Some results of Zoltán Ésik on regular languages

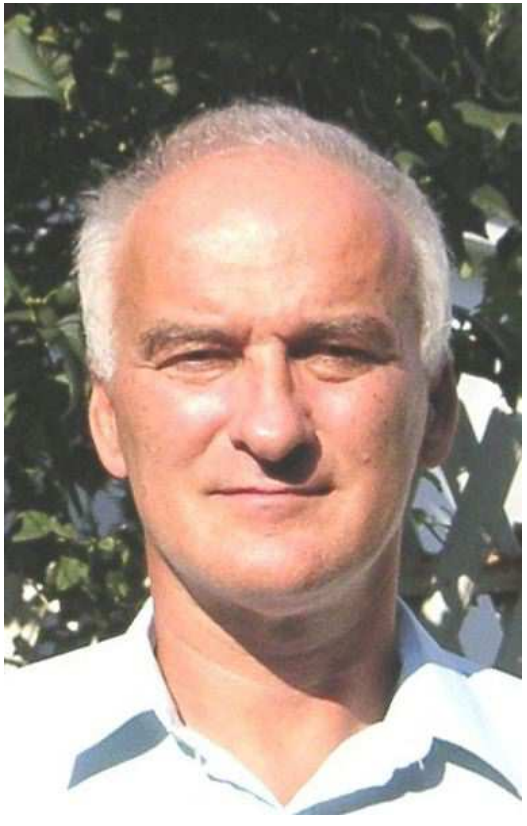
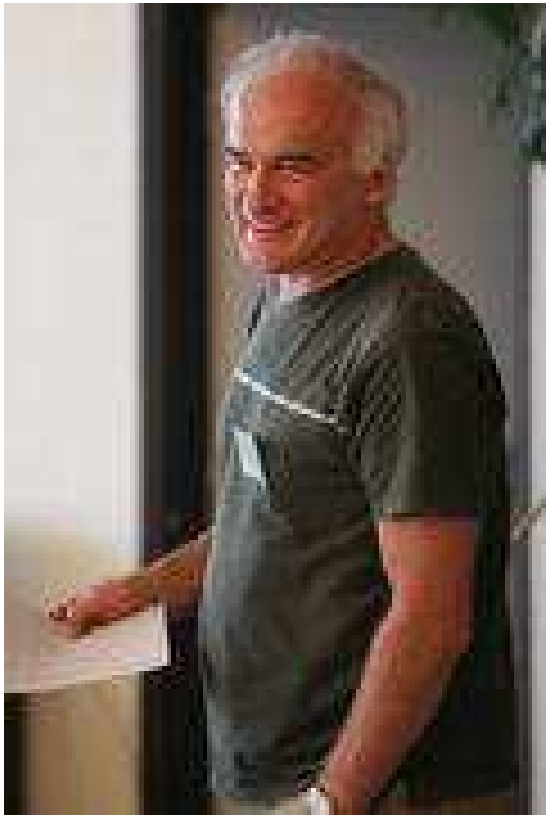
Jean-Éric Pin[1]

[1]IRIF, CNRS and University Paris Diderot

FCT, September 2017, Bordeaux

Zoltán Ésik

Zoltán Ésik passed away in Reykjavik, Iceland, on Wednesday, 25 May 2016.

Publications of Zoltán Ésik

Over 250 scientific works

- 2 books
  - Iteration Theories: The Equational Logic of Iterative Processes (with S. Bloom, 1993)
  - Modern Automata Theory (with W. Kuich, 2013)
- 32 edited volumes,
- 135 (+ at least 2) journal papers,
- 4 book chapters,
- 86 conference papers,
- 7 papers in other edited volumes.

Other tributes

Obituary for Zoltán Ésik by L. Aceto and A. Ingólfsdóttir (BEATCS 120, October 2016)

Logic and Automata Theory, A tribute to Zoltán Ésik (satellite workshop of CSL 2017, Stockholm, August 25, 2017):

- M. Bojanczyk: Algebras for tree languages,
- S. Ivan: Iterative, iteration and Conway semirings,
- W. Thomas: On iteration in logic,
- P. Weil: About recognizable languages of finite trees.

