Chapter 2
Introduction

The fundamental results of Büchi [Büc62] and Rabin [Rab69] state that the monadic second-order (MSO) theory of the ω-chain $(\omega, \leqslant)$ and of the complete binary tree $\left(\{0, 1\}^*, \preceq, \leqslant_{\text{lex}}\right)$ is decidable. In both cases the proof relies on a class of finite automata with expressive power equivalent to MSO. Because of effective closure properties and decidability of the emptiness problem, the languages of ω-words and infinite trees definable in MSO are called *regular*. For a broad introduction to the field of regular languages of infinite objects see [Tho96, PP04, TL93].

Since a single ω-word or infinite tree may not have any finite representation, one has to deal with *actual infinity* when studying languages of such objects. In particular, even the set of ω-words over a two-letter alphabet has cardinality continuum. This is the source of strong relationships between properties of regular languages of infinite objects and descriptive set theory. These relationships have a form of synergy: descriptive set theory motivates new problems and methods in automata theory but on the other hand, automata theory introduces natural examples for classical topological concepts.

Recently there has been a number of papers studying these relationships. Properties of regular languages of infinite trees have been studied in [NW03, AN07, ADMN08, Mur08], the Borel complexity of MSO-definable sets of branches of one infinite tree was estimated in [BNR+10], finally the Borel and Wadge complexity of languages of ω-words recognised by various models of computation was estimated in [DFR01, Fin06, CDFM09, DFR13, FS14]. It is worth mentioning that in most of the above cases it turns out that there are languages definable in respective formalisms that are *complete* for the studied topological classes. It shows that these languages are in some sense *representative*. Also, there are some results studying more general set theoretic properties of definable languages. For instance, expressibility of cardinality of sets in MSO was studied in [BKR11], and measurability of regular languages of infinite trees was settled in [GMMS14].

The results of the thesis are based on [HS12, FMS13, BIS13, BS13, Skr14, BGMS14] and the technical report [MS14].

© Springer-Verlag Berlin Heidelberg 2016

M. Skrzypczak, *Set Theoretic Methods in Automata Theory*, LNCS 9802
DOI: 10.1007/978-3-662-52947-8_2

2.1 Motivations

The following list presents problems studied in the thesis. Most of them have the form of a question about *descriptive complexity* — given a regular language L, is there a description of L that is *simple* in a certain sense.

2.1.1 Definability in WMSO

The first question asks how to effectively decide if a given regular language is definable in some logic weaker than MSO. There are two natural candidates for such logics: first-order logic (FO) and weak monadic second-order logic (WMSO) where the set quantification is restricted to finite sets.

In the case of ω-words, definability in FO was solved by Thomas [Tho79] using the methods of Schutzenberger [Sch65] and McNaughton Papert [MP71]. The definability in WMSO trivialises in this case, since every ω-regular language is WMSO-definable.

The problem of definability in WMSO for regular languages of infinite trees is considered as one of the central problems in the area. Recently, there has been some slight progress for various restricted classes of languages. However, the problem in its full generality seems to be out of reach of the currently known methods.

The thesis presents solutions to the problem of WMSO-definability for certain restricted classes of regular languages of infinite trees: for unambiguous Büchi automata in Chap. 3, for general Büchi automata in Chap. 4, for game automata in Chap. 5, and for languages of thin trees in Chap. 6.

2.1.2 Index Problem

Another *complexity question* studied in the thesis asks about the *index* of a given regular language of infinite trees L: for a given pair (i, j) is there an alternating top-down parity tree automaton that recognises L and uses only priorities among $\{i, i+1, \ldots, j\}$? It turns out that in the case of languages of infinite trees that are bisimulation-invariant (i.e. definable in μ-calculus, see [JW96]), the index corresponds precisely to the alternation of fixpoints used in the definition of a language [Niw97]. Therefore, the index problem can be seen as a variant of a quantifier alternation question: how many alternations of quantifiers are needed to define a given language.

The decidability of the index problem for general languages of infinite trees is open. As shown in [Bra98, Arn99], the index hierarchy is strict — there are regular languages of infinite trees that cannot be recognised by any automaton of small index. As shown by Rabin [Rab70], the index problem and definability in WMSO are closely related: a regular language of infinite trees is definable in WMSO if and only if both

the language and the complement are recognisable by an alternating automaton with Büchi acceptance condition (i.e. condition of the form "infinitely many *accepting* states").

The thesis provides a solution of the index problem for the class of regular languages of infinite trees recognisable by game automata (see Chap. 5). This is the first reasonable class of languages for which the index problem is known to be decidable, that contains languages arbitrarily high in the alternating index hierarchy. Additionally, an effective collapse of index for languages recognisable by unambiguous automata is provided in Chap. 3: it is proved that if an automaton is unambiguous and of certain index then the language recognised by the automaton is lower in the index hierarchy. Although the presented collapse is small, to the author's best knowledge this is the first result that utilizes the fact that a given automaton is unambiguous to give upper bounds on the index of the recognised language.

