalence
undecidable equivalence

2NFT

$$W \mapsto W^*$$

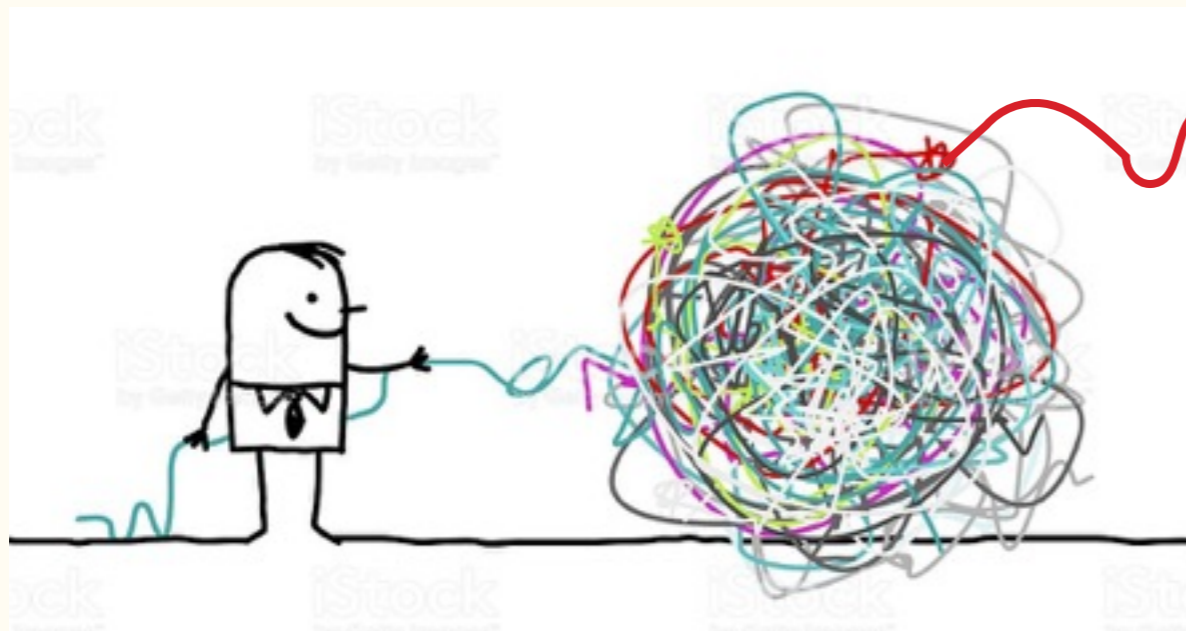
$$uv \mapsto vu$$

NSST = NMSOT

Equivalence problem

A transducer is **functional** (single-valued) if every input has at most one output.

- ❖ [Griffiths'68]: Equivalence of 1NFT is undecidable.
- ❖ [Gurari'82]: Equivalence of 2DFT (DSST [Alur-Cerny]) is PSPACE-c.
- ❖ [Gurari-Ibarra'83]: Equivalence of **functional** 1NFT is in PTime.
- ❖ [Culik-Karhumäki'87] Equivalence of **functional** 2NFT is decidable.
(PSPACE-c, because of equivalence of 2NFA is in PSPACE, Vardi'89)
- ❖ [Alur-Deshmukh'11] Equivalence of **functional** NSST is PSPACE-c.



Word

functions

Functionality

A transducer is **functional** if every input has at most one output.

