LABORATOIRE BORDELAIS DE RECHERCHE EN INFORMATIQUE
UNIV. BORDEAUX, LaBRI, CNRS - UMR 5800, F-33400 TALENCE, FRANCE

Diego FIGUEIRA
Tel: +33 540006900
Fax: +33 540006669
E-mail: diego.figueira@labri.fr

Talence, September 19, 2023

Review report on the PhD thesis by **Rafał Stefański**

The Single-use Restriction for Register Automata and Transducers over Infinite Alphabets

under the supervision of Mikołaj Bojańczyk.

I have never met Rafał Stefański but I know Mikołaj Bojańczyk, with whom I have worked more than a decade ago. I do not feel a conflict of interest writing this report.

The doctoral thesis of Rafał Stefański contributes to the foundational theory on sequences over infinite domains, and on transductions between such sequences. The underlying framework for the formalisms advanced in the thesis is the theory of *sets with atoms* and *orbit-finiteness* which, loosely speaking, allow to generalize constructions to all kind of sets which may be infinite, but finite "modulo permutations".

The thesis proposes and studies subclasses of automata with registers over sets with atoms (called *deterministic orbit-finite automata*) and of its extension as transducers. The definitions are based on the key notion of "single-use functions" (Section 2.2) and k-fold-use functions. Several automata models stem from this defintion, such as the "single-use automata" (Section 2.3) and "single-use Mealy machine" (Section 3.1). Single-use functions are shown to be equivalent to "single-use decision trees", which is a key fact to show that single-use functions over polynomial orbit finite sets are always orbit-finite (Theorem 5). It is further shown that the "non-copy restriction" is, in some sense, maximal, since adding a "copy function" yields already the class of *all* finitely supported functions between polynomial orbit-finite sets.

The thesis then studies language acceptors, showing for example how the expressive power of automata models change depending on the acceptance condition for defining single-use automata: either by configuration, by equivariance-acceptance function, or by single-use function. The thesis proposes a definition of "single-use automata" (slightly) generalizing the model of "single-use register automata", arguably of simpler definition. It is shown that the focus on *single-use automata* is well-justified by its equivalence to orbit-finite-

monoids (as introduced by Bojańczyk in 2013) and to its two-way counterpart of *two-way single-use automata* (Theorem 6). The definition admits two equivalent formulations. The first one by specifying in advance the number of input letter copies needed. The second one uses k-fold use functions into trees (introduced to this end). To show the equivalence between the three formalisms, the thesis first shows that two-way single-use automata are recognized by orbit-finite monoids, in a construction akin to Shepherdson's for translating two-way finite word automata into one-way automata. And that, in turn, showing that languages recognized by polynomial orbit-finite monoids can be translated into one-way single-use automata.

The thesis later explores other automata definitions, this time for defining *transductions*. The focus of study is first put in "single-use Mealy machines", a class of length-preserving transducers. Perhaps the main technical contribution of this thesis is that single-use Mealy machines enjoy a Krohn-Rhodes decomposition intro "prime languages". The prime languages include the ones on finite alphabets (from the Krohn-Rhodes theorem in the classical setting) plus two other sorts of languagess:

- single-use propagations, and
- letter-to-letter "single-use" homomorphisms.

The proof that single-use Mealy machines can always be decomposed into primes is far from obvious and uses the theory of single-use machines developed in Chapter 2. The fact of finding the right definitions for the primes, in particular of single-use homomorphism seems also non-trivial. Further, it necessitates a characterization of when orbit-finite semigroup transduction are definable by single-use Mealy machines, via a notion of "locality", in the sense that local-semigroup transductions can be defined as compositions of primes (and additionally that they characterize single-use Mealy machines).

In his thesis, Stefański also studies other issues related to local semigroup transductions, such as definability and a suitable definition for the "syntactic semigroup". The thesis also defines "rational transductions" over polynomial orbit-finite alphabets using an algebraic approach, and proves that, once again, a Krohn-Rhodes decomposition holds in this case.

Lastly, the thesis investigates two-way transductions. It defines four equivalent models:

1. single-use two-way transducers,
2. single-use streaming string transducers,
3. regular list transductions with atoms, and
4. compositions of single-use two-way primes (based on the Krohn-Rhodes decomposition theorems proved before).

I find the robustness of this class of transducers quite remarkable, especially in the context of infinite alphabets, in which the usual rule of thumb is that "all models are incomparable". The thesis shows that this class is further closed under compositions and has decidable equivalence. In view of the robustness and properties enjoyed by this class, the author argues, I think with reason, that this class deserves the name of "regular transductions with atoms". The thesis finishes with some interesting future research lines.

I believe that the main contribution of this thesis, namely the exploration of automata and transducers under a "single-use" regime, and the many properties enjoyed by these, is very relevant since it establishes fundamental models extending the ones on finite words which preserve most desirable properties. The proofs are rigorous and correct to the best that I could check. Several proofs are innovative and clearly document a high degree of mathematical ability and maturity.

The main contributions of the thesis are based on the paper "Single-Use Automata and Transducers for Infinite Alphabets" which appeared in the top-notch conference ICALP in 2020. I would like to stress that the cited paper is very rich in results —the online version with "full proofs" is a technically dense 60-page long manuscript— and was well-received by the community.

The thesis is generally well-written, well-organized, and didactic. I have particularly appreciated the many useful examples, which are important to grasp an intuition in this rather abstract setting of sets with atoms (for instance, the examples on single-use functions from Section 2.2.2).

For the reasons above, **I deem the thesis as sufficient to grant a PhD**.

Diego FIGUEIRA
CR CNRS HDR
Univ Bordeaux