Concatenation Hierarchies: New Bottle, Old Wine

Thomas Place(✉) and Marc Zeitoun(✉)

LaBRI, Bordeaux University, Bordeaux, France
{tplace,mz}@labri.fr

Abstract. We survey progress made in the understanding of concatenation hierarchies of regular languages during the last decades. This paper is an extended abstract meant to serve as a precursor of a forthcoming long version.

1 Historical Background and Motivations

Our objective in this extended abstract is to outline progress obtained during the last 50 years about *concatenation hierarchies of regular languages* over a fixed, finite alphabet A. Such hierarchies were considered in order to understand the interplay between two basic constructs used to build regular languages: Boolean operations and concatenation. The story started with Kleene's theorem [12], one of the core results in automata theory. It states that languages of finite words recognized by finite automata are exactly the ones that can be described by regular expressions, *i.e.*, are built from the singleton languages and the empty set using a finite number of times operations among three basic ones: union, concatenation, and iteration (also known as Kleene star).

As Kleene's theorem provides another syntax for regular languages, it makes it possible to classify them according to the hardness of describing a language by such an expression. The notion of star-height was designed for this purpose. The *star-height* of a regular expression is its maximum number of nested Kleene stars. The *star-height* of a regular language is the minimum among the star-heights of all regular expressions that define the language. Since there are languages of arbitrary star-height [7,8], this makes the notion an appropriate complexity measure, and justifies the question of computing the star-height of a regular language (it was raised by Eggan [8], see also Brzozowski [4]).

Given a regular language and a natural number n, is there an expression of star-height n defining the language?

This question, called the *star-height problem*, is an instance of a *membership problem*. Given a class $\mathcal{C}$ of regular languages, the membership problem for $\mathcal{C}$ simply asks whether $\mathcal{C}$ is a decidable class, that is:

INPUT:	A regular language L
OUTPUT:	Does L belong to $\mathcal{C}$?

Funded by the DeLTA project (ANR-16-CE40-0007).

P. Weil (Ed.): CSR 2017, LNCS 10304, pp. 25–37, 2017.
DOI: 10.1007/978-3-319-58747-9_5

Thus, the star-height problem asks whether membership is decidable for each class $\mathcal{H}_n$, where $\mathcal{H}_n$ is the class of languages having star-height n. It was first solved by Hashiguchi [10], but it took several more years to obtain simpler proofs, see *e.g.*, [3,11].

Kleene's theorem also implies that adding *complement* to our set of basic operations does not make it possible to define more languages. Therefore, instead of just considering regular expressions, one may consider *generalized* regular expressions, where complement is allowed (in addition to union, concatenation and Kleene star). This yields the notion of *generalized star-height*, which is defined as the star-height, but replacing "regular expression" by "generalized regular expression". One may then ask the very same question: is there an algorithm to compute the *generalized star-height* of a regular language? Despite its very simple statement, this question, also raised by Brzozowski [4], is still open. Even more, one does not know whether there exists a regular language of generalized star-height greater than 1. In other words, membership is open for the class of languages of generalized star-height 1 (see [18] for a historical presentation).

This makes it relevant to already focus on languages of star height 0, *i.e.*, that can be described using only union, concatenation and Boolean operations (including complement), but *without* the Kleene star. Such languages are called *star-free*. Surprisingly, even this restricted problem turned out to be difficult. It was solved by Schützenberger [30] in a seminal paper.

Theorem 1 (Schützenberger [30]). *Membership is decidable for the class of star-free languages.*

The star-free languages rose to prominence due to their numerous characterizations, and in particular, the logical one, which is due to McNaughton and Papert [15]. Observe that one may describe languages with logical sentences. Indeed, any word may be viewed as a logical structure made of a linearly ordered sequence of positions, each one carrying a label. In first-order logic over words (denoted by FO($<$)), one may quantify these positions, compare them with a predicate "$<$" interpreted as the (strict) linear order, and check their labels (for any letter a, a unary predicate P_a selecting positions with label a is available). Each FO($<$) sentence states a property over words and defines the language of all words that satisfy it.

Theorem 2 (McNaughton-Papert [15]). *For any regular language L, the following properties are equivalent:*

- *L is star-free.*
- *L can be defined by an FO($<$) sentence.*

Let us point out that this connection is rather intuitive. Indeed, there is a clear correspondence between union, intersection and complement for star-free languages and the Boolean connectives in FO($<$) sentences. Moreover, concatenation corresponds to existential quantification.