2.1.3 Bi-unambiguous Languages

One of the difficulties when working with MSO on infinite trees arises from the fact that deterministic automata are too weak to recognise all regular languages. The subclass of regular languages of infinite trees recognisable by deterministic automata seems to be much more tractable [KSV96, NW98, NW03, NW05, Mur08]. Unambiguous automata can be seen as a natural class of automata in-between deterministic and non-deterministic ones. A non-deterministic automaton is *unambiguous* if it has at most one accepting run on every input. As shown by Niwiński and Walukiewicz [NW96], there are regular languages of infinite trees that are inherently ambiguous — there is no unambiguous automaton recognising them. Very little is known about unambiguous languages, for instance it is not known how to decide if a given regular language of infinite trees is recognisable by some unambiguous automaton.

The thesis characterizes the class of bi-unambiguous languages (i.e. languages L such that both L and the complement L^c are unambiguous) as those that can be recognised by finite *prophetic thin algebras*. This theorem constitutes a link between the algebraic framework for thin trees from [Idz12] and languages of general infinite trees. Also, it provides an algebraic way of recognition for a non-trivial class of regular languages of infinite trees.

The following new conjecture has arisen when studying properties of prophetic thin algebras.

Conjecture 2.1. The relation $\varphi'(x, Z)$ expressing that $x \in Z$ and Z is contained in a thin tree does not admit MSO-definable uniformization of the first variable x. In other words, there is no MSO-definable choice function in the class of thin trees.

This conjecture is a strengthening of the theorem of Gurevich and Shelah [GS83] stating that there is no MSO-definable choice function on the complete binary tree. Unfortunately, the conjecture is left open, however some equivalent statements are

provided. Also, it is shown that the conjecture implies that it is decidable if a given regular language of infinite trees is bi-unambiguous. Additionally, the conjecture implies that bi-unambiguous languages constitute a very reasonable class (a pseudo-variety from the algebraic point of view).

2.1.4 Borel Languages

The index hierarchy for automata on infinite trees turns out to be closely related to topological hierarchies from descriptive set theory (see for instance [Arn99]). These relations motivate a number of interesting questions, one of them is the following conjecture, stated over 20 years ago.

Conjecture 2.2 (Skurczyński [Sku93]). If a regular language of infinite trees is Borel then it is WMSO-definable.

The converse implication is known to be true: every WMSO-definable language is Borel. Therefore, the conjecture says in fact that a regular language of infinite trees is Borel if and only if it is WMSO-definable. It would mean that if a language is regular and topologically simple then it is also "descriptively" simple. It can also be seen as an automata theoretic counterpart of the relation between the lightface and boldface hierarchies, see [Mos80, Theorem 3E.4].

The conjecture has been proved only in the special case of deterministic languages [NW03]. The thesis provides proofs of the conjecture for wider classes of languages: recognisable by game automata in Chap. 5 and for languages of thin trees in Chap. 6. Additionally, a potential strategy of proving the conjecture for Büchi automata is presented in Chap. 4, unfortunately some additional pumping argument is missing in that case.

2.1.5 Topological Complexity vs. Decidability

In general, there is no direct relationship between decidability of a logic and topological complexity of languages it defines. For instance, the FO theory of the structure of arithmetic $(\omega, \leqslant, +, *)$ is undecidable, while it defines only Borel languages of ω-words. On the other hand one can construct a trivial logic that defines some particular language of very high topological complexity. However, as observed by Shelah [She75] (see also [GS82]) in the case of MSO, the topological complexity and decidability are strongly related: the MSO theory of $(\mathbb{R}, \leqslant)$ is undecidable, however, by Rabin's theorem [Rab69], the theory becomes decidable if we restrict the set quantification to $\mathbf{\Sigma}^0_2$-sets.

These ideas are used in Chaps. 9 and 10 to study decidability of MSO logic equipped with an additional quantifier U (as introduced by Bojańczyk [Boj04] and denoted

MSO+U). Chapter 9 studies topological complexity of languages of ω-words definable in MSO+U. It is shown that the topological complexity of these languages is as high as possible: examples of languages lying arbitrarily high in the projective hierarchy are given. Already this fact implies that there is no *simple* automata model capturing the expressive power of MSO+U on ω-words.

This topological observation is further developed in Chap. 10 to prove that a certain variant of MSO on the Cantor set $\{\mathtt{L}, \mathtt{R}\}^\omega$ (called proj-MSO) can be reduced to the MSO+U theory of the complete binary tree. As shown in [BGMS14], the proj-MSO theory is not decidable in the standard sense (see Theorem 10.88). Therefore, the presented reduction shows that MSO+U is also not decidable in this sense.

