Classifying Regular Languages by a Split Game

Qiqi Yan*
BASICS Laboratory,
Department of Computer Science and Engineering,
Shanghai Jiao Tong University,
200240, Shanghai, P.R. China
contact@yanqiqi.net,
http://www.yanqiqi.net

April 19, 2006

Abstract

In this paper, we introduce a variant of the Ehrenfeucht-Fraïssé game from logic which is useful for analyzing the expressive power of classes of generalized regular expressions. An extension of the game to generalized ω-regular expressions is also established. To gain insight into how the game can be applied to attack the long-standing generalized star height 2 problem, we propose and solve a related but easier problem, the omega power problem. Namely we show that omega powers, together with boolean combinations and concatenations, are not sufficient to express the class of ω-regular languages.

Keywords: split game, EF game, regular expression, star height, omega power

1 Introduction

It is natural to classify regular languages by star height (see [1] for a historical survey), which is the minimum nesting depth of stars of a regular expression representing the language. In the definition of regular expressions, if only union, concatenation and star are allowed as basic operators ("restricted star height" in this context), Eggan [2] showed that languages of arbitrary restricted star height exist, and Hashiguchi [3] showed that the restricted star height of a given regular language can be computed effectively. If complement is also considered as a basic operator (which we assume throughout rest of the paper), the notion of star height ("generalized star height") seems to be more interesting. It is known that the star-free languages, i.e. languages of star height 0, have various characterizations, like having finite aperiodic syntactic monoids [4] or being first

*Supported by NSFC No. 60273050.

order logic definable [5], and hierarchies of star-free languages like the dot-depth hierarchy have also been extensively studied in literature. From these results, it follows that languages of star height 1 exist (e.g. $(aa)^*$) and the class of languages of star height 0 is decidable. However, beyond that, it is still open if a language of star height > 1 exists! We prefer to state this "star height 2 problem" [6] as an inexpressibility problem:

Star Height 2 Problem: Does there exist a regular language $\mathcal{L}$ such that no regular expression of star height[1] ≤ 1 represents $\mathcal{L}$?

Understanding of this problem is important for the even tougher *star height problem*, i.e., is there an algorithm that computes the star height of a given regular language?

The difficulty of the star height 2 problem might come from the fact that the regular expressions of star height ≤ 1 are surprisingly expressive. In fact most related work (like [7][8][9][6]) resulted in larger and larger classes of languages of star height ≤ 1, while eliminating previous candidates for languages of star height 2. We conjecture that languages of star height 2 exist and what is lacking is some tool for such inexpressibility results.

In formal logic, the Ehrenfeucht-Fraïssé game has proved to be a powerful tool for inexpressibility results. It has been successfully applied to the dot-depth hierarchy [10] within star-free languages. Thomas first showed that the dot-depth hierarchy corresponds with the quantifier alternation hierarchy of first-order logic [11]. It follows that the Ehrenfeucht-Fraïssé game [12] can be applied to give a simpler proof [13][14] for the strictness of the dot-depth hierarchy [15]. Such results gave us a hint. If the star height hierarchy can be characterized by some natural hierarchy in logic, then one might be able to derive a variant of the Ehrenfeucht-Fraïssé game to solve problems on star height as well. So Thomas considered a natural hierarchy of regular languages based on weak monadic second order quantifiers, but the hierarchy was proved to collapse [11]. Another possibility is from the characterization of regular languages by FO(mTC) [16], i.e., first order logic equipped with monadic transitive closure operators. However, mTC operators seem to be too expressive to characterize stars, as two such operators are sufficient to define arbitrary regular languages[2] [19]. For the same reason, the monadic partition logic in [20] is also not suitable.

Although no logical characterization of star height has been found, the star height problem is more like a logic problem in nature. Recall that regular expressions have operators $\cup, \cap$ and $\sim$ (complement), which naturally correspond with logical connectives $\vee, \wedge$ and $\neg$. So in fact regular expressions can be seen as the formulas of some special logic. With this idea in mind, we derive a "split game", which gives a characterization of the star height hierarchy. It follows

[1] The star height of a regular expression is the nesting depth of stars in the expression.

[2] This leaves with one possibility. If one can use the Ehrenfeucht-Fraïssé game for FO(TC) [17] [18] to show that some regular language cannot be defined by a FO(mTC) formula with no nesting of mTC operators, then the star height 2 problem is solved. Unfortunately, it is well-known in finite model theory that games like the ones for FO(TC) are difficult to play when the structures are ordered (like word structures).

that separation of the hierarchy is reduced to finding suitable winning strategies for one of the players in the corresponding game. (However, we don't yet know how to use this new approach to solve the star height 2 problem, whose solution still seems far from reach.) The split game can also be used to give characterizations for some other hierarchies of regular languages, including the dot-depth hierarchy. In general, it gives rise to a combinatorial method for studying the expressive power of classes of regular expressions, and obtaining inexpressibility results.

We also establish an extension of the split game to ω-regular languages (see [21] for a survey) in order to study the expressive power of the omega power operator, which is also far from being well-understood like the star operator. Using such an extension, we propose and solve the omega power problem. Namely we show that omega powers, together with boolean combinations and concatenations, are not sufficient to express the class of ω-regular languages. Besides its own interest, we argue that this problem can be seen as a simplified version of the star height 2 problem, and such result might give us some insight for attacking the star height 2 problem.

1.1 Related Work

After sending the draft of this paper to colleagues, Jean-Eric Pin informed us that a similar game was proposed by Wolfgang Thomas in an unpublished note. We thank J.E. Pin for giving us this hint and W. Thomas for providing us that note. Compared to Thomas's game, our treatment is somewhat more natural and general, and our proof of the completeness theorem is much simpler.

2 Regular Expressions and Classes

Fix a finite alphabet Σ, regular languages are built from $\emptyset, \{\epsilon\}$ (ϵ denotes the empty word) and letter sets $\{a\}$ ($a \in \Sigma$) using boolean combination, concatenation and star. Correspondingly, regular languages are represented by (generalized) regular expressions, which are built from symbols $\emptyset, \epsilon, a$ ($a \in \Sigma$) using boolean symbols $\cup, \cap, \sim$, concatenation dot $\cdot$ and star $*$. Two regular expressions are equivalent if they represent the same language.

The (generalized) star height $h(\phi)$ of a regular expression ϕ is the nesting depth of stars in ϕ. E.g., $h(((a^* \cdot b)^* \cdot c)^* \cap \sim c^*) = 3$. The *(generalized) star height* $h(\mathcal{L})$ of a regular language $\mathcal{L}$ is the minimum of $h(\phi)$ with ϕ a regular expression representing $\mathcal{L}$.

