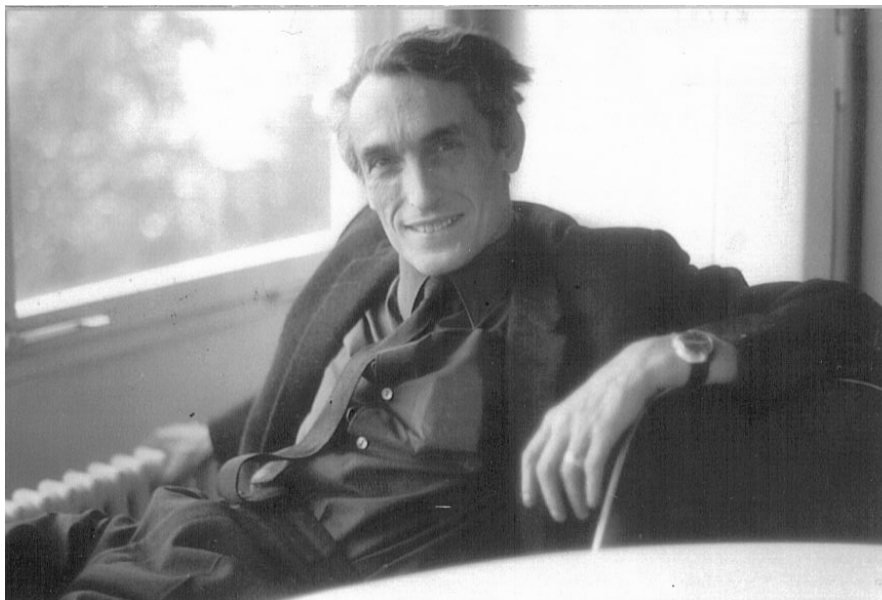

Conference dedicated to the scientific legacy of Marcel-Paul Schützenberger



Program, Abstracts and Information



Monday, March 21: Abstracts

Mireille Bousquet-Mélou, *Lattice walks and algebraic power series*

It is well-known that many algebraic power series occur when counting lattice walks. The archetypal case is that of Dyck paths: they can be seen as walks with steps $+1$ and -1 , starting and ending at 0 , and confined to the non-negative half-line. Of course, they are in bijection with words of the Dyck language, which is context-free.

More algebraic series occur when considering walks in two dimensions: for instance, walks with steps $(1, 1)$, $(0, -1)$, $(-1, 0)$, starting and ending at $(0, 0)$ and confined to the first quadrant. We shall discuss a number of such examples, compare them with the Dyck problem, and explain that some of them are more algebraic than others — even though all are counted by algebraic series.

We will also recast these problems in the systematic study of lattice walks confined to cones.

Volker Diekert, *Rational expressions and bounded synchronization delay*

In 1965 Schützenberger published his famous result that star-free languages and aperiodic languages coincide over finite words, often written as $SF = AP$. This result was extended to infinite words by Perrin in the mid 1980s. In 1973 Schützenberger presented another (and less known) characterization of aperiodic languages in terms of rational expressions where the use of the star operation is restricted to prefix codes with bounded synchronization delay and no complementation is used. The latter class has been denoted as SD . Hence, $SF = AP = SD$.

Kufleitner and the author extended $AP = SD$ to infinite words using what they call “local-divisor technique”. As a byproduct they simplified Schützenberger’s original proof for finite words substantially. In my talk I will present the local-divisor proof of $AP = SD$. As a matter of fact $SF = AP$ becomes an immediate corollary. Thus, even if one is interested in the connection between star-free and aperiodic languages, the new proof for $AP = SD$ might be of interest.

Sergey Fomin, *Noncommutative Schur functions*

Schur functions form the most important basis in the ring of symmetric functions, and the problem of expanding various families of symmetric functions with respect to this basis arises in many mathematical contexts. I will explain how several instances of this problem can be reformulated, and sometimes solved, using noncommutative analogues of Schur functions. This approach can be viewed as a generalization of Schützenberger’s proof of the Littlewood-Richardson Rule (using Schur functions in noncommutative variables satisfying the Knuth relations). This is joint work with Jonah Blasiak.

