Expressiveness of Monadic Second-Order Logics on Infinite Trees of Arbitrary Branching Degree

MSc Thesis *(Afstudeerscriptie)*

written by

Fabio Zanasi
(born December 17th, 1988 in Modena, Italy)

under the supervision of **Dr Alessandro Facchini** and **Prof Dr Yde Venema**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

Msc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:
August 31st, 2012

Members of the Thesis Committee:
Prof Dr Johan van Benthem
Dr Alessandro Facchini
Dr Helle Hansen
Prof Dr Benedikt Löwe
Prof Dr Yde Venema

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

In this thesis we study the expressive power of variants of monadic second-order logic (*MSO*) on infinite trees by means of automata. In particular we are interested in weak *MSO* and well-founded *MSO*, where the second-order quantifiers range respectively over finite sets and over subsets of well-founded trees. On finitely branching trees, weak and well-founded *MSO* have the same expressive power and are both strictly weaker than *MSO*. The associated class of automata (called weak *MSO*-automata) is a restriction of the class characterizing *MSO*-expressivity.

We show that, on trees with arbitrary branching degree, weak *MSO*-automata characterize the expressive power of well-founded *MSO*, which turns out to be incomparable with weak *MSO*. Indeed, in this generalized setting, weak *MSO* gives an account of properties of the 'horizontal dimension' of trees, which cannot be described by means of *MSO* or well-founded *MSO* formulae.

In analogy with the result of Janin and Walukiewicz for *MSO* and the modal μ-calculus, this raises the issue of which modal logic captures the bisimulation-invariant fragment of well-founded *MSO* and weak *MSO*. We show that the alternation-free fragment of the modal μ-calculus and the bisimulation-invariant fragment of well-founded *MSO* have the same expressive power on trees of arbitrary branching degree. We motivate the conjecture that weak *MSO* modulo bisimulation collapses inside *MSO* and well-founded *MSO*.

Contents

Introduction

Monadic second-order logic (*MSO*) is an expressive specification language in which first-order logic is extended with quantification over sets. By adding a successor relation R to the language, path quantification, reachability and other properties of transition systems can be described in *MSO*. We are interested in a particular kind of transition system, namely trees without leaves. In the sequel, we use the name *tree* to refer to such infinite structures.

In the 60's, Rabin [27] proved the decidability of *MSO* on binary trees. This landmark result was obtained by an automata-theoretic characterization of the expressive power of *MSO* on these structures. The idea is to define a class of automata $\mathcal{C}$ such that, for each formula $\varphi \in MSO$, we can construct an automaton $\mathbb{A}_\varphi$ in $\mathcal{C}$ accepting exactly the binary trees in which φ is true. Viceversa, for each automaton $\mathbb{A}$ in $\mathcal{C}$ we can find a formula $\varphi_{\mathbb{A}} \in MSO$ that is true exactly in the binary trees that are accepted by $\mathbb{A}$.

Rabin's work became of direct interest for computer scientists one decade later, when it was realized that (infinite) trees can be used as models of the behavior of nonterminating systems [26]. In this framework, *MSO* plays the role of an assembly-like language into which most specification languages (such as temporal logics and the modal μ-calculus) can be compiled [6] [10]. The underlying automata theory associated with *MSO* has been developed consequently, extending Rabin's characterization result to more general classes of structures.

In the 90's, the work of Walukiewicz [33] [32] has provided a very general framework for investigating *MSO* by means of automata. In particular, in [33] a class of automata was introduced, which captures the expressive power of *MSO* on trees with arbitrary (also infinite) branching degree. We will present them under the name of *MSO-automata*.

Parallel to these developments, variants of *MSO* have also received attention. Among them, *weak monadic second-order logic* (*WMSO*) is a quite appealing specification language, being computationally more manageable than *MSO* [21], but as expressive as *MSO* on simple structures such as streams [19]. Syntax and semantics of *WMSO* are defined as for *MSO*, but for second-order quantifiers, which are restricted to range over finite sets only. An automata characterization of *WMSO* on binary trees has been proposed by Rabin [28], to show that *WMSO* is strictly less expressive than *MSO* on this class of structures. Automata for *WMSO* on binary trees have been further investigated in the '80s by Muller, Saoudi and Schupp [24], introducing the notion of *weak alternating tree automaton*.

On structures that are more general than binary trees, automata theory for *WMSO* is less settled. It is a folklore result that weak alternating tree automata can be suitably generalized to serve as a characterization of the expressive power of *WMSO* on finitely branching trees. We call *weak MSO-automata* the resulting class of automata, being essentially *MSO*-automata where further constraints have been imposed on the structure of each run. This automata characterization leads to the result that *WMSO* is weaker than *MSO* also on finitely branching trees.

In this thesis, we consider the question of how *MSO* and *WMSO* relate on a more general class of structures, namely trees of arbitrary branching degree. The motivation is given by a simple observation on *MSO*-automata, namely that they are not able to distinguish between trees with finite or infinite branching degree. This *Finite Branching Property* can be rephrased on the side of logic, by saying that each *MSO*-formula that is true in some tree is true in a finitely branching tree. As a consequence of that, the landscape of connections between *MSO* and *WMSO* on arbitrarily branching trees radically changes with respect to the case of finitely branching trees.

- We can define a *WMSO*-formula that is only true in trees that are not finitely branching, meaning that *WMSO* does not have the Finite Branching Property. It follows that *WMSO* is no more weaker than *MSO* on trees of arbitrary branching degree, but the two logics have *incomparable* expressive power.

- On the side of automata, weak *MSO*-automata happen to have the Finite Branching Property, being a restricted version of *MSO*-automata. It follows that, contrary to the case of finitely branching trees, weak *MSO*-automata cannot serve as a characterization of *WMSO* on the more general class of structures.

The outcome of this analysis is that the setting of finitely branching trees does not give the complete picture of *WMSO*-expressivity, on the side of logic, and of definability by weak *MSO*-automata, on the side of automata. In this thesis we will be mainly interested in investigating the second question, that is, the theory of weak *MSO*-automata on trees of arbitrary branching degree. The structure of our work can be outlined as follows.