1.1 The Dot-Depth and the Straubing-Thérien hierarchies

Just as the star-height measures how complex a regular language is, a natural complexity for star-free languages is the number of alternations between the concatenation product and the complement operation that are required to build a given language from basic star-free languages. This led Brzozowski and Cohen [5] to introduce in the 70s a hierarchy of classes of regular languages, called the *dot-depth hierarchy*. This hierarchy classifies all star-free languages into full levels, indexed by natural numbers: 0, 1, 2,..., and half-levels, indexed by half natural numbers: $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, etc. Roughly speaking, level $n \in \mathbb{N}$ consists of all languages that can be expressed by a star-free expression having n alternations between concatenation and Boolean operations.

More formally, the hierarchy is built by using, alternately, two closure operations starting from level 0: *Boolean closure* and *polynomial closure*. Given a class of languages $\mathcal{C}$, its *Boolean closure* (denoted by $Bool(\mathcal{C})$) is the smallest Boolean algebra containing $\mathcal{C}$. Polynomial closure is slightly more complicated as it involves *marked concatenation*. Given two languages L_1 and L_2, a marked concatenation of L_1 with L_2 is a language of the form,

$$L_1 a L_2 \qquad \text{for some}\, a \in A.$$

We may now define the polynomial closure of $\mathcal{C}$ (denoted by $Pol(\mathcal{C})$) as the smallest class of languages containing $\mathcal{C}$ and closed under union, intersection and marked concatenation (*i.e.*, $L_1 a L_2 \in \mathcal{C}$ for any $L_1, L_2 \in \mathcal{C}$ and any $a \in A$).

The dot-depth hierarchy is now defined as follows:

- Level 0 is the class $\{\emptyset, \{\varepsilon\}, A^+, A^*\}$ (where A is the working alphabet).
- Each *half-level* is the *polynomial closure* of the previous full level: for any natural number $n \in \mathbb{N}$, level $n + \frac{1}{2}$ is the polynomial closure of level n.
- Each *full level* is the *Boolean closure* of the previous half-level: for any natural number $n \in \mathbb{N}$, level $n + 1$ is the Boolean closure of level $n + \frac{1}{2}$.

A side remark is that the above definitions are not the original ones. First, the historical definition of the dot-depth hierarchy started from another class of languages for level 0. However, both definitions coincide at level 1 and above. Next, the polynomial closure of a class $\mathcal{C}$ was historically defined as the smallest class containing $\mathcal{C}$ and closed under both union and marked concatenation. This original definition is intuitively weaker: it does not explicitly require $Pol(\mathcal{C})$ to be closed under intersection. However, it was shown by Arfi [1,2] that the two definitions are equivalent (provided that the class $\mathcal{C}$ satisfy some standard closure properties, which are always fulfilled for classes within concatenation hierarchies). This was also shown later by Pin [17]. We will present an alternative, elementary proof in the full version of this paper.

Clearly, the union of all levels in the dot-depth hierarchy is the whole class of star-free languages. Moreover, it was shown by Brzozowski and Knast that the dot-depth hierarchy is strict: any level contains strictly more languages than the previous one.

Theorem 3 (Brzozowski and Knast [6]). *The dot-depth hierarchy is strict when the alphabet contains at least two letters.*

This shows in particular that in general, Boolean closure does not preserve the property of being polynomially closed, and conversely. In other words, classes built using Boolean and polynomial closure do not satisfy the same closure properties: typically, when $\mathcal{C}$ is a class of languages, $Pol(\mathcal{C})$ is closed under marked concatenation but **not** under complement, while $Bool(\mathcal{C})$ is closed under complement but **not** under marked concatenation.

The fact that the hierarchy is strict motivates the investigation of the membership problem for all levels.

Problem 4 (Membership for the dot-depth hierarchy). For a fixed level in the dot-depth hierarchy, is the membership problem decidable for this level?

Using the framework developed by Schützenberger in his proof for deciding whether a language is star-free, Knast [13] established that level 1 has decidable membership, via a quite intricate proof from the combinatorial point of view.

Theorem 5 (Knast [13]). *Level 1 in the dot-depth hierarchy has decidable membership.*

The case of half levels required to adapt Schützenberger's methodology, since it was designed to deal with Boolean algebras only (recall that half-levels are *not* Boolean algebras, otherwise the hierarchy would collapse). This was achieved by Pin and Weil [21–23] and by Glaßer and Schmitz [9].

Theorem 6 (Pin and Weil [21–23], Glaßer and Schmitz [9]). *Levels $\frac{1}{2}$ and $\frac{3}{2}$ in the dot-depth hierarchy have decidable membership.*

One may now wonder why, in the definition of the dot-depth hierarchy, level 0 is $\{\emptyset, \{\varepsilon\}, A^+, A^*\}$. It would be natural to start from $\{\emptyset, A^*\}$, and to apply the very same construction for higher levels. This is exactly how the Straubing-Thérien hierarchy is defined. It was introduced independently by Straubing [33] and Thérien [35]. Its definition follows the same scheme as that of the dot-depth hierarchy, except that level 0 is $\{\emptyset, A^*\}$.