Checking functionality

1NFT: [Schützenberger'75, Gurari-Ibarra'83] PTime

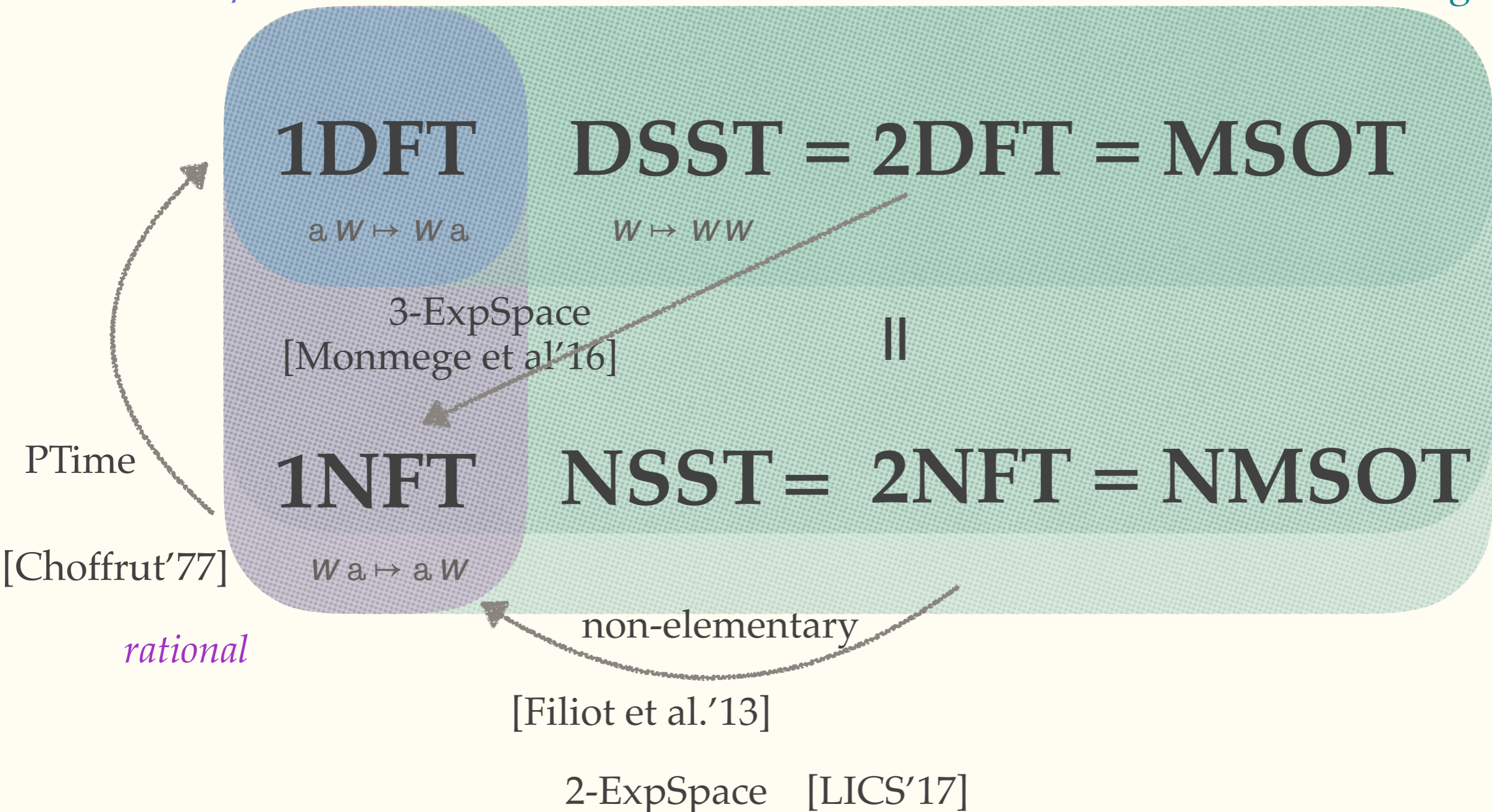
2NFT: [Culik-Karhumäki'87] decidable (actually PSPACE-c)

NSST: [Alur-Deshmukh'11] PSPACE-c

Functional transductions

subsequential

regular



Translations and recent results

- From DSST to 2DFT: **PTime**

(based on new results [Dartois, Fournier, Jecker, Lhote 2017])