Definition 1 *For p a finite word over the operator alphabet $OP = \{\odot, \circledast\}$, the* expression class *(or simply* class*) $\mathcal{C}(p)$ is inductively defined as:*

- *$\mathcal{C}(\epsilon)$ is the closure of the class of basic expressions $\emptyset, \epsilon, a (a \in \Sigma)$ under $\cup, \cap$ and $\sim$.*
- *$\mathcal{C}(\odot p)$ is the closure of the class of expressions $\mathcal{C}(p) \cup \{\phi \cdot \psi : \phi, \psi \in \mathcal{C}(p)\}$ under $\cup, \cap$ and $\sim$.*

- $\mathcal{C}(\circledast p)$ *is the closure of the class of expressions* $\mathcal{C}(p) \cup \{\phi^* : \phi \in \mathcal{C}(p)\}$ *under* $\cup, \cap$ *and* $\sim$.

Corollary 2 *For* ϕ *a regular expression and* $m \geq 0$, $h(\phi) \leq m$ *if and only if* ϕ *is in* $\bigcup_{n \geq 0} \mathcal{C}((\odot^n \circledast)^m \odot^n)$.

A key fact about $\mathcal{C}(p)$ is that:

Lemma 3 *Every expression class* $\mathcal{C}(p)$ *is finite, up to equivalence.*

Proof. We prove by induction on p. Recall that Σ is finite. So $\mathcal{C}(\epsilon)$ is finite up to equivalence. If $p = \odot q$ for some q, then $\mathcal{C}(p)$ is generated using booleans from $\mathcal{C}(q)$ and the set $\{\phi \cdot \psi \mid \phi, \psi \in \mathcal{C}(q)\}$, both finite up to equivalence by induction hypothesis. So $\mathcal{C}(p)$ is also finite up to equivalence. The case $p = \circledast q$ is similar. ■

We adopt the Tarskian notation $u \models \phi$ denoting that word u is in the language represented by expression ϕ. For $u, v \in \Sigma^*$ and class $\mathcal{C}(p)$, we write $u \equiv_p v$ if for every $\phi \in \mathcal{C}(p)$, $u \models \phi$ iff $v \models \phi$. We also define the regular expression χ_u^p as $\bigcap\{\phi \in \mathcal{C}(p) : u \models \phi\}$. Note that such definition is valid by Lemma 3. The following lemma states that χ_u^p characterizes the $\equiv_p$ equivalence class of u.

Lemma 4 *For every* $u, v \in \Sigma^*$ *and class* $\mathcal{C}(p)$, $v \models \chi_u^p$ *iff* $u \equiv_p v$.

Proof. Clearly if $u \equiv_p v$ then $v \models \chi_u^p$. For the converse, assume $v \models \chi_u^p$. For every $\phi \in \mathcal{C}(p)$, if $u \models \phi$ then ϕ is a conjunct of χ_u^p and thus $v \models \phi$. If $u \not\models \phi$, then $\sim \phi$ is a conjunct of χ_u^p and thus $v \models \sim \phi$. Together we have $u \equiv_p v$. ■

3 The Split Game

In this section we introduce our split game. Some techniques used in designing the game are borrowed from the variants of the Ehrenfeucht-Fraïssé game in logic, e.g. Grädel's FO(TC) game [17].

Definition 5 *For every expression class* $\mathcal{C}(p)$ *and words* $u, v \in \Sigma^*$, *in the following we inductively define the game* $\mathcal{G}_p$ *over* (u, v), *which is played by players Samson and Delilah with initial configuration* (u, v):

- *If* $p = \epsilon$, *then this is the trivial game where Delilah wins if* $u \equiv_\epsilon v$ *(i.e. either* $|u|, |v| \geq 2$, *or both* $|u| = |v| \leq 1$ *and* $u = v$*) and Samson wins otherwise.*
- *If* $p = aq$ *for some* $a \in OP$, *then Samson can choose to play a split round if* $a = \odot$ *or play a* $*$*-split round if* $a = \circledast$. *He can also choose to play an empty round instead, i.e. to do nothing. After the round is played, the players continue to play* $\mathcal{G}_q$ *with the updated configuration.*

split round: *Samson splits u into $u = u_1 \cdot u_2$* [3] *and then Delilah splits v into $v = v_1 \cdot v_2$. Next Samson chooses i from $1, 2$ and configuration of the game is updated to (u_i, v_i). Samson can also choose to play such a round with u, v interchanged.*

$*$-split round: *Samson splits u into $u = u_1 \cdot \ldots \cdot u_m$ for some $m \geq 1$ and then Delilah splits v into $v = v_1 \cdot \ldots \cdot v_n$ for some $n \geq 1$. Next Samson chooses some j from $1 \ldots n$, and then Delilah responds with some i from $1 \ldots m$. Configuration of the game is updated to (u_i, v_j). Samson can also choose to play such a round with u, v interchanged.*

We write $u \sim_p v$ if Delilah has a winning strategy for $\mathcal{G}_p$ over (u, v). Note that $\mathcal{G}_p$ is a finite game of perfect information. Exactly one of the players has a winning strategy. We will assume that players always follow their winning strategies, if there exist.

Note that in the split game, the words that the games are played on are repeatedly shortened. In contrast, in the Ehrenfeucht-Fraïssé games (and also in Thomas's concatenation games), all the information about the playground are kept. Such difference is crucial in making our game work.

Lemma 6 *For every $u, v \in \Sigma^*$ and class $\mathcal{C}(p)$, $u \equiv_p v$ if and only if $u \sim_p v$.*

Proof. The case $p = \epsilon$ is trivial by definition. Assume that the claim holds for all proper suffix q of p.

Suppose $u \not\equiv_p v$. We describe a winning strategy of Samson for $\mathcal{G}_p$ over (u, v). If $u \not\equiv_q v$ for some proper suffix q of p, then Samson can win by first playing some empty rounds till the $\mathcal{G}_q$ game, and then applying his winning strategy which exists by induction hypothesis. So we assume the minimality of p, that is $u \equiv_q v$ for all proper suffix q of p. As $u \not\equiv_p v$, some $\phi \in \mathcal{C}(p)$ distinguishes u, v. If $\phi = \sim \psi_1$ or $\psi_1 \cup \psi_2$ or $\psi_1 \cap \psi_2$, then at least one of ψ_1, ψ_2 (also in $\mathcal{C}(p)$) distinguishes u, v. So we assume that the main operator of ϕ is not boolean. W.l.o.g. let $u \models \phi$ and $v \not\models \phi$. There are two cases:

$\phi = \psi_1 \cdot \psi_2$: By the minimality of p, ϕ is in $\mathcal{C}(p) \backslash \mathcal{C}(q)$, and so by the definition of $\mathcal{C}(p)$, ψ_1, ψ_2 are in $\mathcal{C}(q)$ with $p = \odot q$. As $u \models \psi_1 \cdot \psi_2$, Samson can split u into $u_1 \cdot u_2$ such that $u_i \models \psi_i$ for each $i = 1, 2$. As $v \not\models \psi_1 \cdot \psi_2$, no matter how Dilelah splits v into $v_1 \cdot v_2$, we have $v_i \not\models \psi_i$ for some $i = 1, 2$. Let Samson choose this i. So $u_i \not\equiv_q v_i$ and he can win the remaining $\mathcal{G}_q$ game over (u_i, v_i).