Outline

To avoid redundancy with the CSL workshop, I will only focus on a very small part of Zoltán's scientific work, related to regular languages:

- A solution to a twenty year old conjecture on the shuffle operation, obtained by Zoltán jointly with Imre Simon in 1998
- Zoltán's algebraic study of various fragments of logic on words,
- Some results on commutative languages obtained by Zoltán, J. Almeida and myself.

Preliminaries

A language L of A^* is recognised by a monoid M if there exists a monoid morphism $f : A^* \to M$ and a subset P of M such that $L = f^{-1}(P)$.

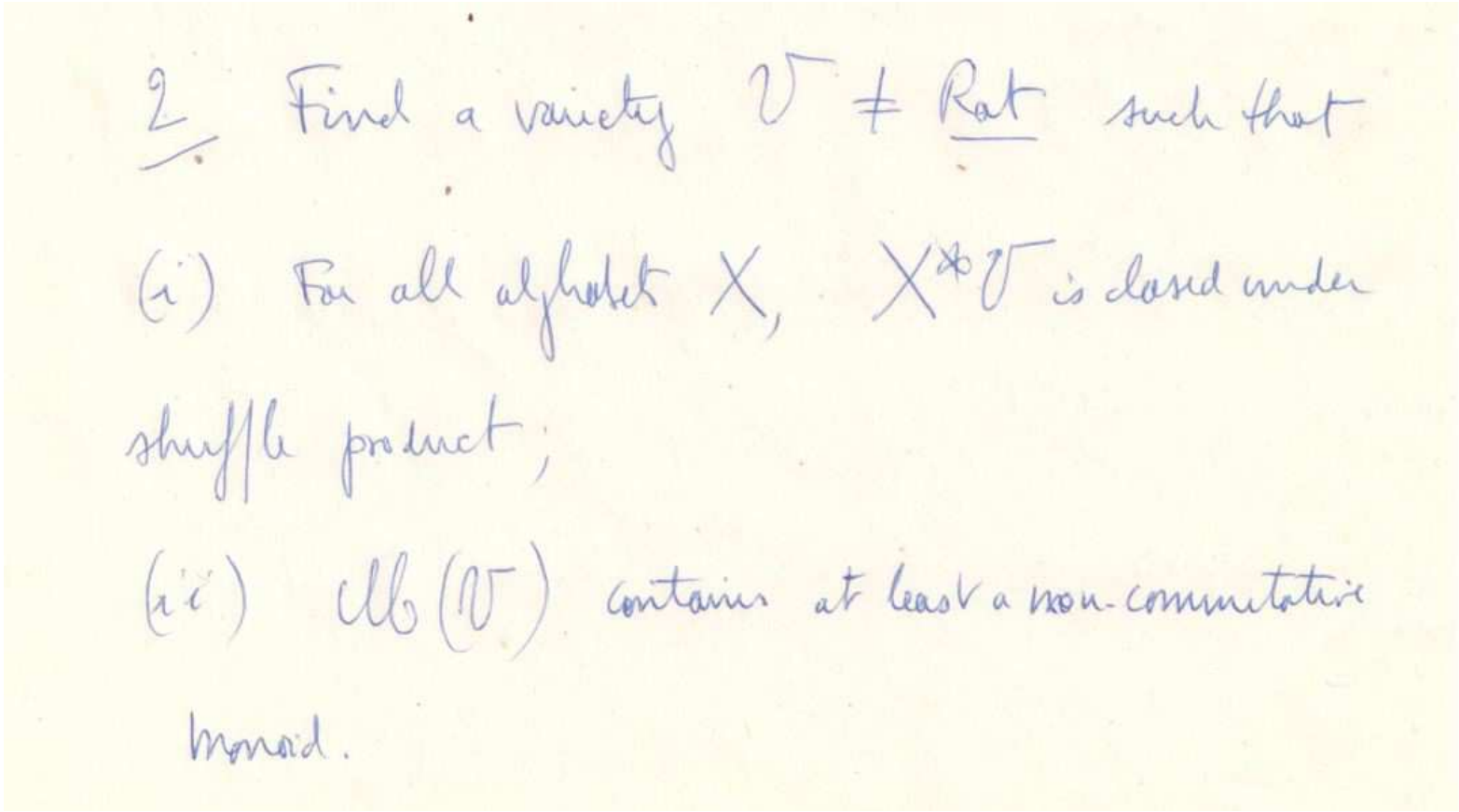
Just like there is a minimal DFA, there is a minimal monoid recognising a language, called its syntactic monoid.