- ❖ decompose DSST as 2DFT o 1DFT (poly-size)
- ❖ 1DFT can be made reversible with quadratic blow-up
- ❖ composition with reversible 1DFT in PTime

- From functional 2NFT to (reversible) 2DFT: **ExpTime** [DFJL'17]

- From 2DFT to DSST: **ExpTime**

Recent results

[Dartois, Fournier, Jecker, Lhote 2017]

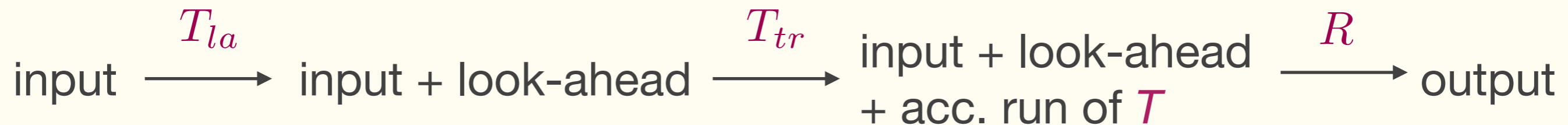
- ❖ composition of reversible 2DFT in PTime (easy)
- ❖ any 2DFT T can be made reversible with exponential blow-up:

exp-size, co-deterministic “look-ahead” 1NFT T_{la}

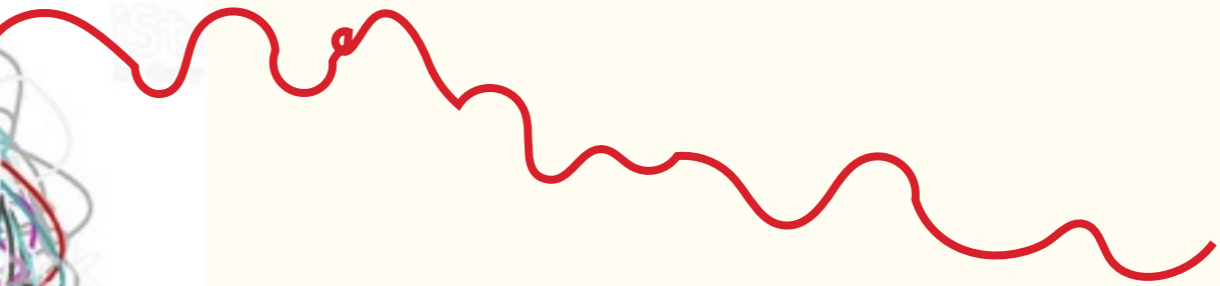
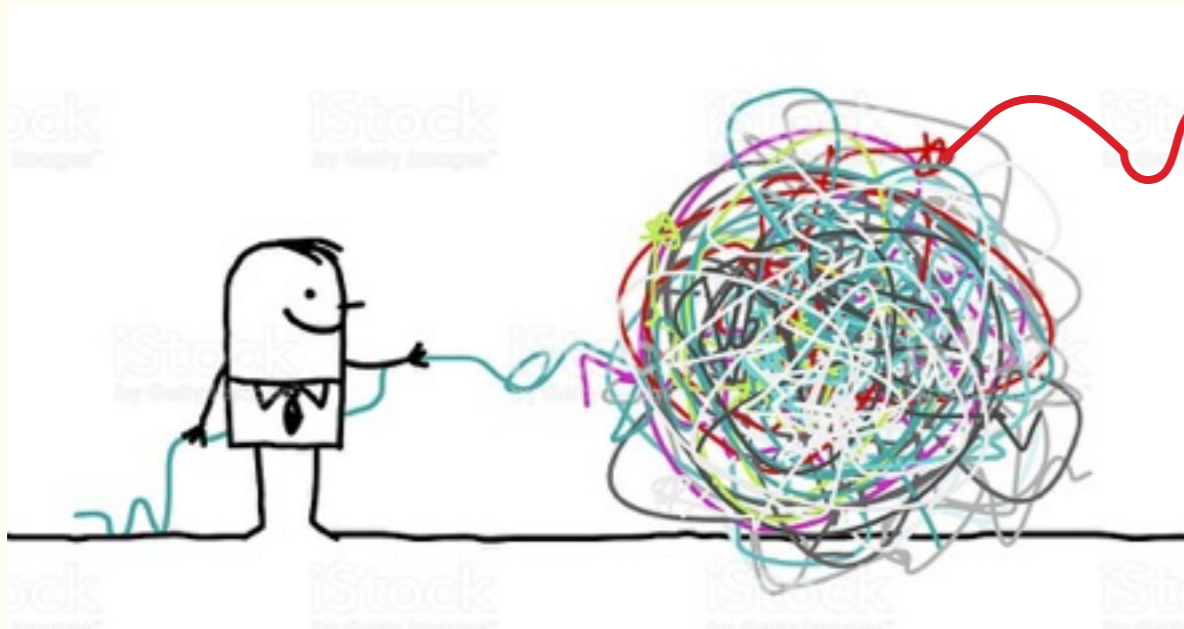
exp-size 1DFT T_{tr} outputs acc. run