Laurence Danlos, *Natural Language Processing 60 years after the Chomsky-Schützenberger hierarchy*

Chomsky and Schützenberger (1956) laid the foundation of formal grammars and introduced Context Free Grammars (CFG). This groundbreaking paper led to formal linguistic research. However, if the syntax analysis of a computer program can be based only on a CFG (with procedures to construct meaning), analysis of a natural language text requires more complex formal systems.

In this survey lecture, I will present various models introduced in linguistics over the years, starting from Chomsky's early definitions of generative grammars, up to Tree Adjoining Grammars (Joshi 1985), which are "middy context-sensitive".

Middy context-sensitive grammars are appropriate for modeling natural languages, however they are "competence" grammars. They do not model "performance" and therefore are not robust enough to deal with the numerous "deviant" phenomena that are found in "real" texts, e. g. journalistic texts. Currently, the scientific community in Natural Language Processing (NLP) focuses on real text analysis with methods that I will sketch.

Eliahu Shamir, *Parental view of the context-free birth and evolution*

The study of Formal Languages arose as a confluence of linguistics models (Chomsky's hierarchy), formulations and compilation of programming languages (automata and machine models), Combinatorics of strings and Marcel-Paul's influential contribution: algebraic approach, via formal power series.

During the years 1957-1963 a lively Boston-Jerusalem correspondence produced seminal articles and monographs, focusing on context free (CF) as a central model, due to merge of several distinct characterizations, rich mathematics and relevance of applications.

The sixties saw a spread of research to persons, centers, articles, textbooks adopting and promoting the set-pattern of formal language study up to firm inclusion in basic CS education. After Eilenberg's books, monoid structure, semi-rings, weighted Automata and further algebraic approaches gained popularity.

Hindsight and impact of several early results: representation theorems, hardness, complexity, algorithms, undecidability adequacy of CF sub-models and super-models, comparing human and machine capacity in grammar learning and parsing, tree banks and repositories.

A recent twist: a grammar-transformation scheme applied to sub and super CF models which have CF skeleton, manages to spread and cyclic-rotate the trees and derived strings. By this transformation "Derivation trees of G" are mapped onto "UNION of derivation tree of several LINEAR grammars", thus enhancing parsing, property tests, and adequacy of applications.

Tuesday, March 22: Abstracts

Mikołaj Bojańczyk, *Recognisable languages over monads*

After Schützenberger's seminal work, the algebraic approach to regular languages has been extended to other structures, such as infinite words or trees. I would like to present a unifying approach to all these algebras, which is to use monads and their Eilenberg-Moore algebras. The idea is that semigroups are Eilenberg-Moore algebras for one monad, ω -semigroups are Eilenberg-Moore algebras for another monad, forest algebras are Eilenberg-Moore algebras for yet another monad, and so on. One can even prove general versions of some theorems, like the existence of syntactic algebras, which are true in many monads at once.

Thomas Place, *Schützenberger's Star Free theorem and what followed*

Schützenberger's Theorem is a seminal result in formal language theory and automata. It gave a deep understanding of a very important class of word languages: the star-free languages. This class is the smallest one which contains all finite languages and is closed under union, complementation and concatenation of languages. It is very prominent in the literature as it may also be defined as the class of languages that may be defined with first-order logic (FO) sentences or equivalently linear temporal logic (LTL) sentences. What Schützenberger proved is that one can decide whether an input regular language is star-free and if so build a star-free description of this language using a generic method.