- After preliminaries, in the second chapter we rephrase Walukiewicz's automata characterization of *MSO* on arbitrarily branching trees. Given a formula $\varphi \in MSO$, we show how an *MSO*-automaton equivalent to φ can be constructed, by induction on its syntactic shape. The flexibility of *MSO*-automata makes relatively easy to prove that the tree languages that they recognize are closed under union and complementation. The hard part is to show that they are closed under projection, corresponding to the case in which φ is of the form $\exists p.\psi$, with p a set-variable. For this purpose, we emphasize the role of the Simulation Theorem, which provides a normal form for *MSO*-automata. The construction which is involved in this result turns out to be an useful benchmark to understand the nature of *MSO*-expressivity.

- The third chapter considers the case of weak *MSO*-automata, defined as a restricted version of the automata introduced in the previous chapter. If *MSO*-automata are tailored to serve as a characterization of *MSO*, now we proceed in the converse direction, tailoring a logic that corresponds to weak *MSO*-automata. This is given as a variant of *MSO*, which we call *well-founded monadic second-order logic* (*WFMSO*). It is defined as *MSO* but for the semantics of second-order quantifiers, that are restricted to range over subsets of well-founded trees only. Analogously to the case of *MSO*, given a formula $\varphi \in WFMSO$, we show by induction how a weak *MSO*-automaton equivalent to φ can be constructed. Once again, the conceptual core of our argument is the case in which φ is of the form $\exists p.\psi$, with p a set-variable. By definition of *WFMSO*, the variable p does not range over arbitrary sets of nodes, but only on the subsets of well-founded trees. To obtain the characterization result, we provide a normal form theorem for weak *MSO*-automata, which is the 'weak' counterpart of the Simulation Theorem, now tailored to the case of *WFMSO*-quantification.

- In the fourth chapter we complete the correspondence between weak *MSO*-automata and *WFMSO*, by showing that each weak *MSO*-automaton is equivalent to some formula $\varphi \in WFMSO$. The proof of this result passes through the introduction of *non-deterministic Büchi automata* (*NDB* automata), which are somehow intermediate between weak *MSO*-automata and *MSO*-automata. We prove that for each tree language $\mathcal{L}$, if both $\mathcal{L}$ and its complement are recognized by *NDB* automata, then $\mathcal{L}$ is defined by some *WFMSO*-formula. This provides another automata characterization for *WFMSO*, in terms of *NDB* automata. Just as for the case of weak *MSO*-automata, also *NDB* automata have a counterpart working on binary trees, which has been introduced by Rabin to characterize *WMSO* on this restricted class of structures [28].

- The fifth chapter brings together all the work we did in the previous part to compare the expressive power of *MSO*, *WMSO* and *WFMSO* on different classes of structures. We argue that automata do not just give an account of these logics, but also reveal which kind of specifications is better expressed by one logic with respect to the others. For instance, we observe that *MSO* is stronger than *WFMSO* and *WMSO* in expressing properties of the *vertical dimension* of trees, such as 'each path has only finitely many nodes whose label includes p'. On the other hand, *WMSO* turns out to be more expressive than *MSO* and *WFMSO* on the *horizontal dimension* of trees, expressing properties such as 'being finitely branching'. This latter property in particular cannot be expressed by means of *MSO* or *WFMSO* formulae, as revealed by a careful analysis of the automata characterization provided for these two logics in the previous chapters. On the base of these observations, we state that, on arbitrarily branching trees, *WFMSO* and *WMSO* are respectively strictly weaker and incomparable with *MSO*. Next, we examine the question of how *WFMSO* and *WMSO* are related. Despite of the fact that they are the same logic on finitely branching trees, we show that they are incomparable on trees of arbitrary branching degree. This is in some sense a refinement of the incomparability result for *MSO* and *WMSO*, *WFMSO* being weaker than *MSO*. In particular, it will follow as a corollary of another characterization result, which we consider one of the main contributions of this thesis: the bisimulation-invariant fragment of *WFMSO* is as expressive as the alternation-free fragment of the modal μ-calculus.

- We also include an appendix, where a game-theoretical argument is supplied to prove that *MSO*-automata are closed under complementation.

Chapter 1

Preliminaries

In this section we introduce some of the preliminaries and fix the notation. We refer to [17], [13] and [5] respectively for the terminology of Set Theory, Model Theory and Order Theory.

1.1 Sets, Functions and Relations

Sets are usually indicated with capital Latin letters X, Y, Z, relations with capital Latin letters R, Q and functions with small Latin letters f, g and h.

Sets Let X be a set. We indicate with $\wp(X)$ the set of subsets of X. For any subset $Y \subseteq X$ of X, we denote with $X \setminus Y$ the set $\{x \in X \mid x \notin Y\}$. If Y is a *finite* subset of X then we write $Y \subseteq_{\omega} X$. If Y is *strictly included* in X, meaning that $X \setminus Y$ is non-empty, we write $Y \subsetneq X$. Given a set Z, we indicate with $X \times Z$ and $X \uplus Z$ respectively the cartesian product and the disjoint union of X and Z. For the set $X \times Z$ we have the usual projection functions $\pi_1 : X \times Z \to X$ and $\pi_2 : X \times Z \to Z$.

Functions The notation $f : X \to Y$ means that f is a function with domain X and codomain Y. We refer to $X \to Y$ as the *type* of f. The domain and codomain of f are also indicated respectively as $Dom(f)$ and $Cod(f)$. For any subset $Z \subseteq X$, the set $f[Z]$ is defined as $\{y \in Y \mid f(x) = y \text{ for some } x \in Z\}$ and we indicate with $f_{\restriction Z} : Z \to Y$ the restriction of the function f to Z. The image $f[X]$ of f on the whole domain X is also denoted with $Ran(f)$. For any singleton set $\{z\}$, we indicate with $f[z \mapsto y]$ the function with domain $X \cup \{z\}$ (where z may be in X) and codomain Y, which is given by

$$f[z \mapsto y](x) \quad := \quad \begin{cases} y & \text{If } x = z, \\ f(x) & \text{Otherwise.} \end{cases}$$

We say that f is *1-1* if $f(x) = f(y)$ implies that $x = y$, for all $x, y \in X$. We say that f is *onto* if $Ran(f) = Y$. The function $f : X \to Y$ is *bijective* if it is both 1-1 and onto. If f is 1-1, the *inverse* of f is the function $f^{-1} : Ran(f) \to X$ assigning to each $y \in Ran(f)$ the unique $x \in X$ such that $y = f(x)$. Given functions $g : X \to Y$ and $h : Y \to Z$, we denote as $h \circ g : X \to Z$ the composition of g and h. Given functions $f : X \to Y$ and $f' : X' \to Y'$, we say that f' *extends* f if X and Y are subsets respectively of X' and Y', and f is equal to $f'_{\restriction X}$. Functions are always assumed total when not specified otherwise.