$\phi = \psi^*$: Similarly, ψ is in $\mathcal{C}(q)$ with $p = \circledast q$. As $u \models \psi^*$, Samson can split u into $u_1 \cdot \ldots \cdot u_m$ such that $u_i \models \psi$ for each $1 \leq i \leq m$. As $v \not\models \psi^*$, no matter how Delilah splits v into $v_1 \cdot \ldots \cdot v_n$, we have $v_j \not\models \psi$ for some $1 \leq j \leq n$. Let Samson choose this j. Whichever i Delilah chooses, $u_i \models \psi$. So $u_i \not\equiv_q v_j$ and Samson can win the remaining $\mathcal{G}_q$ game over (u_i, v_j).

---

[3] We use dots to indicate how words are split. Other dots are usually omitted.

Suppose $u \equiv_p v$. We describe a winning strategy of Delilah. Let $p = aq$ for some $a \in OP$. There are three cases, depending on the type of the first round played in $\mathcal{G}_p$:

empty round: $u \equiv_p v$ implies $u \equiv_q v$, and so Delilah can win the remaining $\mathcal{G}_q$ game.

split round: So $p = \odot q$. Assume that Samson splits u into $u_1 \cdot u_2$. Consider $\phi = \chi^q_{u_1} \cdot \chi^q_{u_2}$ in $\mathcal{C}(p)$. Clearly $u \models \phi$. As $u \equiv_p v$, $v \models \phi$ too. Then Delilah can split v into $v_1 \cdot v_2$ such that $v_1 \models \chi^q_{u_1}, v_2 \models \chi^q_{u_2}$. Whichever i Samson chooses, $u_i \equiv_q v_i$ by Lemma 4 and Delilah can win the remaining $\mathcal{G}_q$ game over (u_i, v_i).

$*$-split round: So $p = \circledast q$. Assume that Samson splits u into $u_1 \cdot \ldots \cdot u_m$. Consider $\phi = \bigcup_{i=1}^m \chi^q_{u_i}$ in $\mathcal{C}(q)$. Clearly $u \models \phi^*$. As $u \equiv_p v$, $v \models \phi^*$ too. Then Delilah can split v into $v_1 \cdot \ldots \cdot v_n$ such that $v_j \models \phi$ for each $1 \leq j \leq n$. Whichever j Samson chooses, $v_j \models \chi^q_{u_i}$ for some $1 \leq i \leq m$. Let Delilah respond with this i. So $u_i \equiv_q v_j$ and she can win the remaining $\mathcal{G}_q$ game over (u_i, v_j).

■

Theorem 7 *For every language $\mathcal{L}$ and class $\mathcal{C}(p)$, the following are equivalent:*

(i) *No regular expression in class $\mathcal{C}(p)$ represents $\mathcal{L}$.*

(ii) *There exist $u \in \mathcal{L}, v \notin \mathcal{L}$ such that Delilah has a winning strategy for $\mathcal{G}_p$ over (u, v).*

Proof. (ii)$\Longrightarrow$(i). Suppose $\neg$(i) and let $\phi \in \mathcal{C}(p)$ represent $\mathcal{L}$. So ϕ distinguishes u, v and $u \not\equiv_p v$. By Lemma 6, Samson can win $\mathcal{G}_p$ over (u, v), contradiction.

(i)$\Longrightarrow$(ii). Suppose $\neg$(ii). Consider $\phi = \bigcup\{\chi^p_u \mid u \in \mathcal{L}\}$ in $\mathcal{C}(p)$ (this definition is valid by Lemma 3). For each $u \in \mathcal{L}$, clearly $u \models \phi$. Conversely, for each $v \in \Sigma^*$, if $v \models \phi$, then $v \models \chi^p_u$ for some $u \in \mathcal{L}$. Thus $u \equiv_p v$ and Delilah can win $\mathcal{G}_p$ over (u, v). So $v \in \mathcal{L}$, or otherwise (ii) is satisfied. Therefore ϕ represents $\mathcal{L}$, contrary to (i). ■

Together with Corollary 2, we have the following characterization of star height.

Theorem 8 *For a regular language $\mathcal{L}$, the following are equivalent:*

(i) *The star height of $\mathcal{L}$ is strictly greater than m.*

(ii) *For each $n \geq 0$, there exist $u_n \in \mathcal{L}$, $v_n \notin \mathcal{L}$ such that Delilah has a winning strategy for $\mathcal{G}_p$ over (u_n, v_n), where $p = (\odot^n \circledast)^m \odot^n$.*

The definitions of $\mathcal{C}(p)$ and $\mathcal{G}_p$ can in fact be refined, by considering $\sim$ as a nontrivial operator. So $\circledsim$ is added into OP and swap rounds are introduced in which u, v are swaped. With appropriate modifications, languages of dot-depth

[10] m are characterized by $\bigcup_{n\geq 0}\mathcal{C}((\circledsim\odot^n)^{m-1})$ and thus can be characterized by the split game. Another way to characterize dot-depth is to introduce n-splits like in Thomas's concatenation game [14]. In general, the split game can be tailored like the Ehrenfeucht-Fraïssé game for specific purposes.

To compare to Thomas's concatenation game [14] in the context of the dot-depth hierarchy, the split game is simpler to use, since fractions of the words irrelevant to future rounds of the game are discarded immediately. Particularly, the split game can help to simplify the presentation of the proof of the strictness of the dot-depth hierarchy in [14].

4 Playing the Game

We are interested in attacking the star height 2 problem using the split game. But what we can give now is just a possible route. By Theorem 8, the problem is reduced to finding for each n two words u_n, v_n distinguished by some fixed regular language such that $u_n \sim_p v_n$ where $p = \odot^n \circledast \odot^n$. For such purpose, we need to first investigate how to construct words that are $\sim_n, \sim_{*,n}$ and $\sim_{n,*,n}$ equivalent step by step. Here $\sim_n, \sim_{*,n}$ and $\sim_{m,*,n}$ are aliases for $\sim_p$ when $p = \odot^n, \circledast\odot^n$ and $\odot^m\circledast\odot^n$ respectively.

Fix n, and set constant T_n to be 2^{n+1}. For word $w \in \Sigma^*$, we say w^m is in the form $w^{\geq k}$ if $m \geq k$. For letter $a \in \Sigma$, we say a^m is in the form $\tilde{a}$ if $m \geq T_n$. The facts below have already appeared in different forms in literature (e.g. [14][22]).