Fact. A language is regular iff its syntactic monoid is finite.

Part I

Shuffle operation

Perrot's conjecture (September 1977)


2. Find a variety $\mathcal{V} \neq$ Rat such that

(i) For all alphabet X, $X^* \mathcal{V}$ is closed under shuffle product;

(ii) $Mb(\mathcal{V})$ contains at least a non-commutative monoid.

Is the variety of all regular languages the unique variety containing a non-commutative language and closed under shuffle?

Varieties

Variety of languages = class of regular languages closed under Boolean operations, quotients and inverses of morphisms.

Examples: Regular languages, star-free languages.

Variety of monoids = class of finite monoids closed under taking submonoids, quotients and finite products.

A language L is commutative if any word obtained by permuting the letters of a word of L is also in L.

A variety of languages is commutative if all of its languages are commutative.

Eilenberg's variety theorem

Given a variety of monoids $\mathbf{V}$, let $\mathcal{V}(\mathbf{V})$ be the class of languages whose syntactic monoid belongs to $\mathbf{V}$.

Given a variety of languages $\mathcal{V}$, let $\mathbf{V}(\mathcal{V})$ be the variety of monoids generated by the syntactic monoids of the languages of $\mathcal{V}$.

Theorem (Eilenberg 1976)

The maps $\mathbf{V} \to \mathcal{V}(\mathbf{V})$ and $\mathcal{V} \to \mathbf{V}(\mathcal{V})$ are mutually inverse, order preserving, bijections between varieties of monoids and varieties of languages.

Shuffle

The shuffle of two words u and v is the set $u \sqcup\!\sqcup v$ of words of the form $u_1 v_1 \cdots u_n v_n$, with $n \geqslant 0$, $u_1 \cdots u_n = u$, $v_1 \cdots v_n = v$.

Example $ab \sqcup\!\sqcup ba = \{ abba, baab, baba, abab \}$

The shuffle of two languages K and L is the set

$$K \sqcup\!\sqcup L = \bigcup_{u \in K,\ v \in L} u \sqcup\!\sqcup v$$

The smallest variety closed under shuffle

For each $a \in A$ and $k \geqslant 0$, let

$$L(a,k) = \{u \in A^* \mid |u|_a = k\}$$

and let $\mathcal{A}com(A^*)$ be the Boolean algebra generated by the languages $L(a,k)$.

Proposition (Perrot 1978)

The variety $\mathcal{A}com$ is the smallest nontrivial variety of languages closed under shuffle. It corresponds to the variety of commutative and aperiodic monoids.

Commutative varieties closed under shuffle

Given a variety of groups $\mathbf{H}$, let $\overline{\mathbf{H}}$ be the variety of monoids all of which subgroups are in $\mathbf{H}$.

Theorem (Perrot 1978)

A commutative variety of languages is closed under shuffle iff the corresponding variety of monoids is of the form $\mathbf{Com} \cap \overline{\mathbf{H}}$.

Another early result

Given a variety of languages $\mathcal{V}$, let $S\mathcal{V}$ be the variety generated by $\mathcal{V}$ and by the languages of the form $L_1 \sqcup\!\sqcup L_2$, where $L_1, L_2 \in \mathcal{V}$.

