Duality and the equational theory of regular languages

The 14th of november, 2010.

Yann Pequignot
yann.pequignot@unil.ch

First year PhD student, *Université de Lausanne, Switzerland.*

Supervisor: Jacques Duparc, *Université de Lausanne, Switzerland.*
Co-supervisor: Jean-Éric Pin, *LIAFA, University Paris-Diderot and CNRS, France.*

Formal languages
The theory of formal languages is basically concerned with the description of properties of sequences of symbols. Since properties of sequences are nothing else than sets of sequences, we have the following definition. Fixing a finite set A of symbols called an alphabet, a (formal) language is a set of finite sequences of symbols in A. In this context, finite sequences are called words and are naturally endowed with the concatenation operation. The set of all words on the alphabet A forms the free monoid A^* for the concatenation with the empty word as the neutral element.

From recognisability to classes of languages
The specification of a language requires an unambiguous description of which word belong to that language. One way to achieve such a specification is by using a machine that recognises the language. An other way to specify a language is through a formula of a logic whose models are words defining that language. In each case any computational model or any logic on words defines a class of languages, namely the class of languages which are respectively recognisable by an instance of this computational model and definable by a formula of this logic. This connection of the theory of formal languages with logic and especially theoretical computer science motivates the study of classes of languages.

Regular languages and algebra
Here we restrict our attention to the simplest model of computation: the finite state automaton. The languages recognised by a finite automaton form the class of *regular languages*. The class of regular languages is also the class of languages definable by a formula of the monadic second order logic on words. However, exploiting the fact that A^* is a monoid leads to an algebraic definition of the class of regular languages. Indeed, a subset $L \subseteq A^*$ is regular if and only if it is saturated by a monoid congruence of finite index. In other terms, a language $L \subseteq A^*$ is regular if and only if there exists a surjective monoid morphism $\varphi : A^* \to M$ onto a finite monoid M that *recognises* L, i.e. such that $\varphi^{-1}(\varphi(L)) = L$. In this sense, the regular languages are the languages recognised by finite monoids.

Classes of regular languages and pseudovarieties
Since we focus here on regular languages, the study of classes of languages restricts to the study of classes of regular languages. By the algebraic characterisation of regular languages, many classes of regular languages appear as the languages recognised by classes of finite monoids. Of particular interest in algebra are the classes of finite monoids closed under taking submonoids, quotient monoids, and finite direct products. Such classes of finite monoids are called *pseudovarieties*. Every pseudovariety of monoids defines on any finite alphabet a class of regular languages which is a boolean algebra, closed by inverse of morphisms and by left and right quotients by words. It is a theorem of Eilenberg [4] that such classes of regular languages are in one-to-one correspondence with pseudovarieties of monoids.

One of the most important results dealing with recognisable languages is an instance of this correspondence. This result is a theorem of Schützenberger [10] which says that the class of star-free languages, which is also the class of languages definable by a first order formula, is associated via this correspondence to the pseudovariety of aperiodic monoids.

Equational descriptions
In order to render membership in various classes of automata decidable, we are interested in equational properties characterising different pseudovarieties of monoids. In contrast to the notion of a pseudovariety, a *variety* of monoids is a class of (not necessarily finite) monoids closed under taking submonoids, quotients, and direct products. It is a theorem of Birkhoff [3] that any variety of monoids can be characterised by a set of equations. For instance, the variety of commutative monoids is described by the equation $xy = yx$. An obvious question in this context, is whether there is a similar characterisation for pseudovarieties. However, fifty years from the theorem of Birkhoff were necessary to find a satisfactory answer in the case of pseudovarieties.

As it was notably observed by Almeida [1, 2] one problem encountered in the quest for equations characterising pseudovarieties is that, as categories of finite monoids, pseudovarieties in general lack free objects. On a finite alphabet A such a free object for finite monoids can be obtained as follows. Define on A^* a metric for which two words are close if a large finite monoid is necessary to distinguish them. The completion of this metric space is an infinite compact monoid called the *free profinite monoid* on A. The points in this metric space are called *profinite words*. Then Reiterman's theorem [9] states that pseudovarieties of finite monoids are characterised by sets of profinite equations, i.e. formal equalities between profinite words.

Finally, the combination of the theorems of Eilenberg and Reiterman leads to an equational description of various classes of regular languages in profinite terms.

Duality and equational theory of regular languages
Roughly speaking, Stone duality [11, 7] states that the category of Boolean algebras with Boolean morphisms is equivalent to the category of Hausdorff compact zero dimensional topological space with continuous maps in which we have simply reverse every arrow. It was observed explicitly by Pippenger [8] that the topological space underlying the free profinite monoid on an alphabet A is the Stone dual of the boolean algebra of regular languages on A^*. This observation led M. Gehrke, S. Grigorieff and J.-É. Pin [6, 5] to consider the combination of Eilenberg's theorem and Reiterman's theorem in terms of duality. The duality perspective is mainly based on the crucial observation that the product operation on the free profinite monoid arise as the dual of supplementary operations on the boolean algebra of regular languages.

References

[1] J. Almeida. Profinite semigroups and applications. *Structural theory of automata, semigroups, and universal algebra*, pages 1–45, 2005.

[2] J. Almeida and P. Weil. Relatively free profinite monoids: an introduction and examples. *Semigroups, formal languages and groups*, 466:73–117, 1995.

[3] G. Birkhoff. On the structure of abstract algebras. *Proc. Cambridge Phil. Soc.*, 31:433–454, 1935.

[4] S. Eilenberg. *Automata, languages, and machines*, volume A et B. Academic Press, 1974.

[5] M. Gehrke. Stone duality and the recognisable languages over an algebra. *Algebra and Coalgebra in Computer Science*, pages 236–250, 2009.

[6] M. Gehrke, S. Grigorieff, and J.É. Pin. Duality and equational theory of regular languages. *Lect. Notes Comp. Sci. (ICALP)*, 5125:246–257, 2008.

[7] P.T. Johnstone. *Stone spaces*. Cambridge Univ Pr, 1986.

[8] N. Pippeger. Regular languages and Stone duality. *Theory of Computing Systems*, 30(2):121–134, 1997.

[9] J. Reiterman. The birkhoff theorem for finite algebras. *Algebra Universalis*, 14(1):1–10, 1982.

[10] M. P. Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.

[11] M.H. Stone. The theory of representation for Boolean algebras. *Transactions of the American Mathematical Society*, 40(1):37–111, 1936.