Quantifiers on languages and codensity monads *

Luca Reggio

IRIF, Université Paris Diderot and Sorbonne Paris Cité, France
luca.reggio@irif.fr

The main content of this talk concerns recent joint work (see [4]) with Mai Gehrke and Daniela Petrişan on the understanding, at the level of recognisers, of the effect of applying a layer of various kinds of quantifiers in the context of logic on words.

Two approaches have been remarkably effective in the study of languages: the algebraic one, and the logical one. Whereas the former relies on the notions of recognition by a monoid and of syntactic monoid of a language, the latter is based on a semantic on finite words. Let us briefly recall these two approaches.

Consider a finite set A (the *alphabet*) and an A-language, i.e. a subset L of the monoid A^* free on A. We say that a monoid M *recognises* the language L provided there is a monoid morphism $\phi\colon A^* \to M$ and $P \subseteq M$ such that $\phi^{-1}(P) = L$. This condition is equivalent to the existence of a homomorphism $A^* \to M$ whose kernel saturates L. The maximal congruence $\sim_L$ on A^* saturating L is defined by $(x, y) \in \sim_L$ if $uxv \in L \Leftrightarrow uyv \in L$ for all $u, v \in A^*$. The quotient $A^*/\sim_L$ is called the *syntactic monoid of* L, and one can define a *regular* language to be one whose syntactic monoid is finite.

It turns out that, beyond the regular case, monoids do not provide a notion of recognition that is fine-grained enough to be useful. This led us to introduce in [3] the notion of a *Boolean space with an internal monoid* (*BM*, for short), which behaves well with respect to recognition in the non-regular setting. A *BM* is a pair (X, M) given by a Boolean space X (i.e, a compact and Hausdorff space that is zero-dimensional) along with a dense subset M carrying a monoid structure, such that $\forall m \in M$ the maps $\lambda_m, \rho_m\colon M \to M$ given by left and right multiplication by m, respectively, can be extended to continuous functions on X. An example is provided by the pair $(\beta(A^*), A^*)$, where $\beta(A^*)$ is the Stone-Čech compactification of the discrete set A^*. Now, define a morphism $(X, M) \to (Y, N)$ to be a continuous function $X \to Y$ whose restriction is a monoid morphism from M to N. Recalling the bijection $L \mapsto \widehat{L}$ between $\mathcal{P}(A^*)$ and the clopens of $\beta(A^*)$, we say that a *BM* (X, M) *recognises* the language L if there is a morphism $\phi\colon (\beta(A^*), A^*) \to (X, M)$ and a clopen subset $C \subseteq X$ such that $\phi^{-1}(C) = \widehat{L}$. This extends the classical definition of recognition in the regular case.

The second approach stems from the interpretation of a word $w \in A^*$, say of length n, as a relational structure on the set $\{1, \ldots, n\}$. These structures are equipped with (interpretations of) unary relations P_a, one for each $a \in A$, selecting the positions in the word w in which the letter a appears. Additional relations, such as the natural order on $\{1, \ldots, n\}$, are sometimes considered in specific situations. Every (first-order, or higher-order) sentence ψ in a language interpretable over words determines a language $L_\psi \subseteq A^*$ consisting of all those words satisfying ψ. However, if $\psi(x)$ is a formula containing a free first-order variable x, in order to be able to interpret the free variable we extend the alphabet to $(A \times \{0, 1\})^*$ and use the more compact notation $a_1 a_2' a_3' \cdots a_n$ for the word $(a_1, 0)(a_2, 1)(a_3, 1) \cdots (a_n, 0) \in (A \times \{0, 1\})^*$. The language

*This work is part of the project DuaLL which has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No.670624). The author also acknowledges financial support from Sorbonne Paris Cité (PhD agreement USPC IDEX – REGGI15RDXMTSPC1GEHRKE).

$L_{\psi(x)} \subseteq (A \times \{0,1\})^*$ is then the collection of all the words in the extended alphabet, with *only one* marked position, in which the formula $\psi(x)$ is satisfied when the variable x points at that position. Finally, one can consider the quantified formula $\exists x.\psi(x)$ which yields the language over the alphabet A^* of all those words $a_1 \cdots a_n$ such that there exists $1 \leq i \leq n$ with $a_1 \cdots a_i' \cdots a_n \in L_{\psi(x)}$. There are other quantifiers of interest in language theory. An example is provided by *modular quantifiers*: a word w satisfies the sentence $\exists_{p \bmod q} x.\psi(x)$ if there are p mod q positions in the word w in which the formula $\psi(x)$ is satisfied.

The question we pose, and answer, is the following: Suppose a language, defined by a formula $\psi(x)$, is recognised by a *BM* (X, M). If Q is some quantifier (e.g. a modular quantifier), how can we construct a *BM* recognising the language associated to the sentence $Qx.\psi(x)$? The question is motivated by open problems on the separation of Boolean circuit complexity classes, where classes of languages are characterised in terms of logic fragments.

The answer employs duality-theoretic and categorical tools. Several quantifiers of interest can be modelled using commutative semirings S (e.g. $S = \mathbb{Z}/q\mathbb{Z}$ for the modular quantifiers) or, from a categorical viewpoint, the free S-semimodule monad on **Set** (=the category of sets and functions). On the way to our answer, we prove that whenever an operation on languages — quantification being a particular case — can be modelled by a finitary *commutative* monad (in the sense of [6]) T on **Set**, then a recogniser for the languages obtained by applying the operation represented by T can be built by means of the *profinite monad* $\widehat{T}$ on the category of Boolean spaces and continuous functions. The profinite monad $\widehat{T}$ associated to T was first defined in [1], building on the ideas introduced in [2], and it is based on the notion of *codensity monad* of a functor which has its origins in the work of Kock in the 60's (see also [5]).

In the case of quantifiers modelled by a finite and commutative semiring S, that is when T is the free S-semimodule monad, we provide a concrete description of the Boolean space $\widehat{T}X$, for X any Boolean space, in terms of certain S-valued *measures* on X. If in addition the semiring S is idempotent (hence a semilattice), $\widehat{T}X$ can be equivalently described as the space of all continuous functions $X \to S$, where S is equipped with the topology of all downsets with respect to its semilattice order. We remark that, in the case $S = \mathbb{2}$ is the two-element Boolean algebra, $\widehat{T}$ is the Vietoris monad on Boolean spaces (already related to the existential quantifier in [3]) and we essentially recover the classical description of the Vietoris space in terms of functions into the Sierpiński space.

References

[1] J. Adámek, L.-T. Chen, S. Milius, and H. Urbat. Profinite monads, profinite equations, and Reiterman's theorem. In *FoSSaCS*, volume 9634 of *Lecture Notes in Computer Science*, pages 531–547. Springer, 2016.

[2] M. Bojańczyk. Recognisable languages over monads. In *DLT*, volume 9168 of *Lecture Notes in Computer Science*, pages 1–13. Springer, 2015.

[3] M. Gehrke, D. Petrişan, and L. Reggio. The Schützenberger product for syntactic spaces. In *ICALP*, volume 55 of *LIPIcs*, pages 112:1–112:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.

[4] M. Gehrke, D. Petrişan, and L. Reggio. Quantifiers on languages and codensity monads. arXiv:1702.08841 [cs.LO], 2017.

[5] D. Gildenhuys and J. F. Kennison. Equational completion, model induced triples and pro-objects. *Journal of Pure and Applied Algebra*, 1(4):317 – 346, 1971.

[6] A. Kock. Monads on symmetric monoidal closed categories. *Journal of The Australian Mathematical Society*, 21:1–10, 1970.