The question of decidability of MSO+U on the infinite trees was posed in [Boj04]. The above line of research proves that this question cannot be answered positively. Somehow surprisingly, the technical hearth of the proof relies on purely topological concepts.

2.1.6 Separation Property

The question of separation asks if it is possible to separate every pair of disjoint languages from some class by a *simple* language. A classical example of such property is the following theorem of Lusin: every pair of disjoint analytic (i.e. $\mathbf{\Sigma}^1_1$) sets can be separated by a Borel set.

The separation property has also been studied for certain classes of regular languages, an example is the following result of Rabin: every pair of disjoint regular languages of infinite trees recognisable by Büchi automata can be separated by a language that is WMSO-definable. Recently, the separation turned out to be crucial step in providing a significant result about the decidability of the *dot-depth* hierarchy, see [PZ14].

In Chap. 11 of the thesis the separation property is studied for certain quantitative extensions of ω-regular languages, namely for ωB- and ωS-regular languages introduced by Bojańczyk and Colcombet [BC06]. It is shown that the ωB- and ωS-regular languages have the separation property with respect to ω-regular languages: every pair of disjoint languages recognisable by ωB- (respectively ωS)-automata can be separated by an ω-regular language. This result is somehow surprising as the models of ωB- and ωS-automata are dual: a language is ωB-regular if and only if its complement is ωS-regular. Usually, exactly one class from a pair of dual classes of sets has the separation property.

2.2 Overview of the Parts

The preliminary Chap. 1 introduces basic notions and known results that will be used later. The rest of the thesis is divided into three parts, each part has three chapters.

All the presented results study related problems of *descriptive complexity*. The respective parts group results of similar type. Most of the chapters present results that are technically independent, in particular they can be read separately. The only technical dependencies are: Chaps. 7 and 8 depend on definitions from Chap. 6; results of Chap. 10 depend on Theorem 2.7 from Chap. 9.

A separate chapter (see page 205) presents conclusions of the whole thesis. In particular, some relationships and similarities between the techniques used in the chapters are discussed.

2.2.1 Part I: Subclasses of Regular Languages

The first part of the thesis studies descriptive complexity questions for restricted classes of regular languages of infinite trees: unambiguous automata in Chap. 3, Büchi automata in Chap. 4, and game automata in Chap. 5. Three main theorems of these chapters are the following.

The first theorem shows how to use the fact that a given automaton is unambiguous to derive a collapse in parity index of the language recognised by it.

Theorem 2.1. *If $\mathcal{A}$ is an unambiguous* min*-parity automaton of index $(0, j)$ then the language $L(\mathcal{A})$ can be recognised by an alternating* Comp$(0, j-1)$*-automaton of size polynomial in the size of $\mathcal{A}$.*

In particular, if $\mathcal{A}$ is Büchi and unambiguous then $L(\mathcal{A})$ is WMSO*-definable.*

The second theorem is based on a theory of certain ranks for Büchi automata. Using these ranks, a characterisation of WMSO-definable languages is given.

Theorem 2.2. *It is decidable if the language of infinite trees recognised by a given non-deterministic Büchi tree automaton is* WMSO*-definable.*

The above result was already proved by Kuperberg and Vanden Boom (see for instance [CKLV13]) using the theory of cost functions. However, as discussed in Chap. 4, the methods developed in the presented proof may be of independent interest since they introduce conceptually new techniques based on ranks of well-founded ω-trees.

Finally, the third theorem shows that both index problems are decidable for *game automata* — a class of alternating automata that extends deterministic ones by allowing certain restricted alternation between the players. Two effective procedures that compute the index of the language recognised by a given game automaton are proposed. Then it is shown that the procedures are correct. For this purpose, upper and lower bounds are given. Interestingly, in the case of the alternating index problem, the lower bounds are based on purely topological methods (namely the topological hardness of languages $W_{i,j}$).

Theorem 2.3. *The non-deterministic index problem is decidable for game automata (i.e. if a game automaton is given as the input). The same holds for the alternating index problem.*

2.2.2 Part II: Thin Algebras

The second part is devoted to a study of *thin algebras* and thin trees, i.e. trees having only countably many infinite branches. In Chap. 6 a characterization of languages of thin trees that are WMSO-definable among all infinite trees is given. Chapter 7 is devoted to the recognition of languages of infinite trees by *prophetic thin algebras*. Finally, Chap. 8 studies Conjecture 2.1 and related uniformization problems on thin trees. Three main theorems of these chapters are the following.