Relations Let X and Y be sets. Given a binary relation $R \subseteq X \times Y$, we define $Dom(R) := \{\pi_1(x,y) \mid (x,y) \in R\}$ and $Ran(R) := \{\pi_2(x,y) \mid (x,y) \in R\}$. For any element $x \in X$, we indicate with $R[x]$ the set $\{y \in Y \mid (x,y) \in R\}$. The relations R^+ and $R^\star$ are defined respectively as the transitive closure of R and the reflexive and transitive closure of R.

Natural Numbers The cardinality of a set X is indicated with $|X|$. Following von Neumann's convention, we denote with ω the set of natural numbers with the usual order $\leq$ and we identify each natural number $i < \omega$ with the set $\{0, 1, 2, \ldots, i-1\}$. For any finite subset Y of ω, we denote with $Max(Y)$ and $Min(Y)$ respectively the largest and the smallest number occurring in Y. A function f whose domain is γ for some $\gamma \leq \omega$ and whose codomain is a set Z is called a *sequence in* Z. The standard notation for $Ran(f)$ is $(z_i)_{i<\gamma}$, where z_i indicates that $f(i) = z$, for each $i < \gamma$. Suppose that the set Z has some order $\preceq$. We say that a sequence $(z_i)_{i<\gamma}$ of elements of Z is *monotone in*

$\leq$ if $z_i \leq z_{i+1}$ for each $i < \gamma$. In some contexts we refer to an infinite sequence $(z_i)_{i<\omega}$ of elements of Z as a $Z-$*stream*. We denote with Z^{ω} the set of all $Z-$streams.

1.2 Trees

Convention 1.1. Throughout this thesis we let P be a fixed set of *propositional letters*, whose elements are denoted with small Latin letters p, q and r. We denote with C the set $\wp(P)$ of *labels* on P. An element of C is usually indicated with the letter c. ◂

Definition 1.2 (Labeled Transition System)**.** A *C-labeled transition system* is a tuple $\mathbb{S} = \langle T, R, V \rangle$, where T is a set, $R : T \times T$ is a binary relation and $V : P \to \wp(T)$ is a function. We say that T is the *carrier*, R is the *successor relation* and V is the *valuation function* of $\mathbb{S}$. For any pair $(s,t) \in R$ we say that s is a *predecessor* of t and t is a *successor* of s. For any pair $(s,t) \in R^+$, we say that s is an *ancestor* of t and t is a *descendant* of s. ◃

We introduce trees as a particular kind of transitions systems.

Definition 1.3 (Tree)**.** A tuple $\mathbb{T} = \langle T, s_I, R, V \rangle$ is a *C-labeled tree* if $\mathbb{T} = \langle T, R, V \rangle$ is a C-labeled transition system, $s_I \in T$ is a distinguished point that has no predecessor, each $s \in T$ that is different from s_I has a unique predecessor and the following identity holds.

$$T \quad = \quad R^{\star}[s_I]$$

The elements of T are called *nodes* and s_I is called the *root* of $\mathbb{T}$. ◃

Subtree Let $\mathbb{T} = \langle T, s_I, R, V \rangle$ be a C-labeled tree. A C-labeled tree $\mathbb{T}' = \langle T', s_I', R', V' \rangle$ is a *subtree* of $\mathbb{T}$ if $T' \subseteq T$, $R' = R \cap (T' \times T')$ and $V'(p) = V(p) \cap T'$ for each $p \in P$. Observe that each subtree of $\mathbb{T}$ is uniquely determined by its carrier. Each node $s \in T$ uniquely defines a subtree of $\mathbb{T}$ with carrier $R^{\star}[s]$ and root s, which we denote with $\mathbb{T}.s$.

Height and Leaf The *height* of a node $s \in T$ is inductively defined as follows: the root is the unique node at height 0; if $s \in T$ is a node of height i, then each $t \in R[s]$ is a node of height $i+1$. Two nodes $s,t \in T$ are *siblings* if there is a node $r \in T$ such that $s \in R[r]$ and $t \in R[r]$. A *leaf* of $\mathbb{T}$ is a node $s \in T$ such that $R[s] = \varnothing$. A tree $\mathbb{T}$ is *leafless* if no node in $\mathbb{T}$ is a leaf.

Path and Branch Let $S \subseteq T$ be a set of nodes. We say that S is a *path* if $S = (s_i)_{i<k}$ for some sequence $(s_i)_{i<k}$ with $k \leq \omega$ and $s_i R s_{i+1}$ for each $i < k$. We say that S is *backwards closed* if $t \in S$ and sRt implies $s \in S$, for all $s,t \in T$. Similarly, S is *frontwards closed* if, for all $s \in S$ with $R[s] \neq \varnothing$, there is some $t \in S$ with sRt. The set S is a *branch* of $\mathbb{T}$ if it is a path and it is both frontwards and backwards closed.

Branching Degree With the terminology *branching degree* we refer to the cardinality of the set $R[s]$ for nodes $s \in T$. A tree $\mathbb{T}$ is *finitely branching* if $R[s]$ is finite for all $s \in T$. For any $k < \omega$, we say that $\mathbb{T}$ is a *k-bounded tree* if $|R[s]| \leq k$ for all $s \in T$. In the specific case in which $|R[s]| = 2$ for all $s \in T$, we say that $\mathbb{T}$ is a *binary tree*. We say that $\mathbb{T}$ is *arbitrarily branching* if there is no special requirement on its branching degree.