Lemma 9 **(i)** *For letter $a \in \Sigma$, words in the form $\tilde{a}$ are $\sim_n$ equivalent.*

(ii) *For word $w \in \Sigma^*$, words in the form $w^{\geq T_n}$ are $\sim_n$ equivalent.*

(iii) *For $u, v, u', v' \in \Sigma^*$, if $u \sim_n v$ and $u' \sim_n v'$, then $uu' \sim_n vv'$. In other words, $\sim_n$ is a congruence relation.*

Proof. For (i), we prove by induction on n that $a^x \sim_n a^y$ if $x, y \geq T_n$. The case $n = 0$ and the case that Samson first plays an empty round are both easy. Assume $n > 0$, $x, y \geq T_n$, and Samson splits a^x into $a^l \cdot a^r$. Let $l \leq r$, then Delilah can split a^y into $a^{\min(l,T_{n-1})} \cdot a^{y-\min(l,T_{n-1})}$. Note that $r \geq T_{n-1}$, $y - \min(l, T_{n-1}) \geq T_{n-1}$ and so $a^r \sim_{n-1} a^{y-\min(l,T_{n-1})}$ by induction hypothesis. One can also easily show that $a^l \sim_{n-1} a^{\min(l,T_{n-1})}$. Thus Delilah can win the remaining game.

(ii) follows from a slightly generalized argument of the proof of (i).

For (iii), we prove by induction on n. The case $n = 0$ is trivial. Suppose that Samson splits uu' into $u_1 \cdot u_2u'$. Since $u \sim_n v$, Delilah can split vv' into $v_1 \cdot v_2v'$ such that $u_1 \sim_{n-1} v_1$ and $u_2 \sim_{n-1} v_2$. Clearly $u' \sim_{n-1} v'$, and so $u_2u' \sim_{n-1} v_2v'$ by induction hypothesis. Then clearly Delilah can win the remaining game. ■

Obviously it is quite flexible to construct words that are $\sim_n$ equivalent. For example, for fixed l, r, all words in the form $a^l b(\tilde{a}b)^{\geq T_n} a^r$ are $\sim_n$ equivalent. Such results will be used freely in the rest of the paper.

For the $\sim_{*,n}$ relation, equivalent words are much more difficult to construct. In fact even the following problem is open. As split rounds are easier for Delilah to play than $*$-split rounds, once we can solve this problem, a solution to the star height 2 problem might be quite near.

Problem 10 *Does there exist a regular language $\mathcal{L}$ such that no regular expression in $\bigcup_{n\geq 0}\mathcal{C}(\circledast\odot^n)$ represents $\mathcal{L}$?*

5 An Extension to ω-Regular Languages

In this section, we establish an extension of the split game to ω-regular languages. This enables us to study the expressive power of the omega power operator by the split game. Concatenation hierarchies for infinite words [23] can also be studied by variants of such extension.

Fix Σ, an ω-word over Σ is an infinite sequence of letters from Σ: $a_1a_2\ldots$. The set of all ω-words over Σ are denoted by Σ^ω. A subset of Σ^ω is called an ω-language. The class of (generalized) ω-regular expressions, which represents ω-regular languages (see [21] for a survey), is the closure of the class of expressions $\{\phi^\omega : \phi \text{ is a regular expression}\} \cup \{\emptyset\}$ under boolean combination (w.r.t. Σ^ω) and left concatenation with a regular expression. Here ϕ^ω, the *omega power* of ϕ, represents the ω-language $\{u_1u_2\ldots \mid u_i \models \phi \text{ and } u_i \neq \epsilon, \text{ for } i \geq 1\}$ and for ϕ a regular expression and ψ an ω-regular expression, $\phi\cdot\psi$, the *left concatenation* of ψ with ϕ, represents the ω-language $\{uv \in \Sigma^\omega : u \models \phi, v \models \psi, u \in \Sigma^*, v \in \Sigma^\omega\}$.

For example, consider the expression $((a\cdot a)\cup b)^\omega \cap \sim((\sim\emptyset)\cdot b\cdot a\cdot(\sim\emptyset))$. A more intuitive way to write this expression is $(aa\cup b)^\omega\backslash(\Sigma^*ba\Sigma^\omega)$. (Note that concatenation and left concatenation share the same $\cdot$ symbol. Also complement w.r.t. Σ^* and complement w.r.t. Σ^ω share the same $\sim$ symbol. But the meaning of a symbol is clear from context, and no ambiguity will arise.) A little thought shows that it is equivalent to $(aa)^*b^\omega$. In fact one can show that this ω-language cannot be represented by an ω-regular expression without the use of any star or omega power operator.

Let $OP_\omega = OP\cup\{\textcircled{\omega}\}$. Expression class $\mathcal{C}^\omega(p)$ is defined similarly for every finite word p over OP_ω if p has at most one occurrence of $\textcircled{\omega}$ only and no $\circledast$ precedes any $\textcircled{\omega}$. Here $\mathcal{C}^\omega(\textcircled{\omega}p)$ is the closure of the class of expressions $\mathcal{C}(p)\cup\{\phi^\omega : \phi\in\mathcal{C}^\omega(p)\}$ under $\cup,\cap$ and $\sim$.

For an ω-word α, an *ω-split* U of α is an infinite sequence of finite words $u_1, u_2, \ldots$ such that $\alpha = u_1\cdot u_2\cdot\ldots$ with each u_i in $\Sigma^*\backslash\{\epsilon\}$. We also regard U as the set $\{u_i : i\geq 1\}$.

For every class $\mathcal{C}^\omega(p)$ and $u, v\in\Sigma^*\cup\Sigma^\omega$, we can correspondingly modify Definition 5 to define $\mathcal{G}_p^\omega$ over (u, v), if u, v are either both in Σ^* or are both in Σ^ω. With the following modifications, an analogue of Theorem 7 can be verified easily:

- Samson has to play an empty round, if $a = \circledast$ and u, v are ω-words or if $a = \textcircled{\omega}$ and u, v are finite words.

- In a split round, if $u, v \in \Sigma^\omega$, players can split each of them into the left concatenation of a finite word and an ω-word, i.e., u_2, v_2 are allowed to be ω-words.
- If $p = \text{Ⓦ}q$, and $u, v \in \Sigma^\omega$, Samson can choose to play an ω-split round in which:

  ω-split round: Samson ω-splits u into $u = u_1 \cdot u_2 \cdot \ldots$ and then Delilah ω-splits v into $v = v_1 \cdot v_2 \cdot \ldots$. Next Samson chooses some $j \geq 1$ and then Delilah responds with some $i \geq 1$. Configuration of the game is updated to (u_i, v_j). Samson can also choose to play such a round with u, v interchanged.