Proposition (Perrot 1978)

If $\mathcal{V}$ contains a non-commutative language, then $S\mathcal{V}(\{a,b\}^*)$ contains the language $(ab)^*$.

Power monoids and shuffle

For each monoid M, the set $\mathcal{P}(M)$ of nonempty subsets of M is a monoid under the product given by

$$XY = \{xy \mid x \in X, y \in Y\}$$

Proposition

If L_1 is recognised by M_1 and L_2 is recognised by M_2, then $L_1 \sqcup\!\sqcup L_2$ is recognised by $\mathcal{P}(M_1 \times M_2)$.

Power monoids and renaming

A morphism from A^* to B^* is a renaming (or length-preserving morphism or litteral morphism) if it maps every letter to a letter.

Example. $f : \{a, b, c\}^* \to \{a, b\}^*$ where $f(a) = a$ and $f(b) = f(c) = b$.

Fact. Let f be a surjective renaming. If L is recognised by M, then $f(L)$ is recognised by $\mathcal{P}(M)$.

If $\mathbf{V}$ is a variety of monoids, let $\mathbf{PV}$ be the variety of monoids generated by the monoids $\mathcal{P}(M)$, where $M \in \mathbf{V}$.

Applying surjective renaming to varieties

Let $\mathcal{V}$ be a variety of languages and let $\mathbf{V}$ be the corresponding variety of monoids. Let $R\mathcal{V}(A^*)$ be the Boolean algebra generated by the languages of the form $f(L)$, where $f : B^* \to A^*$ is a surjective renaming and $L \in \mathcal{V}(B^*)$.

Proposition (Reutenauer 79, Straubing 79)

$R\mathcal{V}$ is a variety of languages and the corresponding variety of monoids is $\mathbf{PV}$.

Varieties containing $(ab)^*$

Proposition (P. 80)

If a variety of languages contains the language $(ab)^*$, then $R\mathcal{V}$ is the variety of all languages.

Shuffle and power monoids

Given a variety of languages $\mathcal{V}$, let $S\mathcal{V}$ be the smallest variety containing $\mathcal{V}$ and closed under shuffle.

Theorem (Esik-Simon 1998)

If $\mathcal{V}$ contains a noncommutative language, then $S\mathcal{V}$ is the class of all regular languages.

Key argument: If $f : A^* \to B^*$ is a surjective renaming and $L \in \mathcal{V}(A^*)$, then $f(L) \in S\mathcal{V}(B^*)$.
It follows that $S\mathcal{V}$ contains $R\mathcal{V}$.

Modeling renaming by shuffle

Let $C = A \cup \{c\}$. Let L be a language of A^* and let

$$L_1 = L \sqcup\!\sqcup c^*, \quad L_2 = L_1 \cap (Ac)^*$$

Then $L_2 = g(L)$ where $g(a_1 \cdots a_k) = a_1 c \cdots a_k c$.

For each $b \in B$, let $f^{-1}(b) = \{a_{i_1}, a_{i_2}, \ldots, a_{i_b}\}$ and let $h : B^* \to C^*$ be the morphism defined by

$$h(b) = a_{i_1} a_{i_2} \cdots a_{i_b} c.$$

Then a magic formula holds:

$$f(L) = h^{-1}(L_2 \sqcup\!\sqcup A^*)$$

Varieties of languages closed under shuffle

The varieties of monoids corresponding to the varieties of languages closed under shuffle are

(1) The trivial variety,

(2) The varieties of the form $\mathbf{Com} \cap \overline{\mathbf{H}}$,

(3) The variety of all finite monoids.

Part II

Logic on words

Two articles

In December 2001, Zoltán Ésik and M. Ito released the BRICS report (subsequently published in 2003):

Temporal logic with cyclic counting and the degree of aperiodicity of finite automata.

in which they enhance temporal logic by adding cyclic counting. In 2002, Zoltán published another BRICS report (published at DLT 2003):

Extended temporal logic on finite words and wreath product of monoids with distinguished generators

where he further developed his idea of enriching temporal logic, in the spirit of Wolper (1983).