Like the dot-depth hierarchy, the Straubing-Thérien hierarchy is strict and spans the whole class of star-free languages. This can be shown by proving that level n in the dot-depth hierarchy sits between levels n and $n+1$ in the Straubing-Thérien hierarchy. This makes the membership problem again relevant for each level in this hierarchy.

Problem 7 (Membership for the Straubing-Thérien hierarchy). For a fixed level in the Straubing-Thérien hierarchy, is the membership problem decidable for this level?

Just as for the dot-depth hierarchy, level 1 in the Straubing-Thérien hierarchy was shown to be decidable (actually before the formal definition of the hierarchy itself), and the first half-levels were solved using the adaptation of the framework of Schützenberger to classes that are not closed under complement.

Theorem 8 (Simon [31,32]). *Level 1 in the Straubing-Thérien hierarchy has decidable membership.*

Theorem 9 (Arfi [1,2], Pin and Weil [21,22]). *Levels $\frac{1}{2}$ and $\frac{3}{2}$ in the Straubing-Thérien hierarchy have decidable membership.*

Both hierarchies are strongly related. First, as we already stated, they are interleaved. More importantly, Straubing established an effective reduction between the membership problems associated to their levels [34].

Theorem 10 (Straubing [34]). *Membership for level $n \in \mathbb{N}$ in the dot-depth hierarchy reduces to membership for level n in the Straubing-Thérien hierarchy.*

This theorem is crucial. Indeed, from a combinatorial view, membership is simpler to deal with for the Straubing-Thérien hierarchy rather than for the dot-depth. This is evidenced by all recent publications on the topic: most results for the dot-depth are indirect. They are corollaries of direct results for the Straubing-Thérien hierarchy via the above theorem. Thus, while the name "dot-depth" remains widely used, the Straubing-Thérien hierarchy is much more prominent.

1.2 Quantifier Alternation Hierarchies

Since star-free languages are exactly those that one can define in first-order logic, it is desirable to refine this correspondence level by level, in each of the hierarchies considered so far. A beautiful result of Thomas [36] establishes indeed such a correspondence, and it is very natural. To present it, we first need to slightly extend the standard signature used in first-order logic over words: we add four new predicates in addition to "$<$" and the unary predicates P_a for $a \in A$:

- The (binary) *successor*, interpreted as the successor between positions.
- The (unary) *minimum*, that selects the leftmost position of the word.
- The (unary) *maximum*, that selects the rightmost position of the word.
- The (nullary) *empty* predicate, which holds for the empty word only.

We denote by $\text{FO}(<, +1, \min, \max, \varepsilon)$ the resulting logic. Notice that these predicates are all definable in $\text{FO}(<)$. Therefore, adding them in the signature does not add to the overall expressive power of first-order logic. In other words, $\text{FO}(<)$ and $\text{FO}(<, +1, \min, \max, \varepsilon)$ are equally expressive. However, this enriched signature makes it possible to define fragments of first-order logic corresponding to levels of the dot-depth hierarchy.

To this end, we classify $\text{FO}(<, +1, \min, \max, \varepsilon)$ sentences by counting their number of quantifier alternations. Given a natural number $n \in \mathbb{N}$, a sentence is said to be "$\Sigma_n(<, +1, \min, \max, \varepsilon)$" (resp. "$\Pi_n(<, +1, \min, \max, \varepsilon)$") when it is an $\text{FO}(<, +1, \min, \max, \varepsilon)$-formula whose prenex normal form has either:

- *Exactly* n blocks of quantifiers, the leftmost being an "$\exists$" (resp. a "$\forall$") block, or
- *Strictly less* than n blocks of quantifiers.

For example, a formula over the signature $(<, +1, \min, \max, \varepsilon, (P_a)_{a \in A})$ whose prenex normal form is

$$\exists x_1 \exists x_2 \, \forall x_3 \, \exists x_4 \; \varphi(x_1, x_2, x_3, x_4) \quad (\varphi \text{ quantifier-free})$$

is $\Sigma_3(<, +1, \min, \max, \varepsilon)$. Observe that while FO($<$) and FO($<, +1, \min, \max, \varepsilon$) have the same expressiveness, the enriched signature increases the expressive power of each individual level.