reversible 2DFT R does the output

make reversible



I. Streaming



Streaming

- ❖ Wealth of research on external memory algorithms
[Mutukrishnan, Henzinger, Aggarwal, Grohe, Magniez]
- ❖ Large input in external memory
- ❖ Random access is more expensive than streaming (= one pass)
- ❖ Few sequential passes acceptable

Streaming string transducers have efficient processing, but still need memory for registers and updates...

... 1DFT and 1NFT are more attractive

Functional transductions

1DFT

$aW \mapsto Wa$

DSST = 2DFT = MSOT

$W \mapsto WW$

||

1NFT

$Wa \mapsto aW$

NSST = 2NFT = NMSOT

non-elementary

[Filiot et al.'13]

2-ExpSpace [LICS'17]

2NFT to 1NFT: example

Fix a regular language R .

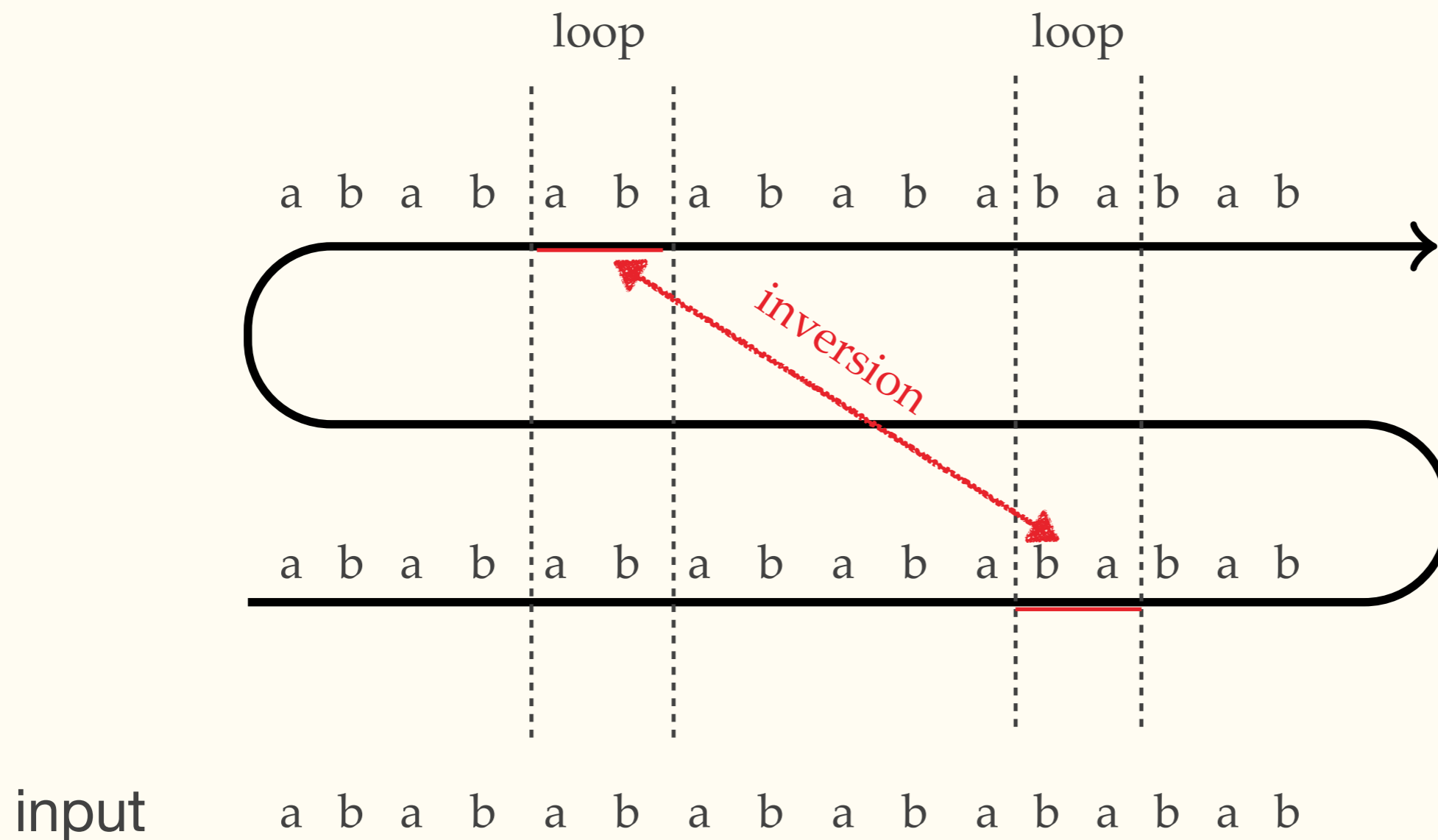
$$F(w) = ww \quad \text{if } w \text{ in } R \quad \text{2DFT}$$