Since this result, researchers have sought to extend and adapt Schützenberger's techniques to obtain similar results for other important classes of languages. A prominent example is the so-called dot-depth hierarchy which classifies star-free languages according to the (required) complexity of their definition. More precisely, the classification is made by counting the number of alternations between concatenation and complementation required to define a star-free language (from a logical point of view this amounts to counting quantifier alternations in first-order formulas). Obtaining results similar to Schützenberger's Theorem for each level on the hierarchy is a problem whose investigation has spanned decades and is still ongoing: only the very first levels are solved.

In this talk, I will introduce this problem and explain why it is both interesting and difficult. I will also present an overview of the known results and what is involved in obtaining them.

Eric Goles, *Dynamics of threshold automata over undirected graphs*

I present a characterization of the dynamical behavior of threshold automata over undirected graphs for arbitrary deterministic updates (parallel, sequential or a combination of both modes) by introducing a parameter taking into account the odd circuits of the graph. From that, it is proven that the network's dynamics admit only fixed point if and only if this parameter is always negative. Else, cycles may appear, and, for some specific updating modes, the periods could be super-polynomial. Finally I present examples as well as sufficient conditions over families of graphs in order to have only fixed point.

Christian Krattenthaler, *Robinson-Schensted-Knuth correspondence, jeu de taquin, and growth diagrams*

One set of important results obtained by Marcel-Paul Schützenberger concerns the correspondence between permutations and standard tableaux of Robinson and Schensted, and its extension to semi-standard tableaux due to Knuth, as well as the closely related "tableau game" that he invented and called "jeu de taquin". The "modern" way to deal with these correspondences is through Fomin's growth diagrams. I shall review all these ideas and then put them into action by establishing a recent conjecture of Sophie Burrill on a certain bijective relation between standard tableaux and oscillating tableaux.

Valentin Féray, *Schützenberger's jeu de taquin*

The jeu de taquin is a simple combinatorial operation on integer tableaux introduced and studied by Marcel-Paul Schützenberger. When iterating this procedure, some surprising invariance properties appear, in particular in connection with Robinson-Schensted algorithm. In a first part of the talk, I will present (some of) these properties.

In a second part of the talk, I will mention some recent results involving jeu de taquin in enumerative, algebraic and probabilistic combinatorics.

Gilles Schaeffer, *Algebraic decomposition of maps*

I will use the excuse of a recent combinatorial result on planar maps, the enumeration of corner polyhedra graphs, to illustrate how "Schützenberger's methodology" underlies the path from a collection of elegant formulas of W.T. Tutte from the 60's to recent algorithmic and probabilistic developments around these objects.

Wednesday, March 23: Abstracts

Arturo Carpi, *Incompleteness and synchronization in finite automata and codes.*

The notions of completeness and synchronization play a central role in Computer Science since they appear in the study of several problems on variable length codes and on finite automata. According to a well-known result of Schützenberger, the property of completeness provides an algebraic characterization of finite maximal codes, which are the objects used in Information Theory to construct optimal sequential codings. The property of synchronization plays a natural role for the formal description of fault-tolerant systems. In particular, in the context of Information Theory, this property is relevant for the construction of decoders that are able to efficiently cope with decoding errors caused by noise during the data transmission.

Problems that naturally arise in these contexts concern the minimal length of an incomplete word in an incomplete set of words as well as the minimal length of a synchronizing word for a synchronizing automaton or for a synchronizing code.

In this framework, a prominent place is taken by the fifty-year old Černý conjecture which hypothesizes that any deterministic n -state automaton with a reset word has a reset word of length $(n - 1)^2$.

We will survey some results concerning these two problems as well as some recent developments concerning finite sets of words.

Thomas Colcombet, *Regular cost functions*

Regular cost functions form a quantitative extension to regular languages, in which, instead of languages, functions from words to $N \cup \{\infty\}$ are considered. The specificity of cost functions is that functions are considered modulo an "up to boundedness" equivalence relation, that preserves the existence of bounds over all sets of inputs, while forgetting the specific values taken. The *Cost function* are equivalence classes for this relation.