Note also that the negation of a $\Sigma_n(<, +1, \min, \max, \varepsilon)$ sentence is not a $\Sigma_n(<, +1, \min, \max, \varepsilon)$ sentence in general (it is a $\Pi_n(<, +1, \min, \max, \varepsilon)$ sentence), and the corresponding classes of languages are not closed under complement. It is therefore meaningful to define $\mathcal{B}\Sigma_n(<, +1, \min, \max, \varepsilon)$ sentences as Boolean combinations of $\Sigma_n(<, +1, \min, \max, \varepsilon)$ and $\Pi_n(<, +1, \min, \max, \varepsilon)$ sentences. This gives a strict hierarchy of classes of languages depicted in Fig. 1, where, slightly abusing notation, each level denotes the class of languages defined by the corresponding set of formulas.

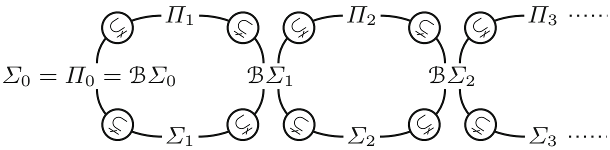

Fig. 1. Quantifier alternation hierarchy

The correspondence discovered by Thomas relates levels in the dot-depth hierarchy and levels in the quantifier alternation hierarchy of first-order logic, over the signature $(<, +1, \min, \max, \varepsilon, (P_a)_{a \in A})$.

Theorem 11 (Thomas [36]). *For any alphabet A, any $n \in \mathbb{N}$ and any language $L \subseteq A^*$, the two following properties hold:*

1. *L has dot-depth n iff L can be defined by a $\mathcal{B}\Sigma_n(<, +1, \mathit{min}, \mathit{max}, \varepsilon)$ sentence.*
2. *L has dot-depth $n + \frac{1}{2}$ iff L can be defined by a $\Sigma_{n+1}(<, +1, \mathit{min}, \mathit{max}, \varepsilon)$ sentence.*

Some years later, a similar correspondence was established between levels in the Straubing-Thérien hierarchy and levels in the quantifier alternation hierarchy over the original signature $(<, (P_a)_{a \in A})$. Such levels are defined analogously as for the enriched signature, and denoted by $\mathcal{B}\Sigma_n(<)$, $\Sigma_n(<)$, etc.

Theorem 12 (Perrin and Pin [16]). *For any alphabet A, any $n \in \mathbb{N}$ and any language $L \subseteq A^*$, the two following properties hold:*

1. *L has level n in the Straubing-Thérien hierarchy iff L can be defined by a $\mathcal{B}\Sigma_n(<)$ sentence.*
2. *L has level $n + \frac{1}{2}$ in the Straubing-Thérien hierarchy iff L can be defined by a $\Sigma_{n+1}(<)$ sentence.*

2 Generic Concatenation Hierarchies

Since the dot-depth and Straubing-Thérien hierarchies follow the very same construction scheme and enjoy similar properties, it is natural to generalize the definition. We will therefore define a generic notion of concatenation hierarchy. Such hierarchies should still classify languages according to the required number of alternations between concatenation and Boolean operations that are needed to define them. The only parameter in the construction is level 0, which is now any class of languages $\mathcal{C}$ satisfying some mild hypotheses (such as being a Boolean algebra). This parameter $\mathcal{C}$ is called the *basis* of the hierarchy. Once $\mathcal{C}$ is fixed, the construction process is uniform, exactly the same as for the two hierarchies we have already presented:

- Level 0 is the basis (*i.e.*, our parameter class $\mathcal{C}$).
- Each *half-level* is the *polynomial closure* of the previous full level: for any natural number $n \in \mathbb{N}$, level $n + \frac{1}{2}$ is the polynomial closure of level n.
- Each *full-level* is the *Boolean closure* of the previous half-level: for any natural number $n \in \mathbb{N}$, level $n + 1$ is the Boolean closure of level $n + \frac{1}{2}$.

For $q \in \mathbb{N}$ or $q \in \frac{1}{2} + \mathbb{N}$, let $\mathcal{C}[q]$ denote level q of the concatenation hierarchy of basis $\mathcal{C}$. By definition, we have $\mathcal{C}[n] \subseteq \mathcal{C}[n + \frac{1}{2}] \subseteq \mathcal{C}[n + 1]$ for any $n \in \mathbb{N}$. However, note that these inclusions need not be strict. For instance, if the basis is closed under Boolean operations and marked concatenation (such as the class of star-free languages), the associated hierarchy collapses at level 0. Of course the interesting hierarchies are the strict ones. We give a graphical representation of the construction process of a concatenation hierarchy in Fig. 2 below.

$$\underset{\text{(basis)}}{0} \xrightarrow{Pol} \tfrac{1}{2} \xrightarrow[Bool]{} 1 \xrightarrow{Pol} \tfrac{3}{2} \xrightarrow[Bool]{} 2 \xrightarrow{Pol} \tfrac{5}{2} \xrightarrow[Bool]{} 3 \xrightarrow{Pol} \tfrac{7}{2} \cdots\cdots$$

Fig. 2. A concatenation hierarchy

Notice that not all concatenations hierarchies are classifications of the star-free languages. Indeed, the generic definition now makes it possible to define hierarchies containing languages which are not star-free: it suffices to choose a basis containing such languages. The most famous one is the *group hierarchy* of Margolis and Pin [14], whose basis is the class of all regular languages recognized by an automaton in which every letter induces a permutation on the states.