Well-foundedness The tree $\mathbb{T}$ is *well-founded* if every path in $\mathbb{T}$ is finite. We denote with $WF(\mathbb{T})$ the set of well-founded subtrees of $\mathbb{T}$. A set of nodes $S \subseteq T$ is *well-closed* if $S \subseteq S'$, where S' is the carrier of some well-founded subtree of $\mathbb{T}$. We use the notation $WC(T)$ to indicate the set of well-closed subsets of T.

Frontier and Prefix We say that G is a *frontier* of $\mathbb{T}$ if $G \cap E$ is a singleton for every branch E of $\mathbb{T}$. A set S is a *prefix* of $\mathbb{T}$ if there exists a frontier G of $\mathbb{T}$ such that $S = \{s \in T \mid sR^{\star}t \text{ for some } t \in G\}$. Observe that each prefix S of $\mathbb{T}$ is the carrier of a well-founded backwards closed subtree of $\mathbb{T}$, with the property that for each node $s \in S$ either none or all successors of s are in S. It is easy to see that every prefix is uniquely determined by a frontier and viceversa. If S is a prefix, we denote with $Ft(S)$ the associated frontier. Given two frontiers G_1 and G_2 of $\mathbb{T}$, we write $G_1 < G_2$ if, for every branch E in $\mathbb{T}$, given $s_1 \in G_1 \cap E$ and $s_2 \in G_2 \cap E$, we have that $s_1 R^+ s_2$. Analogously, $G_1 \leq G_2$ holds if, for every branch E in $\mathbb{T}$, given $s_1 \in G_1 \cap E$ and $s_2 \in G_2 \cap E$, we have that $s_1 R^{\star} s_2$.

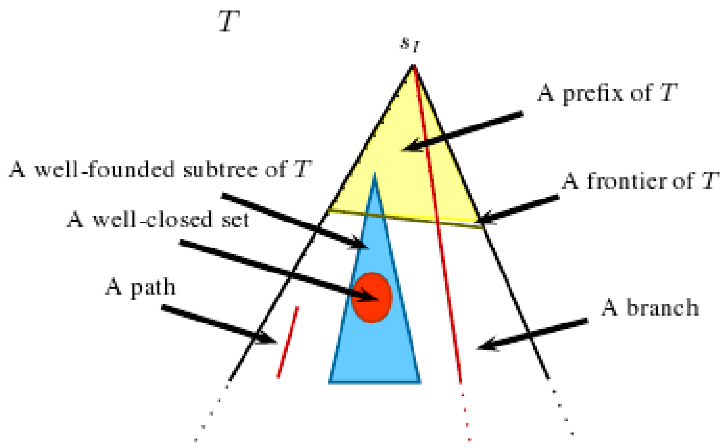

Figure 1.1: naming of parts.

p-variant Let p be a propositional letter (not necessarily in P). Given $\mathbb{T} = \langle T, s_I, R, V \rangle$, suppose that $\mathbb{T}^p = \langle T, s_I, R, V^p \rangle$ is a $\wp(P \cup \{p\})$-labeled tree such that $V^p : P \cup \{p\} \to \wp(T)$ is given as $V[p \mapsto S]$ for some $S \in \wp(T)$. We refer to $\mathbb{T}^p$ as a *p-variant* of $\mathbb{T}$. A p-variant $\mathbb{T}^p$ of $\mathbb{T}$ is *well-closed* if $V^p(p) \in WC(T)$. Similarly, $\mathbb{T}^p$ is a *finite* p-variant if $V^p(p) \subseteq_\omega T$. For a given set $S \in \wp(T)$, we denote with $\mathbb{T}[p \mapsto S]$ the p-variant $\mathbb{T}^p = \langle T, s_I, R, V^p \rangle$ of $\mathbb{T}$ obtained by putting $V^p = V[p \mapsto S]$.

Remark 1.4 (Coalgebraic Presentation [31])**.** It will be convenient to introduce an alternative presentation of trees, where the evaluation function and the successor relation are specified from the 'local' point of view of a node. Given a C-labeled tree $\mathbb{T} = \langle T, s_I, V, R \rangle$, we can represent $V : P \to \wp(T)$ as a *labeling function* $\sigma_C : T \to C$ and $R \subseteq T \times T$ as a *successor function* $\sigma_R : T \to \wp(T)$. Given a node $s \in T$, we call $\sigma_C(s)$ the *label* of s and for each $p \in \sigma_C(s)$ we say that s is *labeled with* p. Since σ_C and σ_R have the same domain, we can encode them as a single function $\sigma : T \to \wp(T) \times C$, assigning to each node the set of its successors and its label. Then we can represent $\mathbb{T}$ as a tuple $\langle T, s_I, \sigma \rangle$. Throughout this thesis we will mainly use this presentation for trees. ◂

Bisimulation is a notion of behavioral equivalence between processes [3]. Roughly, two processes are bisimilar when their behavior is indistinguishable from the point of view of an external observer. Transition systems are a mathematical model for processes and bisimulation is usually rendered as a binary relation. In the sequel we define this notion for the case of trees.

Definition 1.5 (Bisimulation)**.** Given C-labeled trees $\mathbb{T} = \langle T, s_I, \sigma \rangle$ and $\mathbb{T}' = \langle T', s'_I, \sigma' \rangle$, a *bisimulation* is a relation $Z \subseteq T \times T'$ such that for all $(t, t') \in Z$ the following holds:

- $\sigma_C(t) = \sigma'_C(t')$;
- for all $s \in \sigma_R(t)$ there is $s' \in \sigma'_R(t')$ such that $(s, s') \in Z$;
- for all $s' \in \sigma'_R(t')$ there is $s \in \sigma_R(t)$ such that $(s, s') \in Z$.