F can be implemented by some 1NFT iff there is some B such that every word of R has period B .

Example: $R = (ab)^*$ 1DFT outputs “abab” for each “ab”

$$F(w) = ww \quad \text{if } w \text{ in } R$$

F can be implemented by some 1NFT iff for some **bounded** integer B: every word in R has period B.



The result (LICS'17):

Given a functional **2NFT** T :

- ❖ it is decidable in 2-ExpSpace whether an equivalent **1NFT** exists
- ❖ if “yes”: construction of 3-exp size equivalent **1NFT**
- ❖ if T is **sweeping**: one exponential less

Lower bounds

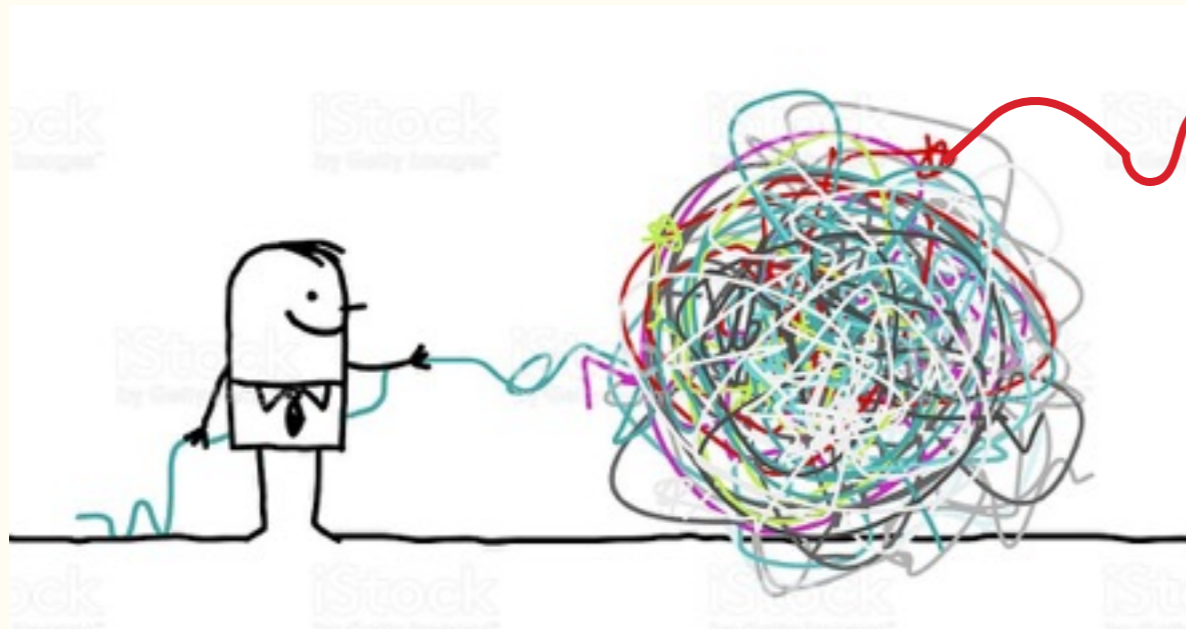
- ❖ PSPACE for the decision procedure
- ❖ 2EXP for the size of the output (1NFT)