The theory can then be developed in a way very reminiscent to regular languages. In particular, many formalisms turn out to be effectively equivalent for describing cost functions: algebraic (stabilisation monoids), automata theoretic (B- and S-automata), expression based (B- and S-regular expressions) and logical (cost monadic-second order logic). Such cost functions are named *regular*. Decision procedures testing the equality of regular cost functions or the existence of bounds, are also available. All these results strictly generalize the ones for regular languages.

The resulting theory offers a toolbox of representation and decidability results that deals with combinations of regular behaviours and boundedness considerations. It has been developed over words, infinite words, trees, and (partially) infinite trees, and has been successfully used for solving problems in automata theory, databases, logic, or verification.

The theory of regular cost functions takes its root in several other works, such as Hashiguchi's work on the star-height problem and distance automata, Leung's topological proof for it, Simon's work on matrices over the tropical semiring and the finite power property, Kirsten's distance desert automata approach, or Bojanczyk and Colcombet's study of MSO+U. It is also related to Schützenberger's weighted automata, and characterizations 'à la Schützenberger' of subclasses of regular cost functions have also been established.

Claudia Malvenuto, *Bijections and Surjections*

During the last 20 years, it was realized that certain combinatorial objects (combinatorial Hopf algebras, to be precise) underly many mathematical theories. This phenomenon occurs in almost all fields of mathematics, from numerical analysis to theoretical physics or number theory including renormalization. The talk will survey some of these developments, and then focus largely on two of the most emblematic and universal such objects, namely the higher algebraic structures that can be constructed out of permutations, and out of surjections.

The link is made through the notion of special poset (equivalent to labelled poset of Stanley): linear extensions of a poset can be seen as bijections, while the generating function of a poset P with respect to Stanley's classical definition of P -partitions associated to a special poset is a quasi-symmetric: in fact, it is a homomorphism between the Hopf algebra of labelled posets and that of quasi-symmetric functions; while linearisation is a homomorphism onto the Hopf algebra of permutations.

The aim is to generalize this frame to preorders, which are in one-to-one correspondence with finite topologies: the objects corresponding to bijections are surjections: they can be seen as linear extensions of a preorder and are encoded by packed words. We can hence define the notion of T -partitions associated to a finite topology T , and deduce a Hopf algebra morphism from a new Hopf algebra on topologies to the Hopf algebra of packed words.

This is joint work with L. Foissy and F. Patras.

Denis Thérien, *Complexity-theoretic stories around two theorems of Schützenberger*

MP Schützenberger has given us two of the most beautiful and deepest theorems in the theory of rational languages, providing algebraic characterizations for the combinatorial operations of concatenation and unambiguous concatenation respectively. When the mechanism of recognition-by-homomorphism is generalized to program-over-monoid, these theorems give insights into questions that belong to computational complexity, in particular in the theory of shallow boolean circuits.

Howard Straubing, *Expansions, Decompositions, and Two-variable Logic*

We will discuss (without too much technical detail) the block product of semigroups, a decomposition that is closely related to the Schützenberger product, an expansion (in Rhodes's terminology) of finite semigroups. This construction is in turn rooted in early work (1961) of Schützenberger on bilateral transducers. Seen in this light the famous theorem on the equivalence of star-free languages and aperiodic monoids (1965) is a kind of two-sided counterpart of the Krohn-Rhodes Theorem, discovered independently, and at roughly the same time.

If we turn these block product decompositions inside-out (associate them left-to-right instead of right-to-left) we fall on another great discovery of Schützenberger, –and the main point of our talk– the variety of DA monoids (1976). This has a remarkable consequence in logic –it exactly separates properties of words definable by first-order formulas with two variables from those definable with more variables.