Theorem 11 *For every ω-language $\mathcal{L}$ and class $\mathcal{C}^\omega(p)$, the following are equivalent:*

(i) *No expression in $\mathcal{C}^\omega(p)$ represents $\mathcal{L}$.*

(ii) *There exist $u \in \mathcal{L}, v \notin \mathcal{L}$ such that Delilah has a winning strategy for $\mathcal{G}_p^\omega$ over (u, v).*

6 The Omega Power Problem

6.1 The Problem

We propose to consider the following problem, which we will solve in the next subsection:

Omega Power Problem: Is every ω-regular language representable by some ω-regular expression in the set $\mathcal{E}_\omega = \bigcup_{n\geq 0} \mathcal{C}^\omega(\odot^n \text{Ⓦ} \odot^n)$?

In other words, this problem asks if omega powers, together with boolean combinations and concatenations (including left concatenations), but without stars, are sufficient to express the class of ω-regular languages. This problem itself is interesting. Previously we don't really have a good understanding of the power of omega power. On the other hand, this problem is also somewhat related to the star height 2 problem.

Recall that the star height 2 problem asks if every regular language is representable by some expression in $\bigcup_{n\geq 0} \mathcal{C}(\odot^n \circledast \odot^n)$. So by definition these two problems are very similar. A negative solution to the omega power problem would also follow the similar steps as described at the beginning of Section 4. Also, the rules for playing the $*$-split rounds and the ω-split rounds are quite similar. So there can be some common strategies for playing the games.

Some readers might think that the omega power problem is easy to solve unlike the star height 2 problem. They might consider proving that ω-languages like $(aa)^* b^\omega$ are not representable in $\mathcal{E}_\omega$. Apparently, $(aa)^*$ is not equivalent to any expression without a star and so the use of star seems to be inevitable.

However, this is not true. As illustrated in Section 5, $(aa)^*b^\omega$ is in fact representable in $\mathcal{E}_\omega$ by clever use of negations. In general, the class $\mathcal{E}_\omega$ is very expressive because of the existence of negations. This is similar to that the class of expressions of star height ≤ 1 is surprisingly expressive.

There is another interesting way to look at these two problems. Finite words are words with both left and right ends, while ω-words are words without right ends. So loosely speaking, the omega power problem can be seen as a "right ends removed" version of the star height 2 problem. By such "removal" of right ends, we obtain a simpler and managable problem because abilities like modulus counting [8] of regular expressions are weakened.

Of couse, a solution to the omega power problem would rely on some properties that only ω-languages have. But we might still learn strategies for playing the game from such a solution, which can be useful for attacking the star height 2 problem.

6.2 The Proof

In this subsection, we settle the omega power problem negatively.

Let $n \geq 0$ be fixed. Similar to the star height 2 problem, it suffices to construct ω-words α_n, β_n distinguished by some fixed ω-regular language with $\alpha_n \sim_{n,\omega,n} \beta_n$. Here notations $u \sim_n v, u \sim_{\omega,n} v$ and $u \sim_{m,\omega,n} v$ are introduced like in Section 4, with the associated games called (n)-game, (ω, n)-game and (m, ω, n)-game respectively. Note that the $\sim_n$ relation here extends the $\sim_n$ relation in Section 4 by allowing u, v to be ω-words.

We focus on the (ω, n)-game over (α, β) with α, β in the form $a^L b(a^{\geq P_n} b)^\omega$ for some fixed L. Here $P_n = Q_n + 3T_n$ where $T_n = 2^{n+1}$ as before and Q_n is from the following lemma which can be easily proved using elementary number theory.

Lemma 12 *For each $n \geq 0$, there is an integer Q_n such that for each subset $\{m_1, \ldots, m_q\}$ of $\{1, 2, \ldots, T_n - 1\}$ with greatest common divisor d, for each r a multiple of d, if $r \geq Q_n$, then r is equal to a sum $\Sigma_{i=1}^q k_i m_i$ with each k_i a nonnegative integer.*

If Samson does not start with an ω-split round, clearly α, β are $\sim_n$ equivalent and Delilah can win. So w.l.o.g. we assume that Samson first makes an ω-split $U : u_1 \cdot u_2 \cdot \ldots$ of α.

We classify the words in U into types. For a word in U in the form $a^l b(\tilde{a}b)^m a^r$, we say it is of type $\overline{t_l t_m t_r}$, where t_l (or t_m, t_r, respectively) is 1 if l (or m, r, respectively) $\geq T_n$, and 0 otherwise. A word in U in the form a^m is of type $\overline{1}$ if $m \geq T_n$ and $\overline{0}$ otherwise. All types are exhausted in this list: $\overline{0}, \overline{000}, \overline{0?1}, \overline{010}, \overline{1}, \overline{1??}$, where ? means either 0 or 1. We use I_U to denote the smallest i such that u_i is not a $\overline{0}$ word.

Normal ω-Splits We say U is *normal* if it contains no word of type $\overline{1}$ or $\overline{1??}$ and it contains at least one $\overline{0}$ word. If U's normal, U's *characteristic* is the greatest common divisor of the lengths of U's $\overline{0}$ words. We first prove that if U is not normal, then Delilah can win the remaining (n)-game, and so we can assume that U is always normal.

Suppose U is not normal. There are three cases: (1) Some $\overline{1??}$ word in the form $\tilde{a}b(\tilde{a}b)^m a^r$ is in U. $u_1 \dots u_{I_U}$ must be in the form $a^{t_1}, \dots, a^{t_q}, a^{l_1}b(\tilde{a}b)^{m_1}a^{r_1}$ with $q \geq 0$. Then Delilah can win by ω-splitting β into the form $a^{t_1} \cdot \dots \cdot a^{t_q} \cdot a^{l_1}b(\tilde{a}b)^{m_1}a^{\min(r_1,T_n)} \cdot \tilde{a}b(\tilde{a}b)^m a^{\min(r,T_n)} \cdot \tilde{a}b(\tilde{a}b)^m a^{\min(r,T_n)} \cdot \dots$. (2) Some $\overline{1}$ word is in U. Then Delilah can win by ω-splitting β into the form $a^{t_1} \cdot \dots \cdot a^{t_q} \cdot a^{l_1}b(\tilde{a}b)^{m_1}a^{\min(r_1,T_n)} \cdot \tilde{a} \cdot a^{\min(l_1,T_n)}b(\tilde{a}b)^{m_1}a^{\min(r_1,T_n)} \cdot \tilde{a} \cdot \dots$. (3) U does not contain a $\overline{0}$ word. Assume (1) and (2) do not hold. Then as each a segment(except the first) of α has length $\geq P_n$ except the first, every u_i must be a $\overline{0?1}$ word. So u_1 is in the form $a^L b(a^{\geq P_n}b)^m\tilde{a}$ and Delilah can win by ω-splitting β into words all in such form.