Remark: The problem is undecidable without functionality [FSTTCS'15]

Open problems

- ❖ PSPACE lower bound for decision procedure “2NFT to 1NFT” - better lower bound?
- ❖ Better upper bound?
- ❖ Better complexity for “2NFT to 1DFT”?
- ❖ Extension from functional 2NFT to k-valued 2NFT?

II. Minimizing



passes

Sweeping transducers

- ❖ Sweeping: left-to-right, right-to-left passes
- ❖ less expressive than 2NFT:

example: reversing a list

$$u_1 \# u_2 \# \cdots u_n \longrightarrow u_n \# \cdots u_2 \# u_1$$

From 2NFT to Sweeping

- ❖ Given functional 2NFT \mathbf{T} and integer \mathbf{k} . It is decidable in 2ExpSPACE (poly space in \mathbf{k}) if \mathbf{T} is equivalent to some \mathbf{k} -pass sweeping transducer.
 - ❖ Given functional 2NFT \mathbf{T} . If \mathbf{T} is equivalent to some \mathbf{k} -pass sweeping transducer, then we can assume that \mathbf{k} is exponentially bounded.
- ❖ Given functional 2NFT \mathbf{T} . It is decidable in 2ExpSPACE if \mathbf{T} is equivalent to some sweeping transducer.

Sweeping transducers vs. SST

- ❖ Given a functional sweeping transducer T . Let k be minimal such that T is equivalent to some k -pass sweeping transducer. Then k can be computed in ExpSPACE.
- ❖ Tight connection between sweeping transducers and **concatenation-free SSTs**:
 $2k$ passes = k registers
- ❖ Given a functional, **concatenation-free** SST T . Let k be minimal such that T is equivalent to some k -register concatenation-free SST. Then k can be computed in 2-ExpSPACE.

Open questions

- ❖ Compute minimal number of registers for deterministic SST
- ❖ Decomposition theorem for k -valued SST?
- ❖ Decidability of equivalence for k -valued SST?