The following result, which will be shown in the full version of this paper, generalizes Theorem 3 to any concatenation hierarchy whose basis is finite.

Theorem 13. *Let $\mathcal{C}$ be a finite Boolean algebra of regular languages over an alphabet of size at least 2. Then, the concatenation hierarchy of basis $\mathcal{C}$ is strict.*

Again, this theorem justifies the quest for algorithms deciding membership in levels of the hierarchy of basis $\mathcal{C}$.

Quantifier Alternation Hierarchies

The correspondence between star-free languages and first-order logic established by McNaughton and Papert in Theorem 2 can be lifted not only to the dot-depth and the Straubing-Thérien hierarchies (Theorems 11 and 12), but also to arbitrary concatenation hierarchies: for any basis $\mathcal{C}$, we associate a well-chosen first-order signature (also denoted by $\mathcal{C}$) such that the concatenation hierarchy of basis $\mathcal{C}$ and the quantifier alternation hierarchy within the variant $\text{FO}(\mathcal{C})$ of first-order logic equipped with this signature correspond. This signature contains all label predicates: for any $a \in A$, we have a unary predicate (denoted by "P_a") which is interpreted as the unary relation selecting all positions whose label is a. Moreover, for any language $L \in \mathcal{C}$, we add four predicates. To define them, we introduce the following notation: if $w = a_1 \cdots a_n$ is a word of length n, we denote by $w[i, j]$ its infix $a_i \cdots a_j$ (which is empty if $i > j$), and we let $w]i, j] = w[i+1, j]$, $w[i, j[= w[i, j-1]$ and $w]i, j[= w[i+1, j-1]$. We are now able to finish our description of the signature (associated to) $\mathcal{C}$. In addition to the strict order and the letter predicates, we add the following predicates for each language $L \in \mathcal{C}$:

– A binary predicate I_L. Its interpretation is as follows: given a word w and two positions i, j in w, $I_L(i, j)$ holds when $i < j$ and the infix $w]i, j[$ is in L.
– A unary predicate P_L. Its interpretation is as follows: given a word w and a position i in w, $P_L(i)$ holds when the prefix $w[1, i[$ is in L.
– A unary predicate S_L. Its interpretation is as follows: given a word w and a position i in w, $S_L(i)$ holds when the suffix $w]i, |w|]$ is in L.
– A nullary predicate N_L. Its interpretation is as follows: given a word w, N_L holds when w is in L.

Recall that we abuse notation and identify $\mathcal{C}$ with this signature. In other words, we denote by $\text{FO}(\mathcal{C})$ the associated variant of first-order logic.

We are now ready to state a generic correspondence between the concatenation hierarchy of basis $\mathcal{C}$ and the quantifier alternation hierarchy within $\text{FO}(\mathcal{C})$. We need an additional condition on $\mathcal{C}$: it should be closed under left and right quotients. That is, if L belongs to $\mathcal{C}$, then for any $a \in A$, so do its left and right quotients $a^{-1}L = \{w \in A^* \mid aw \in L\}$ and $La^{-1} = \{w \in A^* \mid wa \in L\}$.

Theorem 14. *Let $\mathcal{C}$ be a Boolean algebra of regular languages which is closed under left and right quotients. Then, for any finite alphabet A, any $n \in \mathbb{N}$ and any language $L \subseteq A^*$, the two following properties hold:*

1. $L \in \mathcal{C}[n]$ if and only if L can be defined by a $\mathcal{B}\Sigma_n(\mathcal{C})$ sentence.
2. $L \in \mathcal{C}[n + \frac{1}{2}]$ if and only if L can be defined by a $\Sigma_{n+1}(\mathcal{C})$ sentence.

3 Decision Problems

The membership problem for concatenation hierarchies is not well understood. For instance, although the dot-depth hierarchy has been given a lot of attention since 1971, obtaining membership algorithms for all of its levels remains one of the most famous open problems in automata theory. It has been under investigation for decades but progress is slow: as we explained above, the first known result is due to Knast [13] and yields an algorithm for dot-depth 1. Algorithms were later found for the half-levels $\frac{1}{2}$ in [21,22] and $\frac{3}{2}$ in [9,23]. However, it took more than thirty years to obtain an algorithm for the next full level: dot-depth 2 (see [25]). Furthermore, the problem is still open for dot-depth 3.