(References: The bilateral transducers, later termed 'sequential bimachines' by Eilenberg, are introduced in Schützenberger, 'A remark on finite transducers' (1961). The Schützenberger product and the equivalence of star-free languages and aperiodic monoids is from Schützenberger, 'On finite monoids having only trivial subgroups' (1965). To quote again from Eilenberg, 'this is next to Kleene's Theorem, probably the most important result dealing with recognizable sets'. The

variety DA is introduced in Schützenberger, ‘Sur le produit de concaténation non-ambigu’ (1976). The connection with two-variable logic is from Thérien and Wilke, ‘Over words, two variables are as powerful as one quantifier alternation’ (1998), and the approach through the inside-out iterated block product in Straubing and Thérien, ‘Weakly iterated block products of finite monoid.’ (2002.)

Olivier Carton, *Algebraic characterization of logics on countable linear orderings*

In this talk, I will discuss some notion of recognizability for sets of words indexed by countable linear orderings. This topic can be seen as a natural continuation of a long line of research that aims to generalize finite words to infinite objects. I will present an algebraic model for recognizing languages of countable words and show that this notion of recognizability is effectively equivalent to definability in MSO. Then I will present some characterizations by algebraic means of some fragments of MSO like FO.

Donald Knuth, *Reminiscences of Marco*

In this talk I will speak of some ”souvenirs” I have of discussions with Marco about our common interests in formal languages and combinatorial algorithms.

Thursday, March 24: Abstracts

Véronique Bruyère, *On algorithmic aspects of infinite games played on weighted graphs*

Reactive computer systems (like power plants, plane navigation systems, ABS systems, ...) bear inherent complexity due to continuous interactions with the environment in which they evolve. Formal techniques using mathematical models have been advocated to help to their systematic design. One well-studied formal technique is model checking which automatically checks whether a model of a system meets a given specification. A more challenging goal is the automatic synthesis of a correct system from the given specification, that is, a system that enforces the specification no matter how the environment behaves. The mathematical framework used for the synthesis problem is game theory and in particular infinite two-player games played on graphs where the two players are the system and the environment, the vertices of the graph model their possible states, and the edges model their possible interactions. Synthesizing a correct system then means constructing a winning strategy of the system such that, whatever the strategy of the environment, the outcome of the play satisfies the specification. The fundamental algorithmic questions are therefore: Can we decide whether a player has a winning strategy? Can we construct such a strategy, and if possible a simple one? What are the complexities of these problems? Answers to these questions have been obtained for many kinds of qualitative specifications like reachability objectives, Büchi objectives, ... In this talk (in the form of a survey), we focus on recent developments about games played on weighted graphs with quantitative objectives, in a way to treat specifications like never dropping out of fuel, ensuring a suitable mean response time, ... We begin with the simple case of unidimensional weighted graphs, and then continue with the more ambitious multidimensional case to deal with situations where the reactive system has to satisfy several quantitative objectives instead of one. We also consider the extension to multiplayer games where several (instead of two) players have their own objectives that are not necessarily antagonistic.

Stefano Crespi Reghizzi, *An enduring trail of language characterizations via homomorphisms and local languages*

The Chomsky-Schützenberger theorem opened a third way between the existing two methods of defining languages, namely generation by grammar and recognition by automaton. The corresponding view has permeated many later studies on context-free and other formal language families, and is also remindful of the surface versus deep structure distinction in linguistics. Many would agree that the whole algebraic theory of languages, in particular the Abstract Family of Languages movement, originated from the theorem, but here we prefer to focus on some closely related developments to be classified as: refinements of the theorem, extensions to language families that include the context-free, and analogous proposals to characterize other families (e.g., the recursively enumerable) by intersection and homomorphism. Among refinements, we mention the restatements that avoid cancellation of Dyck parentheses and that use a grammar-independent alphabet of parentheses. Among extensions, an example is the case of automata that use more than one push-down stack. Since its publication and until now, a steady flow of articles referring to the theorem have appeared, which altogether constitute a sort of cross section of the formal language theoretical research space. On the other hand, algorithmic and practical applications to language processing have been rare, and the unperishable value of the theorem is more as an aid for thinking about language structures rather than as a designing method. To this communication has participated Pierluigi San Pietro.