An extension of Eilenberg's variety theorem

These papers provide an algebraic characterization of the expressive power of these logics. The novelty is that the corresponding classes of languages do not form a variety: they are are closed under inverses of renamings, but not under inverses of morphisms.

A similar idea was developed independently and at the same time by Straubing (2002). This gave rise to the theory of $\mathcal{C}$-varieties, which is an extension of Eilenberg's variety theory.

$\mathcal{C}$-morphisms

Let $\mathcal{C}$ be a class of morphisms closed under composition containing the renamings.

Examples of such classes $\mathcal{C}$:

- All morphisms
- Renamings ($\varphi(A) \subseteq B$)
- Length increasing ($\varphi(A) \subseteq B^+$)
- Length decreasing ($\varphi(A) \subseteq B \cup \{1\}$)
- Uniform ($\varphi(A) \subseteq B^k$ for some fixed k)

$\mathcal{C}$-varieties of languages

A class of languages $\mathcal{K}$ is closed under inverses of $\mathcal{C}$-morphisms if, for each $\mathcal{C}$-morphism $\varphi : A^* \to B^*$, the condition $L \in \mathcal{K}(B^*)$ implies $\varphi^{-1}(L) \in \mathcal{K}(A^*)$.

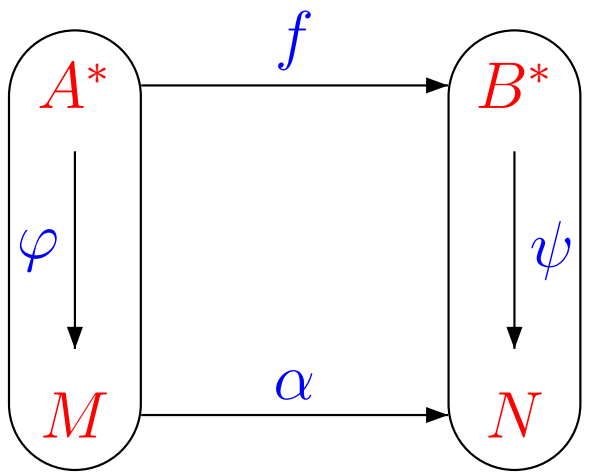
A $\mathcal{C}$-variety of languages is a class of regular languages closed under Boolean operations, quotients and inverses of $\mathcal{C}$-morphisms.

- Eilenberg's $+$-varieties can be seen as length-increasing varieties.
- The languages of generalized star-height $\leqslant n$ form a length-decreasing variety of languages.

The category of $\mathcal{C}$-stamps

Objects: Stamps $=$ Surjective morphisms from a free monoid onto a finite monoid.

$\mathcal{C}$-morphisms: A pair (f, α), where $f : A^* \to B^*$ is in $\mathcal{C}$, $\alpha : M \to N$ is a monoid morphism, and $\psi \circ f = \alpha \circ \varphi$.



Extended variety theorem

A $\mathcal{C}$-variety of stamps is a class of stamps closed under taking sub-objects, quotient objects and finite product in the category of $\mathcal{C}$-stamps.

Theorem (Ésik - Straubing)

$\mathcal{C}$-varieties of languages are in bijective correspondence with $\mathcal{C}$-varieties of stamps.

$\mathcal{C}$-identities

A class E of profinite equations is closed under $\mathcal{C}$-morphism if, for each $\mathcal{C}$-morphism $\varphi : A^* \to B^*$, $u \to v \in E(A)$ implies $\varphi(u) \to \varphi(v) \in E(B)$. Such equations are called $\mathcal{C}$-identities.

Theorem (Esik, Straubing + Kunc 2003)

A class of regular languages is $\mathcal{C}$-variety iff it can be defined by a set of profinite $\mathcal{C}$-identities.