The trees $\mathbb{T}$ and $\mathbb{T}'$ are *bisimilar* if there is a bisimulation $Z \subseteq T \times T'$ including (s_I, s'_I). We write $\mathbb{T} \rightleftarrows \mathbb{T}'$ to indicate that $\mathbb{T}$ and $\mathbb{T}'$ are bisimilar. ◁

Given a tree $\mathbb{T}$, the *ω-expansion* $\mathbb{T}_\omega$ of $\mathbb{T}$ is a canonical instance of a tree which is bisimilar to $\mathbb{T}$. Intuitively, $\mathbb{T}_\omega$ is given as a tree with the same root of $\mathbb{T}$ and ω copies of any other node $s \in T$.

Definition 1.6 (ω-expansion)**.** Given a C-labeled trees $\mathbb{T} = \langle T, s_I, \sigma \rangle$, the *$\omega$-expansion* of $\mathbb{T}$ is a C-labeled tree $\mathbb{T}_\omega = \langle T_\omega, (s_I, 0), \sigma^\omega \rangle$ defined as follows.

- The carrier T_ω is given as $((T \setminus \{s_I\}) \times \omega) \cup \{(s_I, 0)\}$.
- For each node $(s, i) \in T_\omega$, the label $\sigma^\omega_C(s, i)$ of (s, i) is just $\sigma_C(s)$, and the set $\sigma^\omega_R(s, i)$ of its successors is given as $\sigma_R(s) \times \omega$. ◁

Remark 1.7. Each C-labeled tree $\mathbb{T}$ is bisimilar to its ω-expansion $\mathbb{T}_\omega$. A bisimulation relation $Z \subseteq T \times T_\omega$ witnessing this fact can be defined by putting

$$Z := \{(s,(s,i)) \mid s \in (T \setminus \{s_I\}) \text{ and } i < \omega\} \cup \{(s_I,(s_I,0))\}.$$

In words, Z links each node $s \in T$ to all the copies of s in the ω-expansion. ◂

Convention 1.8. Throughout this thesis, every tree $\mathbb{T}$ that we consider is *leafless* and *C-labeled* if not specified otherwise. ◂

1.3 Monadic Second-Order Logics on Trees

Definition 1.9 (Syntax)**.** The *monadic second-order language* on P is defined by the grammar

$$\varphi \quad ::= \quad p \sqsubseteq q \mid R(p,q) \mid \neg\varphi \mid \varphi \vee \varphi \mid \exists p.\varphi, \tag{1.1}$$

where p and q are letters from P. Given a formula φ of the monadic second-order language, we denote with $FV(\varphi)$ and $BV(\varphi)$ respectively the set of *free* and *bound* letters occurring in φ, defined as expected. We also adopt the standard convention that no letter occurs both free and bound in φ, that is, $FV(\varphi)$ and $BV(\varphi)$ are disjoint.

The language of *monadic second-order logic* (*MSO_P*), *weak monadic second-order logic* (*$WMSO_P$*) and *well-founded monadic second-order logic* (*$WFMSO_P$*) is the monadic second-order language on P. We omit the subscript P when this is clear from the context. ◃

Definition 1.10 (Semantics)**.** Let $\mathbb{T} = \langle T, s_I, V, R\rangle$ be a C-labeled tree and φ a formula of the monadic second-order language on P. The semantics of *MSO* is given by the following clauses, defining a truth relation ⊨ between $\mathbb{T}$ and φ, by induction on φ. If $\mathbb{T} \models \varphi$ holds, then we say that φ is *true* in $\mathbb{T}$.

$\mathbb{T} \models p \sqsubseteq q$	*iff*	$V(p) \subseteq V(q)$
$\mathbb{T} \models R(p,q)$	*iff*	for all $s \in V(p)$ there is some $t \in V(q)$ with sRt
$\mathbb{T} \models \neg\varphi$	*iff*	$\mathbb{T} \not\models \varphi$
$\mathbb{T} \models \varphi \vee \psi$	*iff*	$\mathbb{T} \models \varphi$ or $\mathbb{T} \models \psi$
$\mathbb{T} \models \exists p.\varphi$	*iff*	there is a p-variant $\mathbb{T}^p$ of $\mathbb{T}$ such that $\mathbb{T}^p \models \varphi$

The semantics of *WMSO* is given as the semantics of *MSO* but for the clause of the existential quantifier, which is replaced by the following clause.

$$\mathbb{T} \models \exists p.\varphi \quad \textit{iff} \quad \text{there is a finite } p\text{-variant } \mathbb{T}^p \text{ of } \mathbb{T} \text{ such that } \mathbb{T}^p \models \varphi$$

The semantics of *WFMSO* is given as the semantics of *MSO* but for the clause of the existential quantifier, which is replaced by the following clause.

$$\mathbb{T} \models \exists p.\varphi \quad \textit{iff} \quad \text{there is a well-closed } p\text{-variant } \mathbb{T}^p \text{ of } \mathbb{T} \text{ such that } \mathbb{T}^p \models \varphi$$

Let $\varphi \in MSO$ be a formula. We denote with $\|\varphi\|_P$ the set of C-labeled trees $\mathbb{T}$ such that $\mathbb{T} \models \varphi$. The subscript P is omitted when the set P of propositional letters is clear from the context. ◃

Remark 1.11. The monadic second-order language is a *one-sorted* language: the only variables appearing are the letters from the set P, which are interpreted over *sets*. This definition is very convenient for the automata-theoretic perspective that we will consider throughout this thesis. Perhaps a different version of the monadic second-order language may have been expected, with *two* sorts of variables. For instance, given a set *Var* of individual variables and the usual set P of set variables, consider the language defined by the following grammar.

$$\varphi \quad ::= \quad x \approx y \mid x \in p \mid xRy \mid \neg\varphi \mid \varphi \vee \varphi \mid \exists x.\varphi \mid \exists p.\varphi \tag{1.2}$$

Variables x and y are from the set *Var* and p is from the set P. We can provide a semantic interpretation on trees for formulae of this language in a completely standard way: ≈ is interpreted as equality of nodes, R as the successor relation, and $\exists x$ and $\exists p$ denote respectively first-order quantification (that is, quantification over nodes) and second-order quantification (that is, quantification over sets of nodes).