Alain Connes, *Le monde de la caractéristique 1*

G rard Berry, *Automata, circuits, electricity, intuitionistic logic, and determinism*

Every student exposed to finite automata learns that they exist in two forms, non-deterministic and deterministic, and that one of the first actions to perform is to transform non-deterministic automata into deterministic ones, maybe at the price of an exponential size explosion. The talk will first question this common belief, using ideas from digital circuits. In circuit design, there is a trivial way to encode deterministic automata into gates : performing one-hot encoding, i.e., using one register fed by an or-gate for each state, implementing a transition with an and-gate fed by the source state register output and the wire for the letter of the transition, and connecting this and-gate to the or-gate that feeds the target state register. But this encoding is linear the size of the deterministic automaton and thus becomes rapidly unpractical. The alternative encoding of n -states automatas with $\log_2(n)$ registers does not prevent size explosion in the next-state logic because the next-state function may use $\exp(\log_2(n))=n$ gates. But a much better encoding follows from a simple but largely unknown remark : the trivial encoding performed in the one-hot case works perfectly for non-deterministic automata as well, yielding a deterministic, linear-size, and efficient circuit that completely avoids determinization-related size explosion. The intrinsically synchronous and deterministic concurrency of electrical propagation actually performs the classical subset construction on-the-fly and at lightspeed. Thus, the word "non-determinism" that dates back to the times where the only way to compute was sequentially with a single pencil might have led many people to a wrong track at least for circuits. Electrical non-deterministic automata are indeed deterministic by construction, and the generated circuits can also be trivially implemented in software.

Once this idea is understood, the second part of the talk shows how to add more powerful programming primitives to complement the sequence, choice and loop. For instance, concurrency and preemption available in the Esterel language have two simultaneous qualities : providing the programmer with much more compact elegant source code, and yielding circuits that turn out to be smaller and faster than those designed by hand once adequately optimized, at least for the control-intensive applications. We finally study what happens if the generated circuit may contain combinational loops, which is natural for Esterel and other high-level languages. While they are rejected by most synthesis tools, such cyclic circuits may be more natural and much smaller than acyclic ones for the same function as shown in the late 80's by Malik. But not all cyclic circuits are deterministic or even make sense ($x=x$ or $x=\text{not } x$ for instance). We state a beautiful result that characterizes the good circuits by relating logic and physics : a cyclic circuit is electrically deterministic for all delays of gates and wires if and only if the solution of its equations can be computed using only intuitionistic Boolean logic, i.e.. not using excluded middle. The proof by Mendler (2012) consists in relating intuitionistic Boolean logic with a continuous temporal logic of wire voltage stabilization.

Friday, March 25: Abstracts

Jim Haglund, *Combinatorial problems connected to Macdonald Polynomials*

We begin by discussing the combinatorial formula for Macdonald polynomials due to Haiman, Loehr, and the speaker, and show how it implies a famous result of Lascoux and Schutzenberger on the Schur expansion of Hall-Littlewood polynomials. We also overview some more recent results and conjectures involving Macdonald polynomials and combinatorial expressions for various characters.

Cedric Join, *The model-free control: a mathematical tool for industrial applications*

Applied academic publications are quite often based on an intricate mathematical modelling. Differential equations and probability distributions, which are so efficient in pure physics, remain most popular, although it is difficult, if not impossible, to write them down in most concrete situations. The situation looks quite different however in the industrial world, where such a modelling is avoided most of the time. In order to analyse as precisely as possible this dichotomy, this talk will almost exclusively be devoted to control engineering, which is sometimes called *informatique industrielle* in France.

Control theory is being taught and developed mainly in the departments of applied mathematics and of electrical and computer engineering. Its aim is to regulate a plant, or more generally any type of “system,” via actuators thanks to measures obtained by sensors. A “good” control law permits to reach “good” performances.