The result for level 2 is based on a new approach, which is the key idea we wish to convey in this survey. The approach relies on two main features:

1. It is generic to *all* concatenations hierarchies whose basis is *finite* (which is the case of the dot-depth and of the Straubing-Thérien hierarchies).
2. We consider decision problems which are *more general* than membership. While recent papers on the topic actually consider several such problems (see [28] for a global picture), we will focus on the simplest one: *separation*.

Let us define the separation problem. Consider a class of languages $\mathcal{C}$. Given two languages L_0 and L_1, we say that L_0 is $\mathcal{C}$-separable from L_1 if and only if there exists a third language $K \in \mathcal{C}$ such that $L_1 \subseteq K$ and $L_2 \cap K = \emptyset$. The separation problem for $\mathcal{C}$ is as follows:

INPUT: Two regular languages L_0 and L_1
OUTPUT: Is L_0 $\mathcal{C}$-separable from L_1?

The main reason why this problem is interesting is that solving it requires (and therefore, brings) much insight about the class $\mathcal{C}$. In particular, membership for $\mathcal{C}$ reduces to separation for $\mathcal{C}$. More interesting, if one has an algorithm in hand to decide *separation* for a given half-level in a concatenation hierarchy, then one can use it to obtain a new one deciding *membership* for the *next* half-level. This is what we formally state in the next theorem, which is essentially a result of [25] (note however that while the proof argument of [25] is generic to all hierarchies, the statement itself in [25] is specific to the Straubing-Thérien hierarchy).

Theorem 15. *Consider a basis $\mathcal{C}$ which is a Boolean algebra of regular languages closed under left and right quotients. Then, for any natural number $n \geq 1$, there exists an effective reduction from the membership problem for level $\mathcal{C}[n+\frac{1}{2}]$ to the separation problem for level $\mathcal{C}[n-\frac{1}{2}]$.*

This result is completed by the following theorem, which summarizes the recent results that have been obtained regarding the separation problem for low levels within concatenation hierarchies. The first two items are taken from [29] and the third one is an unpublished generalization of a result of [24] (which states that separation for level $\frac{5}{2}$ in the Straubing-Thérien hierarchy is decidable).

Theorem 16. *Consider an arbitrary* finite *Boolean algebra $\mathcal{C}$ which is closed under left and right quotients. Then the following results hold:*

1. *$Pol(\mathcal{C})$-separation is decidable.*
2. *$BPol(\mathcal{C})$-separation is decidable.*
3. *$Pol(BPol(\mathcal{C}))$-separation is decidable.*

Altogether, this yields that for any concatenation hierarchy whose basis is *finite*, levels $\frac{1}{2}$, 1 and $\frac{3}{2}$ have decidable separation. Moreover, this can be combined with Theorem 15 to obtain the decidability of membership for level $\frac{5}{2}$.

These results are generic to all concatenations hierarchies whose basis is finite. However, in the special case of the dot-depth and Straubing-Thérien hierarchies, one can do better and lift them to levels 2 and $\frac{5}{2}$ for separation (and thus to level $\frac{7}{2}$ for membership). These stronger results are based on a specific property of the Straubing-Thérien hierarchy: its level $\frac{3}{2}$ is also level $\frac{1}{2}$ in another concatenation hierarchy having a finite basis. Let us explain this statement in more details.

Back to the Dot-Depth and Straubing-Thérien Hierarchies

In this final part, we explain why one may lift all results one level higher in the dot-depth and Straubing-Thérien hierarchies. The argument relies on a theorem of Pin and Straubing [20], which implies that levels $\frac{3}{2}$ and above in the Straubing-Thérien hierarchy are also levels $\frac{1}{2}$ and above in the concatenation hierarchy whose basis is the *finite* class AT of alphabet testable languages, defined below. While simple, this result is crucial: it allows us to lift the separation results of Theorem 16 to levels 2 and $\frac{5}{2}$ of the Straubing-Thérien hierarchy.

Let us define the class AT of *alphabet testable languages*. It consists of all Boolean combinations of languages of the form,

$$A^*aA^* \quad \text{for some}\, a \in A.$$

Clearly AT is finite, and one may verify that it is a Boolean algebra closed under left and right quotients. It was proved by Pin and Straubing [20] that level $\frac{3}{2}$ in the Straubing-Thérien hierarchy[1] is also the class $Pol(\text{AT})$.

Note that the original formulation of this statement by Pin and Straubing is that level $\frac{3}{2}$ in the Straubing-Thérien hierarchy consists exactly of unions of languages of the form,

$$B_0^*a_1B_1^*a_2B_2^* \cdots a_nB_n^* \quad \text{with} \quad B_0, \dots, B_n \subseteq A.$$

We reformulate this result in the following theorem.