In fact the monadic second-order logics based on languages as in (1.1) or (1.2) are equivalent: the key observation is that an individual variable x can be seen as a set variable p_x whose interpretation is restricted to singletons. The translation from a formula as in (1.2) to a formula as in (1.1) crucially involves formulae $Empty(p)$ and $Singl(p)$ defined by putting

$$\begin{aligned} Empty(p) &:= \forall q\ (p \sqsubseteq q) \\ Singl(p) &:= (\neg(Empty(p))) \wedge \forall q\ (q \sqsubseteq p \rightarrow (Empty(q) \vee p \sqsubseteq q)). \end{aligned}$$

Either if we interpret $Empty(p)$ and $Singl(p)$ according to the semantics of *MSO*, *WMSO* or *WFMSO*, the formula $Empty(p)$ holds in a tree $\mathbb{T}$ when $V(p)$ is the empty set and $Singl(p)$ when $V(p)$ is a singleton. We refer to [31], remark 6.34 for more details on this translation. ◂

1.4 The Modal μ-Calculus

We refer to [31] for a thorough introduction to the modal μ-calculus (μMC).

Definition 1.12 (Syntax)**.** The language of the modal μ-calculus on P is given by the following grammar:

$$\varphi ::= p \mid \neg p \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid \diamond \varphi \mid \Box \varphi \mid \mu p.\varphi \mid \nu p.\varphi,$$

where p is a letter from P, which does not occur under the scope of $\neg$ in $\mu p.\varphi$ and $\nu p.\varphi$. We call μ and ν respectively *least* and *greatest fixpoint operator*. Given a formula $\varphi \in \mu MC$, we define the sets $FV(\varphi)$ and $BV(\varphi)$ of *free* and *bound* variables of φ as expected, with fixpoint operators binding propositional letters analogously to quantifiers of the monadic second-order language. We also adopt the standard convention that $FV(\varphi)$ and $BV(\varphi)$ are disjoint. ◃

Definition 1.13 (Semantics)**.** Given a tree $\mathbb{T} = \langle T, s_I, V, R\rangle$, we inductively define the *meaning* $\|\varphi\|^{\mathbb{T}}$ of a formula $\varphi \in \mu MC$ in $\mathbb{T}$ as follows.

$$\begin{aligned}
\|p\|^{\mathbb{T}} &:= V(p) \\
\|\neg p\|^{\mathbb{T}} &:= T \setminus V(p) \\
\|\psi_1 \wedge \psi_2\|^{\mathbb{T}} &:= \|\psi_1\|^{\mathbb{T}} \cap \|\psi_2\|^{\mathbb{T}} \\
\|\psi_1 \vee \psi_2\|^{\mathbb{T}} &:= \|\psi_1\|^{\mathbb{T}} \cup \|\psi_2\|^{\mathbb{T}} \\
\|\Box\psi\|^{\mathbb{T}} &:= \{s \in T \mid \forall t\ (sRt \Rightarrow t \in \|\psi\|^{\mathbb{T}}\} \\
\|\diamond\psi\|^{\mathbb{T}} &:= \{s \in T \mid \exists t\ (sRt \wedge t \in \|\psi\|^{\mathbb{T}}\} \\
\|\mu p.\psi\|^{\mathbb{T}} &:= \bigcap\{S \subseteq T \mid S \supseteq \|\psi\|^{\mathbb{T}[p\mapsto S]}\} \\
\|\nu p.\psi\|^{\mathbb{T}} &:= \bigcup\{S \subseteq T \mid S \subseteq \|\psi\|^{\mathbb{T}[p\mapsto S]}\}
\end{aligned}$$

We say that φ is *true* in $\mathbb{T}$ - notation $\mathbb{T} \vDash \varphi$ - if the following condition holds.

$$\mathbb{T} \vDash \varphi \quad \textit{iff} \quad s_I \in \|\varphi\|^{\mathbb{T}}.$$

As for the case of *MSO*, we denote with $\|\varphi\|_P$ the set of C-labeled trees $\mathbb{T}$ such that $\mathbb{T} \vDash \varphi$. The subscript P is omitted when the set P of propositional letters is clear from the context. ◃

Formulae of the modal μ-calculus are classified according to their *alternation depth*, which is informally given as the maximal length of a chain of nested alternating least and greatest fixpoint operators [4]. In particular, we are interested in the *alternation-free fragment* of modal μ-calculus, which is the collection of μMC-formulae without nesting of least and greatest fixpoint operators. The study of this fragment is motivated by computational feasibility, the alternation depth being the major factor in the complexity of model-checking algorithms for the modal μ-calculus [8].

Definition 1.14. We define the alternation-free fragment of modal μ-calculus (*AFMC*) as the set of formulae $\varphi \in \mu MC$ with the following property:

⋆ for any two subformulae of φ of the form $\mu p.\psi_1$ and $\nu q.\psi_2$, letters p and q do not occur free respectively in ψ_2 and ψ_1. ◃

Example 1.15. The formula $\mu p.(\Box p \vee (\nu q.(\diamond q \wedge r)))$ is in *AFMC*, because p does not occur free in $\diamond q \wedge r$ and q does not occur free in $\Box p \vee (\nu q.(\diamond q \wedge r))$. Instead, the formula $\mu p.(\nu q.(\Box p \vee r \vee \diamond q))$ is not in *AFMC* because p occurs free in $\Box p \vee r \vee \diamond q$. ◂

1.5 Definability

Tree Languages We usually refer to a set of C-labeled trees $\mathcal{L}$ as a *tree language* on P - or just as a tree language, if P is clear from the context. We indicate with $\overline{\mathcal{L}}$ the *complement* of $\mathcal{L}$, i.e. the set of C-labeled trees that are not in $\mathcal{L}$. A tree language $\mathcal{L}$ is *MSO-definable* if there is a formula $\varphi \in MSO$ such that $\|\varphi\| = \mathcal{L}$. We say that φ *defines* $\mathcal{L}$. Given *MSO*-formulae φ_1 and φ_2, we say that they are *equivalent* - notation $\varphi_1 \equiv \varphi_2$ - if $\|\varphi_1\| = \|\varphi_2\|$. We define in the same way analogous notions of definability for *WMSO*, *WFMSO*, *μMC* and *AFMC*.