Jumping Automata After Samson makes a normal ω-split U of α, if Delilah can win by an ω-split V of β such that each word in V is $\sim_n$ equivalent to some $\overline{0}$ or $\overline{000}$ word in U, we say that Delilah has a *fine win*. When can Delilah have a fine win? It turns out that if L is large enough, then this can be decided by a jumping automaton induced from U. Such idea is made precise in Lemma 15.

Definition 13 *For each normal ω-split U with characteristic Z, its associated (nondeterministic)* jumping automaton $\mathcal{A}$ *is the tuple $(\Sigma_Z, S, s_0, \Delta)$ with alphabet $\Sigma_Z = \{0, 1, \dots, Z-1\}$, state set $S = \Sigma_Z \times \{0, 1, \dots, T_n - 1\}$, initial state $s_0 = \langle 0, 0\rangle$ and $\Delta \subseteq S \times \Sigma_Z \times S$ the transition relation such that: for every $\overline{000}$ word in U, say in the form $a^l b(\tilde{a}b)^m a^r$, Δ contains:*

- $\langle\langle r', 0\rangle, (r' + l) \bmod Z, \langle r \bmod Z, m\rangle\rangle$ *for every* $0 \leq r' < Z$
- $\langle\langle r \bmod Z, t + 1\rangle, k, \langle r \bmod Z, t\rangle\rangle$ *for every* $k \in \Sigma_Z$ *and* $0 \leq t < m$.

For a finite word $k_0k_1 \dots k_{l-1}$ of length l or an ω-word $k_0k_1 \dots$ of length $l = \infty$, a *run* of $\mathcal{A}$ over the word is a state sequence $q_0q_1 \dots q_l \in S^*$ or an infinite state sequence $q_0q_1 \dots \in S^\omega$, respectively, such that $q_0 = s_0$ and $\langle q_i, k_i, q_{i+1}\rangle \in \Delta$ for all $0 \leq i < l$.

Definition 14 *For a word $u = a^{m_0}b \dots ba^{m_{t-1}}ba^{m_t}$ over Σ, the Z-*signature *of u is the word over Σ_Z: $sig_Z(u) = (m_0 \bmod Z)(m_1 \bmod Z) \dots (m_{t-1} \bmod Z)$. $sig_Z(u)$ for u an ω-word is defined similarly. Conversely, for a word $k_0 \dots k_{l-1}$ over Σ_Z, $wrd_Z(k_0 \dots k_{l-1})$ is an arbitrary word w in the form $(a^{\geq P_n}b)^l$ such that $sig_Z(w) = k_0 \dots k_{l-1}$.*

Lemma 15 *In an (ω, n)-game over (α, β), with α, β in the form $a^{\geq P_n - T_n}b(a^{\geq P_n}b)^\omega$, if $U : u_1 \cdot u_2 \cdot \dots$ is a normal ω-split of α of characteristic Z by Samson and $\mathcal{A}$ is the associated jumping automaton, then Delilah has a fine win iff $\mathcal{A}$ has a run over $sig_Z(\beta)$.*

Proof. *Only If.* Suppose Delilah has a fine win, then β can be ω-split into the form (1) below such that for each i, U has a $\overline{000}$ word in the form $a^{l_i}b(\tilde{a}b)^{m_i}a^{r_i}$. The Z-signature of β is as in (2), and a run of $\mathcal{A}$ over $sig_Z(\beta)$ can be constructed correspondingly as in (3). Here we write $p \xrightarrow{k} q$ to denote that state p goes to q by reading k, and $\langle r, m\rangle \to \langle r, 0\rangle$ is a short hand for the state sequence $\langle r, m\rangle, \langle r, m-1\rangle, \ldots, \langle r, 0\rangle$.

(1) $\beta : (a^Z)^* \cdot a^{l_1}b(\tilde{a}b)^{m_1}a^{r_1} \cdot (a^Z)^* \cdot a^{l_2}b(\tilde{a}b)^{m_2}a^{r_2} \cdot (a^Z)^* \cdot \ldots$

(2) $sig_Z(\beta)$: $l_1 \bmod Z, \ldots, (r_1 + l_2) \bmod Z, \ldots, (r_2 + l_3) \bmod Z, \ldots$

(3) ρ: $\langle 0, 0\rangle \xrightarrow{l_1 \bmod Z} \langle r_1, m_1\rangle \longrightarrow \langle r_1, 0\rangle \xrightarrow{(r_1+l_2) \bmod Z} \langle r_2, m_2\rangle \longrightarrow \langle r_2, 0\rangle \ldots$

If. Suppose there is a run ρ of $\mathcal{A}$ in the form of (3) over $sig_Z(\beta)$, then by definition of jumping automata one can conversely ω-split β into the form of (1) such that for each i, U has a $\overline{000}$ word in the form $a^{l_i}b(\tilde{a}b)^{m_i}a^{r_i}$. As to the words in the form $(a^Z)^*$, each of them has length $\geq Q_n$ and thus by Lemma 12 can be further split into $\overline{0}$ words from U. So Delilah has a fine win. ■

Let J_U denote the smallest i such that u_i is of type $\overline{010}$ or $\overline{0?1}$. J_U is ∞ if no such i exists. The *fine part* of α with respect to U is the prefix $u_1 \ldots u_{J_U-1}$, or simply α itself if $J_U = \infty$. Note that α's fine part has been split (or ω-split) into $\overline{0}$ and $\overline{000}$ words of U. An argument similar to the Only If direction of the above proof can be applied to show that $\mathcal{A}$ has a run over the Z-signature of α's fine part.

Winning the (ω, n)-Game We apply jumping automata to construct $\sim_{\omega,n}$ equivalent words. Roughly speaking, for an (ω, n)-game over (α, β) in which Samson makes a normal ω-split of α first, we show that if α has a sufficiently complex subword contained in the fine part of α, then the associated jumping automaton, which has a run over the signature of the fine part of α, would be confused, and then has a run over the signature of β. Then Delilah can win.

For a jumping automaton $\mathcal{A}$ associated to some normal ω-split U with characteristic Z, and a Z-signature $j_1 \ldots j_p$, we say that $w \in \{a, b\}^*$ has the *all-or-none* property with respect to $\mathcal{A}$ and $j_1 \ldots j_p$, if either $\mathcal{A}$ has no run over $j_1 \ldots j_p sig_Z(w)$ or for every infinite Z-signature in the form $j_1 \ldots j_p sig_Z(w) k_1 k_2 \ldots$, $\mathcal{A}$ has a run.