Model-based control, which is mainly investigated in the academic world, has almost never achieved a preeminent position for real applications. The following points are sketching an explanation:

1. Friction, heat and ageing effects, . . . are impossible to quantify accurately.
2. Mass production yields a dispersion of the item characteristics, which might be severe.

This is why PID (Proportional-Integral-Derivative) controllers, which were introduced more than seventy years ago and which do not necessitate the knowledge of any mathematical modelling, are still the bread and butter of industrial control. They need however a fine tuning, which is too often very hard to achieve. The resulting system behaviour might nevertheless be quite poor.

This talk will present model-free control, i.e., an approach which is utilizing new mathematical tools, and which, contrarily to PIDs, is easy to implement. It will be illustrated by several successful applications in most diverse domains, ranging from hydroelectric power plants to intelligent transportation systems.

Jacques Demongeot, *On conjectures of M. P. Schützenberger about genetic regulation : the coding and its expression*

The French school of theoretical biology has been mainly initiated in Poitiers during the sixties by scientists like J. Besson, G. Bouligand, P. Gavaudan, M.P. Schützenberger and R. Thom, launching many new research domains on the fractal dimensions, combinatorial properties of the genetic code and related amino-acids, as well as on the genetic regulation of the biological processes. Presently, the biological science knows that RNA molecules are often involved in the regulation of complex genetic networks as effectors, e.g., activators (small RNAs as transcription factors), inhibitors (micro-RNAs) or hybrids (circular RNAs). M.P. Schützenberger and his colleagues already believed that RNA molecules like RNA codons or small circular RNAs played a central role in the birth of the genetic code and in the regulation of the genetic networks, in particular they have conjectured 3 main open questions:

- Conjecture 1: is the repartition of the 64 codons inside the 21 synonymy classes of the genetic code verifying an optimization principle?
- Conjecture 2: if yes, are there some physico-chemical rules giving birth to a primitive non-degenerate code?
- Conjecture 3: if yes, has this primitive non-degenerate code leaved RNA relics in the present genomes and which role have these relics in the regulation of complex genetic networks ?

After solving the two first conjectures, some examples of such relics and regulated networks will be given showing that i) there exist indeed RNA "relics" that have played an important role during evolution and survive in many genomes, whose probability distributions of their sub-sequences are quantified by their Shannon entropy, and ii) the robustness of the networks they regulate is characterized by their Kolmogorov-Sinai dynamic entropy.

Matthieu Josuat-Vergès, *Eulerian and André polynomials: some recent developments*

In the early 70's, Dominique Foata and Marcel-Paul Schützenberger published two monographs that form an important milestone in enumerative combinatorics: "Théorie géométrique des polynômes eulériens" and "Nombres d'Euler et permutations alternantes". They extended pioneering work by Désiré André on alternating permutations, and named few objects after him: André permutations, André trees, André polynomials, André complexes. After some historical review, I will go into into more recent topics and present a multivariate hook length formula for André trees.

Sylvain Lombardy, *On formal power series realized by two-way automata*

Marcel-Paul Schützenberger introduced weighted automata in 1961. This model has proven to be quite sound: the class of formal power series realized by weighted automata is the class of rational series – obtained from polynomial by sum, Cauchy product and Kleene star –, which is also the class of recognizable series realized by linear representations.

We present in this talk the series realized by two-way weighted automata. Opposite to the case of finite automata, two-way weighted automata are strictly more powerful than one-way ones. They are closely related to the class HadRat of series obtained from rational series by sum, Hadamard product and Hadamard iteration, and to series realized by two-way representations.

We show that, when weights belong to “nice” semirings, two-way automata and two-way representations realize the same class of series, called two-way recognizable series. These series can also be characterized as solutions of some fixed-point equations. By contrast, the class HadRat is smaller; we show that this class is realized by a restricted class of two-way weighted automata: the rotating automata. We also characterize series realized by a larger subclass, the sweeping automata.