Theorem 17 (Pin and Straubing [20]). *Level $\frac{3}{2}$ in the Straubing-Thérien hierarchy is exactly the class $Pol(\text{AT})$. In particular, any level $n \geq \frac{3}{2}$ (half or full) in the Straubing-Thérien hierarchy corresponds exactly to level $n-1$ in the concatenation hierarchy of basis* AT.

The important point here is that while AT is more involved than $\{\emptyset, A^*\}$ as a basis, it remains *finite*. Therefore, Theorem 17 states that any level $n \geq \frac{3}{2}$ in the Straubing-Thérien hierarchy is also level $n-1$ in another hierarchy whose basis is finite. This result is crucial. Indeed, this means that Theorem 16 does not only apply to levels $\frac{1}{2}$, 1 and $\frac{3}{2}$ of the Straubing-Thérien hierarchy but also to levels 2 and $\frac{5}{2}$. Altogether, we get the following corollary.

[1] In fact, the original formulation of Pin and Straubing considers level 2 in the Straubing-Thérien hierarchy and not level $\frac{3}{2}$.

Corollary 18. *The separation problem is decidable for levels* 2 *and* $\frac{5}{2}$ *in the Straubing-Thérien hierarchy. Moreover, the membership problem is decidable for level* $\frac{7}{2}$.

Finally, these results can be lifted to the dot-depth hierarchy using an approach which is similar to the one used by Straubing in Theorem 10. Indeed, recall from Theorem 10 that the Straubing-Thérien hierarchy can be viewed as "more fundamental" than the dot-depth. It turns out that the reduction provided by Straubing can actually be lifted to half-levels [23] and to separation [26].

Theorem 19. *For any level* n *in the dot-depth hierarchy, the following two properties hold:*

- *If membership is decidable for level* n *in the Straubing-Thérien, then it is decidable for level* n *in the dot-depth hierarchy as well.*
- *If separation is decidable for level* n *in the Straubing-Thérien, then it is decidable for level* n *in the dot-depth hierarchy as well.*

Corollary 20. *The separation problem for levels* 2 *and* $\frac{5}{2}$ *in the dot-depth hierarchy are decidable. Moreover, the membership problem is decidable for level* $\frac{7}{2}$.

4 Conclusion

In this extended abstract, we outlined part of the (slow) progress that occurred during the last decades regarding concatenation hierarchies. We refer the reader to the full version of the paper for details, and to [18,19,27,37] for surveys on this fascinating subject.

References

1. Arfi, M.: Polynomial operations on rational languages. In: Brandenburg, F.J., Vidal-Naquet, G., Wirsing, M. (eds.) STACS 1987. LNCS, vol. 247, pp. 198–206. Springer, Heidelberg (1987). doi:10.1007/BFb0039607
2. Arfi, M.: Opérations Polynomiales et Hiérarchies de Concaténation. Theoret. Comput. Sci. **91**(1), 71–84 (1991)
3. Bojanczyk, M.: Star height via games. In: 2015 30th Annual ACM/IEEE Symposium on Logic in Computer Science. IEEE Computer Society 2015, pp. 214–219 (2015)
4. Brzozowski, J.A.: Developments in the Theory of regular Languages. In: IFIP Congress, pp. 29–40 (1980)
5. Brzozowski, J.A., Cohen, R.S.: Dot-depth of star-free events. J. Comput. Syst. Sci. **5**(1), 1–16 (1971)
6. Brzozowski, J.A., Knast, R.: The Dot-depth hierarchy of star-free languages is infinite. J. Comput. Syst. Sci. **16**(1), 37–55 (1978)
7. Dejean, F., Schützenberger, M.P.: On a question of Eggan. Inf. Control **9**(1), 23–25 (1966)
8. Eggan, L.C.: Transition graphs and the star-height of regular events. Michigan Math. J. **10**(4), 385–397 (1963)
9. Glaßer, C., Schmitz, H.: Languages of dot-depth 3/2. Theory Comput. Syst. **42**(2), 256–286 (2007)