Examples of length-multiplying identities

Length-multiplying identities: x and y represent words of the same length.

- $FO[< +MOD]$ = Regular languages in AC^0 [Barrington, Compton, Straubing, Thérien 92]. Identities $(x^{\omega-1}y)^\omega = (x^{\omega-1}y)^{\omega+1}$ [Straubing, Kunc].
- $\Sigma_1[< +MOD]$ = Finite union of languages of the form $(A^d)^* a_1 (A^d)^* a_2 (A^d)^* \cdots\ a_k (A^d)^*$, with $d > 0$: $1 \leqslant x^{\omega-1}y$ and $1 \leqslant yx^{\omega-1}$. [Chaubard, Pin, Straubing 06]

Part III

Back to the shuffle operation

Ordered monoids

An ordered monoid is a monoid equipped with a partial order compatible with the product.

A language L of A^* is recognised by an ordered monoid M if there exists a monoid morphism $f : A^* \to M$ and an upset P of M such that $L = f^{-1}(P)$.

There is a minimal ordered monoid recognising a language, called its syntactic ordered monoid.

Positive varieties

A positive variety of languages is a class of regular languages closed under union, intersection, quotients and inverses of morphisms.

A variety of ordered monoids is a class of finite ordered monoids closed under taking finite products, ordered submonoids and quotients.

Theorem (P. 1995)

There is a bijection between positive varieties of languages and varieties of ordered monoids.

Downset monoids

Let $(M, \leqslant)$ be an ordered monoid. A downset of M is a subset F of M such that if $x \in F$ and $y \leqslant x$ then $y \in F$. The product of two downsets X and Y is the downset

$$XY = \{z \in M \mid z \leqslant xy \text{ for some } x \in X \text{ and } y \in Y\}$$

This operation makes the set $\mathcal{P}^{\downarrow}(M)$ of nonempty downsets of M an ordered monoid, called the downset monoid of M. The order is inclusion.

Surjective renaming and downset monoids

Proposition

Let f be a surjective renaming. If L is recognised by an ordered monoid M, then $f(L)$ is recognised by $\mathcal{P}^{\downarrow}(M)$.

Let $\mathcal{V}$ be a positive variety of languages. Let $R\mathcal{V}(A^*)$ be the lattice generated by the languages of the form $f(L)$, where $f : B^* \to A^*$ is a surjective renaming and $L \in \mathcal{V}(B^*)$. Then $R\mathcal{V}$ is a positive variety of languages.

Applying surjective renaming to positive varieties

Given a variety of ordered monoids $\mathbf{V}$, let $\mathbf{P}^{\downarrow}\mathbf{V}$ be the variety of ordered monoids generated by the monoids of the form $\mathcal{P}^{\downarrow}(M)$, where $M \in \mathbf{V}$.

Let $\mathcal{V}$ be a positive variety of languages and let $\mathbf{V}$ be the corresponding variety of ordered monoids.

Proposition (Polák 02, Cano-Pin 04)

The variety of ordered monoids corresponding to $R\mathcal{V}$ *is* $\mathbf{P}^{\downarrow}\mathbf{V}$.

Positive varieties closed under surjective renaming

Let $\mathcal{V}$ be a positive variety of languages and let $\mathbf{V}$ be the corresponding variety of ordered monoids.

- $\mathcal{V}$ is closed under surjective renaming iff $\mathbf{V} = \mathbf{P}^{\downarrow}\mathbf{V}$
- For all $\mathbf{V}$, $\mathbf{P}^{\downarrow}(\mathbf{P}^{\downarrow}\mathbf{V}) = \mathbf{P}^{\downarrow}\mathbf{V}$. (Note that one may have $\mathbf{P}(\mathbf{PV}) \neq \mathbf{PV}$).
- Several infinite families of fixed points of $\mathbf{P}^{\downarrow}$ are known [Almeida, Cano, Klíma, Pin 2015].