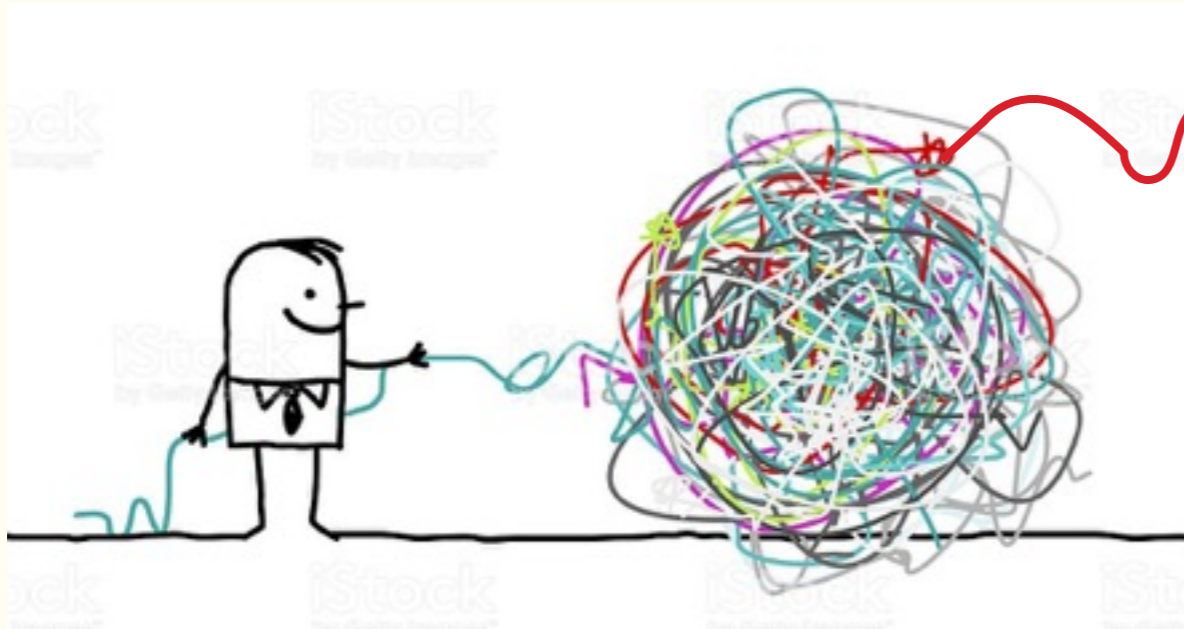
[Weber'96, Sakarovitch, de Souza'08]

Every k -valued 1NFT can be decomposed into k functional 1NFTs.

[Culik, Karhumäki'86]

Equivalence of k -valued 2NFT is decidable.

III. Algebra



Algebra

Long line of research on algebra for regular languages:

- ❖ algebra offers machine-independent characterizations, canonical objects, minimization, decision procedures for subclasses
- ❖ prominent example: decide whether a regular language is star-free

[Schützenberger'65] star-free = aperiodicity

[McNaughton, Papert'71] star-free = first-order logic

Algebra for transducers?

❖ A Myhill-Nerode theorem for **1DFT**... [Choffrut'79]

... thus a canonical (minimal) **1DFT**

❖ **1NFT**

Any **1NFT** is equivalent to the composition $D \circ R$ of a 1DFT D with a co-deterministic 1NFT R . [Elgot, Mezei'65]

Bimachine: DFA L + co-deterministic NFA R +
output function $\text{out}(\text{letter}, L\text{-state}, R\text{-state})$

For every **2NFT** there is a canonical bimachine.

[Reutenauer, Schützenberger'91]

Recent results

- ❖ **1NFT**: equivalent to order-preserving MSOT

Given a 1NFT it is decidable whether it is equivalent to an order-preserving FOT (first-order transduction).

[Filiot, Gauwin, Lhote'16]

proof uses canonical bimachines

- ❖ **2NFT** = MSOT: no decision procedure for FOT so far, but ...

A 2NFT is equivalent to some FOT iff it is equivalent to some aperiodic 2NFT iff it is equivalent to some aperiodic SST.

[Carton, Dartois'15], [Filiot, Krishna, Trivedi'15]

Conclusions

- ❖ This talk presented a selection of current work on word transducers.
- ❖ Goal of current work: develop a robust theory of word transducers and identify genuine algorithmic questions.

Beyond words...

- ❖ Transducers with origin [Bojanczyk'14]: record where the information comes from.

Less combinatorics involved, Myhill-Nerode theorem.

- ❖ Tree transducers: yet another story...

Thank you!

