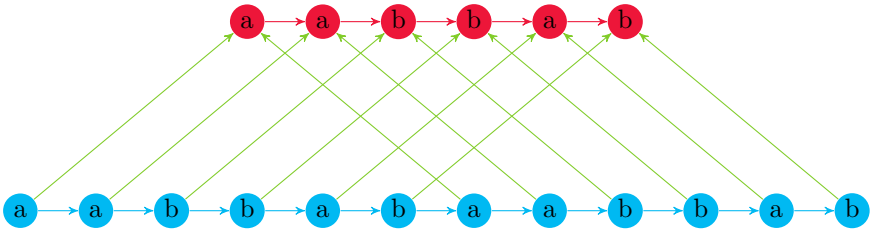
Which classes of origin graphs are generated by transducer?*

Bruno Guillon

Università degli studi di Milano

This presentation is about *string-to-string transductions* with *origin semantics*. A *string-to-string transduction* is a binary relation between strings over fixed input and output alphabets. Examples include the SQUARING transduction $w \mapsto ww$, or the SUBWORD transduction which is the set of pairs (u, v) such that v is a (possibly not connected) subword of u. Note that SQUARING is a function, while SUBWORD is a relation; both types are studied. The *origin semantics of a transduction* (technically speaking, of a device computing it) consists not only of pairs (u, v) of input and output words, but also gives an *origin mapping* that specifies which positions of the input word were used to produce which positions of the output word. For example, suppose that we model the SQUARING transduction by a two-way automaton which does two consecutive left-to-right passes over the input word during each of them it outputs a copy of the input word. In this case, the origin semantics over a particular input word can be visualised as follows:



An object as in the above picture is called an *origin graph*. We define an *origin string-to-string transduction* to be a set of origin graphs. Origin semantics were introduced in [Boj14], where it was shown that existing models of transducers, such as two-way finite state transducers, see for instance [EH01], MSO string-to-string transductions [CE12], or streaming string transducers [AC10, AD11] can be equipped with origin semantics so that they generate not sets of pairs of words, but sets of origin graphs. Furthermore, existing results on equivalence between models remain true when the origin semantics are used [Boj14]. This work aims to study sets of origin graphs that are origin semantics of transducers.

A first result states that given an MSO formula on origin graphs and a non-deterministic streaming string transducer, one can decide if the formula is true in some origin graph realised by the transducer. This result gives a generic

*This paper is part of LIPA, a project funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement N°683080).

framework for deciding questions like: is some input position used at least twice in some output? is the origin mapping order preserving? It is related to the recent work in [DFL17], in which the authors introduced a new logic for origin transductions (they use the term *productions*). The logic used is a strict fragment of MSO which has decidable satisfiability when evaluated over the class of all origin graphs.

The second result is a characterisation of the origin transductions which are realised by streaming string transducer. Our goal is to describe them in a machine independent way. The result states that a set of origin graphs is the origin semantics of some nondeterministic streaming string transducer if and only if it has three properties: *a*) it is MSO-definable as a set of coloured graphs; *b*) it has bounded degree; and *c*) it has bounded crossing, which means intuitively that the origin mapping does not oscillate too much.

We finally discuss how other models of transducer behaves with respect to these properties. In particular we allow ε-transitions and we consider two-way nondeterministic finite state transducers.

This is a joint work with Mikołaj Bojańczyk, Laure Daviaud and Vincent Penelle, that has been accepted at ICALP 2017.

References

[AC10] Rajeev Alur and Pavol Cerný. Expressiveness of streaming string transducers. In *FSTTCS*, pages 1–12, 2010.

[AD11] Rajeev Alur and Jyotirmoy V. Deshmukh. Nondeterministic streaming string transducers. In *International Colloquium on Automata, Languages, and Programming*, pages 1–20. Springer, 2011.

[Boj14] Mikołaj Bojańczyk. Transducers with origin information. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part II*, pages 26–37, 2014.

[CE12] Bruno Courcelle and Joost Engelfriet. *Graph Structure and Monadic Second-Order Logic — A Language-Theoretic Approach*, volume 138 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, 2012.

[DFL17] Luc Dartois, Emmanuel Filiot, and Nathan Lhote. Decidable logics for transductions and data words. *CoRR*, abs/1701.03670, 2017.

[EH01] Joost Engelfriet and Hendrik Jan Hoogeboom. Mso definable string transductions and two-way finite-state transducers. *ACM Trans. Comput. Log.*, 2(2):216–254, 2001.