Lemma 16 *There exists w_n such that: (1) for every normal ω-split U, say with characteristic Z and associated jumping automaton $\mathcal{A}$, for every Z-signature $j_1 \ldots j_p$, w_n satisfies the all-or-none property w.r.t. $\mathcal{A}, j_1 \ldots j_p$. (2) w_n is in the form $(a^{\geq P_n}ba^{\geq P_n}b)^{\geq T_{2n}}$.*

Proof. First one can easily verify that if w has the all-or-none property w.r.t. $\mathcal{A}, j_1 \ldots j_p$, then for every $v \in \Sigma^*$, wv also has the all-or-none property w.r.t. $\mathcal{A}, j_1 \ldots j_p$. Initially we set $w_n = \epsilon$, we will gradually append words to obtain the desired w_n.

Let $\mathcal{A}$ be an jumping automaton associated to some normal ω-split U with characteristic Z. Jumping automata are in fact a special kind of looping automata [24], which can be determinized using subset construction. So we assume here that $\mathcal{A}$ is deterministic. For a state s of $\mathcal{A}$, we say that w has the all-or-none property w.r.t. $\mathcal{A}$ and s, if either $\mathcal{A}$ has no run starting from s over $sig_Z(w)$, or for every Z-signature in the form $sig_Z(w)k_1k_2\ldots$, $\mathcal{A}$ has a run from s over it.

If w_n does not satisfy the all-or-none property w.r.t. $\mathcal{A}, s$, then there is no run of $\mathcal{A}$ from s over some $sig_Z(w_n)k_1k_2\ldots$. Thus $\mathcal{A}$ has no run from s over $sig_Z(w_n)k_1\ldots k_q$ for some $q \geq 0$. Set w_n to be $w_n wrd_Z(k_1\ldots k_q)$ and then w_n satisfies the all-or-none property w.r.t. $\mathcal{A}, s$. Repeat this for every state s of $\mathcal{A}$. Now for every Z-signature $j_1\ldots j_p$, let s be the state $\mathcal{A}$ reaches after reading $j_1\ldots j_p$ (the case that $\mathcal{A}$ has no run over $j_1\ldots j_p$ is trivial). Then the all-or-none property of w_n w.r.t. $\mathcal{A}, j_1\ldots j_p$ follows from its all-or-none property w.r.t. $\mathcal{A}, s$.

Repeat the above for every jumping automaton $\mathcal{A}$ associated to some normal ω-split and then (1) is satisfied. Note that the characteristic of a normal ω-split is bounded by T_n and so there are only finitely many such automata.

As to condition (2), we can simply append enough $wrd_2(0)$'s to w_n to make it into the desired form. ■

We say an ω-word γ in the form $(a^{\geq P_n}b)^\omega$ has the *richness* property if for all x, Z with $0 \leq x < Z < T_n$, there are infinitely many subwords of γ in the form $b(a^Z)^*a^xb$.

Lemma 17 *For every v in the form $a^Lb(a^{\geq P_n}b)^*$, and $\gamma_1, \gamma_2 \in (a^{\geq P_n}b)^\omega$ having the richness property, $vw'_n\gamma_1 \sim_{\omega,n} vw'_n\gamma_2$, where $w'_n = wrd_2(0)^{T_n}w_n$.*

Proof. Let $\alpha = vw'_n\gamma_1$ and $\beta = vw'_n\gamma_2$. As before, it suffices to consider the (ω, n)-game over (α, β) in which Samson first makes a normal ω-split U of α with characteristic Z.

Case 1: vw'_n is a prefix of the fine part of α. Note that vw'_n contains $u_1\ldots u_{I_U}$ as prefix. So u_{I_U} is a $\overline{000}$ word. Let α' and β' be such that $\alpha = u_1\ldots u_{I_U}\alpha', \beta = u_1\ldots u_{I_U}\beta'$. So α', β' are both in the form $a^{\geq P_n - T_n}b(a^{\geq P_n}b)^\omega$. Note that $U' : u_{I_U+1}\cdot u_{I_U+2}\cdot\ldots$ is a normal ω-split of α' and so $\mathcal{A}$ has a run over the Z-signature of the fine part of α' w.r.t. U'. Note that $vwrd_2(0)^{T_n}$ contains $u_1\ldots u_{I_U}$ as prefix, $\alpha' = u'w_n\ldots$ for some u'. So $u'w_n$ is a prefix of the fine part of α' and $\mathcal{A}$ also has a run over $sig_Z(u'w_n)$. Note that $\beta' = u'w_n\ldots$ and thus by the all-or-none property of w_n, $\mathcal{A}$ has a run over $sig_Z(\beta')$. By Lemma 15, β' can be ω-split into words $\sim_n$ equivalent to ones from U'. Together with $u_1\cdot\ldots\cdot u_{I_U}$, these constitute a winning ω-split of β for Delilah.

Case 2: the fine part of α is a proper prefix of xw'_n. Let u' be such that $u_1\ldots u_{J_U-1}u' = xw'_n$. One can show that Delilah can win by splitting the $u_1\ldots u_{J_U-1}$ part of β as Samson did and ω-splitting the rest part into words $\sim_n$ equivalent to u_{J_U} and $\overline{0}$ words from U. There are two subcases. (1) u_{J_U} is a $\overline{0?1}$ word in the form $a^lb(\tilde{a}b)^m\tilde{a}$. $u'\gamma_2$ is in the form $a^lb(\tilde{a}b)^\omega$, and then can be ω-split into the form $a^lb(\tilde{a}b)^m\tilde{a}\cdot a^lb(\tilde{a}b)^m\tilde{a}\cdot\ldots$. Together with $u_1\cdot\ldots\cdot u_{J_U-1}$, these constitute a winning ω-split of $vw'_n\gamma_2$ for Delilah. (2) u_{J_U} is a $\overline{010}$ word in the form $a^l b(\tilde{a}b)^{\geq T_n} a^r$. By the richness property of γ_2, γ_2 can be ω-split into the form $(\tilde{a}b)^{\geq T_n} ba^r \cdot (a^Z)^* \cdot a^l b(\tilde{a}b)^{\geq T_n} a^r \cdot (a^Z)^* \cdot a^l b(\tilde{a}b)^{\geq T_n} a^r \cdot \ldots$, then $u'\gamma_2$ can be ω-split into words in the form $a^l b(\tilde{a}b)^{\geq T_n} a^r$, which are $\sim_n$ equivalent to u_{J_U}, and words in the form $(a^Z)^*$. As each such $(a^Z)^*$ word has length $\geq Q_n$, by Lemma 12, it can be further split into $\overline{0}$ words from U. Together with $u_1 \cdot \ldots \cdot u_{J_U - 1}$, these constitute a winning ω-split of $vw_n\gamma_2$ for Delilah. ■

Finally we turn to play the (n, ω, n)-game, and complete our proof.