The first theorem gives an effective characterisation of regular languages of thin trees that are definable in WMSO among all infinite trees. Additionally, it expresses an upper bound: even if a regular language of thin trees is not WMSO-definable among all infinite trees, it is still topologically simple (i.e. it belongs to $\mathbf{\Pi}^1_1$).

Theorem 2.4. *A regular language of thin trees (i.e. a regular language that contains only thin trees) is either:*

1. $\mathbf{\Pi}^1_1$*-complete among all infinite trees,*
2. WMSO*-definable among all infinite trees (and thus Borel).*

Moreover, it is decidable which of the cases holds.

The second theorem provides an algebraic framework for recognition of a restricted class of regular languages of infinite trees. The idea is to use algebras designed for thin trees to recognise languages of arbitrary infinite trees.

Theorem 2.5. *A language of infinite trees L is recognised by a homomorphism into a finite prophetic thin algebra if and only if L is bi-unambiguous, i.e. both L and the complement L^c can be recognised by unambiguous automata.*

The last theorem consists of three ingredients: an equivalent formulation of Conjecture 2.1, an example of a non-uniformizable relation on thin trees, and an essentially new example of an ambiguous regular language of infinite trees. The non-uniformizable relation uses a concept of *skeleton* — a subset of a thin tree that provides a decomposition of this tree into separate branches.

Theorem 2.6. *Conjecture 2.1 is equivalent to the fact that every finite thin algebra admits some consistent marking on every infinite tree.*

The relation $\varphi(\sigma, t)$ stating that t is a thin tree and σ is a skeleton of t does not admit any MSO*-definable uniformization of σ.*

The language of all thin trees is ambiguous (i.e. it is not recognised by any unambiguous automaton).

Although Conjecture 2.1 is not proved in this thesis, the above non-uniformizability results are of their own interest. In particular, the example about skeletons provides a standalone answer to Rabin's uniformization problem (the problem was solved originally by Gurevich and Shelah in [GS83]).

2.2.3 Part III: Extensions of Regular Languages

The last part of the thesis studies some properties of contemporary *quantitative* developments in automata theory. Topological complexity of MSO+U-definable languages of ω-words is estimated in Chap. 9. Chapter 10 studies consequences of the high topological complexity of MSO+U regarding decidability of this logic on the complete binary tree. Finally, in Chap. 11 the separation property for ωB- and ωS-regular languages is proved. Three main theorems of these chapters are the following.

The first expresses the topological complexity of MSO+U on ω-words.

Theorem 2.7. *There exist languages of ω-words that are definable in* MSO+U *logic and lie arbitrarily high in the projective hierarchy.*

The second theorem uses studies a new variant of MSO (called *proj*-MSO). It is a logic introduced in [BGMS14] where set quantifiers are restricted to projective sets of certain level (fixed explicitly during quantification). For instance, a logic can say "there exists a set X that belongs to $\mathbf{\Sigma}^1_5$ and …".

Theorem 2.8. *The proj-*MSO *theory of $\{\mathtt{L}, \mathtt{R}\}^{\leqslant\omega}$ with prefix $\preceq$ and lexicographic $\leqslant_{\text{lex}}$ orders effectively reduces to the* MSO+U *theory of the complete binary tree $\big(\{\mathtt{L}, \mathtt{R}\}^*, \preceq, \leqslant_{\text{lex}}\big)$.*

*An algorithm deciding the proj-*MSO *theory of $\{\mathtt{L}, \mathtt{R}\}^{\leqslant\omega}$ (together with its proof of correctness) would imply that analytic determinacy fails.*

This result was further extended in [BGMS14] using an adaptation of the technique of Shelah [She75]. It is shown there that under a certain set theoretic assumption (namely that V=L, i.e. we work in the *Gödel's constructible universe*) the proj-MSO theory of $\{\mathtt{L}, \mathtt{R}\}^{\leqslant\omega}$ is undecidable. Therefore, together with the above theorem, V=L implies that the MSO+U theory of the complete binary tree is undecidable.

Finally, the ninth main theorem of the thesis studies *separation property* for languages of ω-words that are recognised by counter automata introduced by Bojańczyk and Colcombet in [BC06].

Theorem 2.9. *If L_1, L_2 are disjoint languages of ω-words both recognised by ωB-(respectively ωS)-automata then there exists an ω-regular language L_{sep} such that*

$$L_1 \subseteq L_{\text{sep}} \quad \text{and} \quad L_2 \subseteq L^{\text{c}}_{\text{sep}}.$$

Additionally, the construction of L_{sep} is effective.



http://www.springer.com/978-3-662-52946-1

Descriptive Set Theoretic Methods in Automata Theory
Decidability and Topological Complexity
Skrzypczak, M.
2016, XIII, 211 p. 18 illus., Softcover
ISBN: 978-3-662-52946-1