Bisimulation Invariance The tree language $\mathcal{L}$ is *bisimulation closed* if $\mathbb{T} \rightleftarrows \mathbb{T}'$ implies that $\mathbb{T} \in \mathcal{L} \Leftrightarrow \mathbb{T}' \in \mathcal{L}$ for each tree $\mathbb{T}$ and $\mathbb{T}'$. A formula $\varphi \in MSO$ is *bisimulation invariant* if $\mathbb{T} \rightleftarrows \mathbb{T}'$ implies that $\mathbb{T} \vDash \varphi \Leftrightarrow \mathbb{T}' \vDash \varphi$ for each tree $\mathbb{T}$ and $\mathbb{T}'$. We define in the same way analogous notions of bisimulation invariance for *WMSO*, *WFMSO*, *μMC* and *AFMC*.

Proposition 1.16. *Each μMC-definable tree language is bisimulation closed.*

Proposition 1.17 (Janin-Walukiewicz Theorem [15])**.** *The class of MSO-definable tree languages that are bisimulation closed coincides with the class of μMC-definable tree languages.*

The following is a corollary of Bradfield's result on the modal μ-calculus [4], showing that the alternation depth hierarchy does not collapse.

Proposition 1.18 ([4])**.** *The class of AFMC-definable tree languages is strictly included in the class of μMC-definable tree languages.*

1.6 Topological Complexity

We are interested in measuring the complexity of tree languages, from a topological point of view. In chapter 5 we are going to use the topological perspective to compare the expressive power of different logics on trees.

Borel Sets Given a topological space (X,τ), the class $Borel(X) \subseteq \wp(X)$ of *Borel sets* of (X,τ) is the smallest collection having all the open sets of (X,τ) as elements and that is closed under the set-theoretical operations of countable union and complementation. A subset $Y \subseteq X$ is *Borel* if it is an element of $Borel(X)$.

Prefix Topology [9] We define a topology on C-labeled trees. For a C-labeled tree $\mathbb{T}$ and a natural number $n < \omega$ the *depth-n prefix* of $\mathbb{T}$, denoted as $\mathbb{T}^{(n)}$, is the subtree of $\mathbb{T}$ with carrier the prefix $\{s \in T \mid s \text{ has height at most } n\}$. We say that two trees $\mathbb{T}$ and $\mathbb{T}'$ are *equivalent up to height n* if $\mathbb{T}^{(n)} = \mathbb{T}'^{(n)}$. We call a set $\mathcal{X}$ of trees *open* if, for each $\mathbb{T} \in \mathcal{X}$, there is a natural number $n \geq 1$ such that, for any tree $\mathbb{T}'$, if $\mathbb{T}^{(n)} = \mathbb{T}'^{(n)}$ then $\mathbb{T}'$ is in $\mathcal{X}$. It can be checked that this definition of open set yields a topology on C-labeled trees, which we call *prefix topology*.

The following results relate topological complexity and logical definability of tree languages.

Proposition 1.19 ([9])**.** *Let $\mathcal{L}$ be a tree language on P. If $\mathcal{L}$ is WMSO-definable then it is a Borel set of the prefix topology.*

Proposition 1.20 ([9])**.** *The tree language on P defined by the formula $\mu q.(\Box q \vee p)$ is not a Borel set of the prefix topology.*

1.7 First-Order Logic

Throughout this section, we fix a finite set A of unary predicates. First, we present a first-order logic on signature given by A. It will be used later to define automata whose transition function is given in terms of first-order sentences.

Definition 1.21. Let *Var* a set of first-order variables. We define $For^+(A)$ as the set of formulae generated by the following grammar.

$$\varphi, \psi \ ::= \ \top \mid \bot \mid x \approx y \mid x \not\approx y \mid a(x) \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \exists x.\varphi \mid \forall x.\varphi \tag{1.3}$$

The variables x and y are from the set *Var*. Intuitively, $For^+(A)$ is the set of first-order formulae on signature A, where unary predicates from A can only occur positively. Given a subset S of A, we introduce the notation

$$\tau_S^+(x) \quad := \quad \bigwedge_{a \in S} a(x).$$

The formula $\tau^+_S(x)$ is called a *positive A-type*. We use the convention that, if S is the empty set, then $\tau^+_S(x)$ is $\top$. Given a finite set $X \subseteq_\omega For^+(A)$ of formulae, $\bigvee X$ is the formula given as the (finite) disjunction of all formulae in X. Given a set $Y \subseteq For^+(A)$ of formulae, $SLatt(Y) = \{\bigvee X \mid X \subseteq_\omega Y\}$ is the collection of all finite disjunctions of formulae in Y. We indicate with $FO^+(A)$ the set of *sentences* from $For^+(A)$. ◁

Definition 1.22 (Semantics)**.** Given a set X, a function $m : A \to \wp(X)$ and a valuation $v : Var \to X$, we inductively define the notion of a formula $\varphi \in For^+(A)$ being *true* in (X,m,v) as follows.

$$
\begin{array}{lll}
(X,m,v) \models \top & & \\
(X,m,v) \not\models \bot & & \\
(X,m,v) \models x \approx y & \textit{iff} & v(x) = v(y) \\
(X,m,v) \models x \not\approx y & \textit{iff} & v(x) \neq v(y) \\
(X,m,v) \models a(x) & \textit{iff} & x \in m(a) \\
(X,m,v) \models \varphi \vee \psi & \textit{iff} & (X,m,v) \models \varphi \text{ or } (X,m,v) \models \psi \\
(X,m,v) \models \varphi \wedge \psi & \textit{iff} & (X,m,v) \models \varphi \text{ and } (X,m,v) \models \psi \\
(X,m,v) \models \exists x.\varphi & \textit{iff} & (X,m,v[x \mapsto s]) \models \varphi \text{ for some } s \in X \\
(X,m,v) \models \forall x.\varphi & \textit{iff} & (X,m,v[x \mapsto s]) \models \varphi \text{ for all } s \in X
\end{array}
$$

The function m is called a *marking*. We say that (X,m,v) is an *A-structure*. ◁

Given φ and ψ in $For^+(A)$, in the sequel we freely use the notation $\varphi \to \psi$ to abbreviate $\neg\varphi \vee \psi$, provided that $\neg\varphi$ can be rewritten into an equivalent formula in $For^+(A)$.