Examples of fixed points of $\mathbf{P}^{\downarrow}$

$x^\omega =$ idempotent power of x

- $[\![xy = yx, 1 \leqslant x]\!]$: commutative, 1 at bottom
- $[\![x^\omega y = yx^\omega, x \leqslant x^2]\!]$,
- $[\![x^\omega y = yx^\omega, 1 \leqslant x]\!]$,
- $[\![x^\omega y^\omega = y^\omega x^\omega, 1 \leqslant x]\!]$
- $[\![x \leqslant u]\!]$, where u is any profinite word,
- $[\![xy \leqslant u]\!]$, where u is any profinite word,
- $[\![x^{\omega+1} y = yx^{\omega+1}, x \leqslant x^{\omega+1}]\!]$,
- $[\![x^{\omega+1} y z^{2^\omega} = z^{2^\omega} y x^{\omega+1}, x \leqslant x^{2^\omega}]\!]$.

Varieties closed under renaming

- There are many positive varieties closed under renaming (fixed points of $\mathbf{P}^{\downarrow}$).
- There is a unique maximal one: it is the maximal positive variety not containing $(ab)^*$ [Cano, Pin 04].
- What about the commutative ones?

Commutative positive varieties

ld = class of length-decreasing morphisms between free monoids: the image of each letter is a letter or the empty word.

Theorem (Almeida, Ésik, Pin 2017)

Every commutative positive ld-variety of languages is a positive variety of languages.

A curious arithmetic interpretation

Mentioned in [Cegielski, Grigorieff, Guessarian 2014] for finite subsets. Setting, for each subset L of $\mathbb{N}$ and each positive integer k,

$$L - 1 = \{n \in \mathbb{N} \mid n + 1 \in L\}$$
$$L \div k = \{n \in \mathbb{N} \mid kn \in L\}$$

Proposition (Substraction allows division)

Let $\mathcal{L}$ be a lattice of regular subsets of $\mathbb{N}$ such that if $L \in \mathcal{L}$, then $L - 1 \in \mathcal{L}$. Then for each positive integer k, $L \in \mathcal{L}$ implies $L \div k \in \mathcal{L}$.

Commutative positive varieties closed under shuffle

Proposition (Almeida, Ésik, Pin 2017)

Let $\mathcal{V}$ be a commutative positive variety of languages and let $\mathbf{V}$ be the corresponding variety of ordered monoids. Are equivalent:

(1) *$\mathcal{V}$ is closed under surjective renaming,*

(2) *$\mathcal{V}$ is closed under shuffle,*

(3) *$\mathcal{V}$ is closed under product over one-letter alphabets,*

(4) $\mathbf{V} = \mathbf{P}^{\downarrow}\mathbf{V}$.

Examples

For each set of natural numbers S, let

$$\mathbf{V}_S = [\![\, xy = yx, x \leqslant x^{n+1} \text{ for all } n \in S \,]\!].$$

and let $\langle S \rangle$ be the submonoid of $(\mathbb{N}, +)$ generated by S.

Proposition

(1) *The positive variety of languages associated to* $\mathbf{V}_S$ *is closed under shuffle.*

(2) $\mathbf{V}_S$ *satisfies* $x \leqslant x^{m+1}$ *iff* $m \in \langle S \rangle$.

(3) $V_S = V_T$ *iff* $\langle S \rangle = \langle T \rangle$.

An open problem: Intermixed languages

Intermixed languages = smallest class of languages containing the singletons and closed under Boolean operations, product and shuffle.

It strictly contains the variety of star-free languages. The language $(abab)^*$ is intermixed but $(aa)^*$ is not.

Intermixed languages do not form a variety of languages, but they form a ld-variety of languages.

Open Problem [Restivo ∼2000]. Give an algebraic characterization of intermixed languages. Is it a decidable class?

Conclusion

I am sure that Zoltán would have liked to further investigate this type of questions among the numerous topics he was interested in.

I deeply miss him, as a scientist and as a personal friend.