10. Hashiguchi, K.: Algorithms for determining relative star height and star height. Inf. Comput. **78**(2), 124–169 (1988)
11. Kirsten, D.: Distance desert automata and the star height problem. RAIRO-Theor. Inf. Appl. **39**(3), 455–509 (2005)
12. Kleene, S.C.: Representation of events in nerve nets and finite automata. In: Shannon, C., McCarthy, J. (eds.) Annals of Mathematics Studies 34, pp. 3–41. Princeton University Press, New Jersey (1956)
13. Knast, R.: A semigroup characterization of dot-depth one languages. RAIRO - Theor. Inform. Appl. **17**(4), 321–330 (1983)
14. Margolis, S.W., Pin, J.E.: Products of group languages. In: Budach, L. (ed.) FCT 1985. LNCS, vol. 199, pp. 285–299. Springer, Heidelberg (1985). doi:10.1007/BFb0028813
15. McNaughton, R., Papert, S.A.: Counter-Free Automata. MIT Press, Cambridge (1971)
16. Perrin, D., Pin, J.É.: First-order logic and star-free sets. J. Comput. Syst. Sci. **32**(3), 393–406 (1986)
17. Pin, J.É.: An explicit formula for the intersection of two polynomials of regular languages. In: Béal, M.-P., Carton, O. (eds.) DLT 2013. LNCS, vol. 7907, pp. 31–45. Springer, Heidelberg (2013). doi:10.1007/978-3-642-38771-5_5
18. Pin, J.-É.: Open problems about regular languages, 35 years later. In: The Role of Theory in Computer Science. Essays Dedicated to Janusz Brzozowski. World Scientific (2017)
19. Pin, J.-É.: The dot-depth hierarchy, 45 years later. In: The Role of Theory in Computer Science. Essays Dedicated to Janusz Brzozowski. World Scientific (2017)
20. Pin, J.-É., Straubing, H.: Monoids of upper triangular boolean matrices. In: Semigroups. Structure and Universal Algebraic Problems, vol. 39, pp. 259–272. North-Holland (1985)
21. Pin, J.-E., Weil, P.: Polynomial closure and unambiguous product. In: Fülöp, Z., Gécseg, F. (eds.) ICALP 1995. LNCS, vol. 944, pp. 348–359. Springer, Heidelberg (1995). doi:10.1007/3-540-60084-1_87
22. Pin, J.É., Weil, P.: Polynomial closure and unambiguous product. Theory Comput. Syst. **30**(4), 383–422 (1997)
23. Pin, J.É., Weil, P.: The wreath product principle for ordered semigroups. Commun. Algebra **30**(12), 5677–5713 (2002)
24. Place, T.: Separating regular languages with two quantifier alternations. In: 30th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2015, pp. 202–213 (2015)
25. Place, T., Zeitoun, M.: Going higher in the first-order quantifier alternation hierarchy on words. In: Esparza, J., Fraigniaud, P., Husfeldt, T., Koutsoupias, E. (eds.) ICALP 2014. LNCS, vol. 8573, pp. 342–353. Springer, Heidelberg (2014). doi:10.1007/978-3-662-43951-7_29
26. Place, T., Zeitoun, M.: Separation and the successor relation. In: 32nd International Symposium on Theoretical Aspects of Computer Science, STACS 2015. Dagstuhl, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, pp. 662–675 (2015)
27. Place, T., Zeitoun, M.: The tale of the quantifier alternation hierarchy of first-order logic over words. SIGLOG news **2**(3), 4–17 (2015)
28. Place, T., Zeitoun, M.: The covering problem: a unified approach for investigating the expressive power of logics. In: Proceedings of the 41st International Symposium on Mathematical Foundations of Computer Science, MFCS 2016, pp. 77:1–77:15 (2016)

29. Place, T., Zeitoun, M.: Separation for dot-depth two. In: 32th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017 (2017)
30. Schützenberger, M.P.: On finite monoids having only trivial subgroups. Inf. Control **8**(2), 190–194 (1965)
31. Simon, I.: Hierarchies of Events of Dot-Depth One. Ph.D. thesis. Waterloo, Ontario, Canada: University of Waterloo, Department of Applied Analysis and Computer Science (1972)
32. Simon, I.: Piecewise testable events. In: Brakhage, H. (ed.) GI-Fachtagung 1975. LNCS, vol. 33, pp. 214–222. Springer, Heidelberg (1975). doi:10.1007/3-540-07407-4_23
33. Straubing, H.: A generalization of the schützenberger product of finite monoids. Theoret. Comput. Sci. **13**(2), 137–150 (1981)
34. Straubing, H.: Finite semigroup varieties of the form V * D. J. Pure Appl. Algebra **36**, 53–94 (1985)
35. Thérien, D.: Classification of finite monoids: the language approach. Theoret. Comput. Sci. **14**(2), 195–208 (1981)
36. Thomas, W.: Classifying regular events in symbolic logic. J. Comput. Syst. Sci. **25**(3), 360–376 (1982)
37. Weil, P.: Concatenation product: a survey. In: Pin, J.E. (ed.) LITP 1988. LNCS, vol. 386, pp. 120–137. Springer, Heidelberg (1989). doi:10.1007/BFb0013116