Lemma 18 *Let γ have the richness property, then for all $x, y \geq 3^m$ and $u \in \Sigma^*$, $\alpha = u(wrd_2(0)w'_n)^x\gamma$ and $\beta = u(wrd_2(0)w'_n)^y\gamma$ are $\sim_{m,\omega,n}$ equivalent.*

Proof. We prove by induction on m. The case $m = 0$ follows from Lemma 17. First suppose that Samson splits α into $u(wrd_2(0)w'_n)^l v \cdot v'(wrd_2(0)w'_n)^r\gamma$ for some $l + r + 1 = x$ and $vv' = wrd_2(0)w'_n$. If $l < r$, then Delilah splits β into $u(wrd_2(0)w'_n)^l v \cdot v'(wrd_2(0)w'_n)^{y-l-1}\gamma$. Both $r \geq 3^{m-1}, y - l - 1 \geq 3^{m-1}$ and by induction hypothesis Delilah can win the remaining $(m-1, \omega, n)$-game. If $l \geq r$, then Delilah splits β into $u(wrd_2(0)w'_n)^{y-r-1} v \cdot v'(wrd_2(0)w'_n)^r\gamma$. Since the players are not allowed to ω-split finite words, the remaining $(m-1, \omega, n)$-game over $(u(wrd_2(0)w'_n)^l v, u(wrd_2(0)w'_n)^{y-r-1} v)$ is in fact an $(m - 1 + n)$-game. Recall w_n is in the form $(\tilde{a}b)^{\geq T_{2n}}$. Both words are in the form $u(\tilde{a}b)^{\geq T_{m-1+n}} v$ and Delilah can win the remaining $(m-1+n)$-game. For the cases that Samson splits α at the u part or the γ part, one can easily show that Delilah can win by a split at the similar position of β. ■

We say $\gamma \in \Sigma^\omega$ in the form $(a^{\geq P_n} b)^\omega$ has the *odd-odd* property if for all $k \geq 1$, the k-th a segment is of odd length iff k is odd. It is easy to construct γ_n which satisfies both the richness property and the odd-odd property. Define $\alpha_n = (wrd_2(0)w'_n)^{3^n+1}\gamma_n$ and $\beta_n = (wrd_2(0)w'_n)^{3^n}\gamma_n$, which are $\sim_{n,\omega,n}$ equivalent by Lemma 18. Recall that w_n (and w'_n) has an even number of b's, one can verify that α_n and β_n are distinguished by the ω-language $(a^*ba^*b)^*((aa)^*b(aa)^*ab)^\omega$. Thus we have completed the proof.

Theorem 19 *There exists an ω-regular language $\mathcal{L}$ such that no ω-regular expression in class $\bigcup_{n\geq 0} \mathcal{C}^\omega(\odot^n \circledcirc \odot^n)$ represents $\mathcal{L}$.*

Acknowledgment

I thank Enshao Shen for his kind guidance and suggestions, and thank Jean-Eric Pin and Wolfgang Thomas for their useful comments.

References

[1] J. Brzozowski, Open problems about regular languages, in: R. V. Book (Ed.), Formal Language Theory, New York: Academic Press, 1980, pp. 23–47.

[2] L. Eggan, Transition graphs and star height of regular events, Michigan Math. J. 10 (4) (1963) 385–397.

[3] K. Hashiguchi, Algorithms for determining relative star height and star height, Inf. Comput. 78 (2) (1988) 124–169.

[4] M. Schützenberger, On finite monoids having only trivial subgroups, Inf. Control 8 (1965) 190–194.

[5] R. McNaughton, S. Papert, Counter-free Automata, MIT Press, Cambridge, USA, 1971.

[6] J. Pin, H. Straubing, D. Thérien, Some results on the generalized star-height problem, Inf. Comput. 101 (2) (1992) 219–250.

[7] W. Henneman, Algebraic theory of automata, Ph.D. thesis, MIT (1971).

[8] W. Thomas, Remark on the star-height-problem, Theor. Comp. Sci. 13 (2) (1981) 231–237.

[9] J. Robson, More languages of generalised star height 1, Theor. Comp. Sci. 106 (1992) 327–335.

[10] R. Cohen, J. Brzozowski, Dot-depth of star-free events, J. Comput. Syst. Sci. 5 (1) (1971) 1–16.

[11] W. Thomas, Classifying regular events in symbolic logic, J. Comput. Syst. Sci. 25 (3) (1982) 360–376.

[12] H. Ebbinghaus, J. Flum, W. Thomas, Mathematical Logic, 2nd Edition, Undergraduate Texts in Mathematics, 1994.

[13] W. Thomas, An application of the Ehrenfeucht–Fraïssé game in formal language theory, Bull. Soc. Math. France, Mem. 16 (1984) 11–21.

[14] W. Thomas, A concatenation game and the dot-depth hierarchy, in: E. Börger (Ed.), Computation Theory and Logic, Vol. 270 of LNCS, 1987, pp. 415–426.

[15] J. Brzozowski, R. Knast, The dot-depth hierarchy of star-free languages is infinite, J. Comput. Syst. Sci. 16 (1) (1978) 37–55.

[16] Y. Bargury, J. Makowsky, The expressive power of transitive closure and 2-way multihead automata, in: 5th Workshop on Computer Science Logic, Vol. 626 of LNCS, 1991, pp. 1–14.

[17] E. Grädel, On transitive closure logic, in: CSL '91: 5th Workshop on Computer Science Logic, Vol. 626 of LNCS, 1991, pp. 149–163.

[18] A. Calò, J. Makowsky, The Ehrenfeucht–Fraïssé games for transitive closure, in: Logical Foundations of Computer Sciences, Tver '92, LNCS, 1992, pp. 57–68.

[19] A. Potthoff, Logische klassifizierung regulärer baumsprachen, Ph.D. thesis, Institut für Informatik u. Prakt. Math., Universität Kiel (1994).

[20] E. Shen, Q. Tian, Monadic partition logics and finite automata, Theor. Comp. Sci. 166 (1-2) (1996) 63–81.

[21] W. Thomas, Automata on infinite objects, in: J. van Leeuwen (Ed.), Handbook of Theoretical Computer Science, Vol. B, Formal models and semantics, Elsevier, 1990, pp. 133–191.

[22] W. Thomas, Languages, automata and logic, in: A. Salomaa, G. Rozenberg (Eds.), Handbook of Formal Languages, Vol. 3, Beyond Words, Springer-Verlag, Berlin, 1997.

[23] D. Perrin, J.E. Pin, First order logic and star free sets, J. Comput. Syst. Sci. 32 (1986) 393–406.

[24] M. Vardi, P. Wolper, Reasoning about infinite computations, Inf. Comput. 115 (1) (1994) 1–37.