Definition 1.23. For sets A and X, let $\mathcal{M}_{A,X}$ be a set of markings of type $A \to \wp(X)$. We define a partial order $\trianglelefteq$ on $\mathcal{M}_{A,X}$ by putting

$$m \trianglelefteq m' \quad \textit{iff} \quad m(a) \subseteq m'(a) \text{ for all } a \in A.$$

◁

Let $\varphi \in FO^+(A)$ be a sentence and (X,m,v) a first-order structure on signature A. Since φ has no free variables, we simply write $(X,m) \models \varphi$ to indicate that φ is true in (X,m,v) for any v. Each sentence in $FO^+(A)$ enjoys a monotonicity property that we introduce with the following remark.

Remark 1.24 (Monotonicity)**.** Let X be a set and $\varphi \in FO^+(A)$ a sentence. We observe the following property of φ, that can be easily verified by induction on φ:

- ⋆ Let $m : A \to \wp(X)$ be a marking such that $(X,m) \models \varphi$. For every marking $m' : A \to \wp(X)$ such that $m \trianglelefteq m'$ we have that $(X,m') \models \varphi$. ◂

Definition 1.25 (Basic Form)**.** Let $B_1 \dots B_k$ and $C_1 \dots C_j$ be sequences of subsets of A, possibly empty if $k = 0$ or $j = 0$. A sentence $\varphi \in FO^+(A)$ is in *basic form* if it is of shape

$$\varphi \quad = \quad \exists x_1 \dots \exists x_k \, (\textit{diff}(\bar{x}) \wedge \bigwedge_{1 \le i \le k} \tau^+_{B_i}(x_i) \wedge \forall z \, (\textit{diff}(\bar{x},z) \to \bigvee_{1 \le l \le j} \tau^+_{C_l}(z))),$$

where each $\tau^+_{B_i}(x_i)$ and $\tau^+_{C_l}(z)$ is a positive A-type, as in definition 1.21, and $\textit{diff}(y_1,\dots,y_n) := \bigwedge_{1 \le m < m' < n}(y_m \not\approx y_{m'})$ is a $For^+(A)$-formula stating that the interpretations of $y_1,\dots,y_n$ are pairwise different. The sentence φ is of the form $\exists x_1 \dots \exists x_k(\psi_1 \wedge \forall z\, \psi_2)$; we refer to ψ_1 and ψ_2 respectively as the *existential* and the *universal* part of φ. We indicate which positive A-types appear in the sentence by saying that φ *depends* on the sequence $B_1 \dots B_k, C_1 \dots C_j$ of subsets of A. We denote with $BF^+(A)$ the set of all sentences from $FO^+(A)$ that are in basic form. ◁

Using Ehrenfeucht-Fraïssé Games [13] it is possible to show that every sentence $\varphi \in FO^+(A)$ can be rewritten into an equivalent disjunction of sentences in $BF^+(A)$.

Proposition 1.26 ([33] - Lemma 38, [11] - Lemma 16.23)**.** *Let $\varphi \in FO^+(A)$ be a sentence. There is a sentence $\varphi' \in SLatt(BF^+(A))$ such that $\varphi \equiv \varphi'$.*

Remark 1.27. A sentence $\varphi \in BF^{+}(A)$ in basic form provides a quite informative picture of each A-structure (X,m) in which is true. By the particular shape of φ, the marking m has the effect of partitioning X into two sets $X^m_\exists$ and $X^m_\forall = X \setminus X^m_\exists$. The set $X^m_\exists$ consists of the witnesses for variables $x_1,\dots,x_k$ in the existential part in φ. By the presence of the subformula $\mathit{diff}(\bar{x})$ in φ, we know that $X^m_\exists$ contains *exactly* k elements $s_1,\dots,s_k$. For i with $1 \le i \le k$, the node $s_i \in X^m_\exists$ is associated with the positive A-type $\tau^+_{B_i}(x_i)$, meaning that $s_i \in \bigcap_{a\in B_i} m(a)$. Analogously, the set $X^m_\forall$ contains all the other elements of X, which are witnesses for the variable z in the universal part of φ. We have that $s \in \bigcap_{b\in C_l} m(b)$, for some positive A-type $\tau^+_{C_l}(z)$ occurring in the universal part of φ.

The A-structure (X,m) is generally a 'redundant' representation of the sentence φ. Each element $t \in X$ witnesses some variable y of φ, either in its existential or universal part, associated with a positive A-type $\tau^+_S(y)$. This means that there is a set $S_t \subseteq A$, such that S is a subset of S_t and t is in $\bigcap_{a\in S_t} m(a)$. The key observation is that t would still be a 'good' witness for the variable y if we do not assign to t any unary predicate in $S_t \setminus S$. Following this intuition, we say that a marking $m^\flat : A \to \wp(X)$ is a *shrinking* of m if the following conditions hold.

1. Given any $s_i \in \{s_1,\dots,s_k\} = X^m_\exists$,
$$\{a \in A \mid s_i \in m^\flat(a)\} \quad = \quad B_i.$$
2. Given any $t \in X \setminus X^m_\exists = X^m_\forall$,
$$\{a \in A \mid t \in m^\flat(a)\} \quad \subseteq \quad \{a \in A \mid t \in m(a)\}.$$
Furthermore, $\{a \in A \mid t \in m^\flat(a)\}$ is a minimal element of $\{C_1,\dots,C_j\}$ with respect to the order $\subseteq$.

It is clear by the syntactic shape of φ that at least one shrinking of m exists. Intuitively, the two conditions express that the A-structure $(X,m^\flat)$ is a 'non-redundant' representation of the sentence φ, obtained by 'contracting' the representation (X,m). The marking $m^\flat$ assigns to each element t of X a subset of A, which is $\subseteq$-minimal among the ones making t a witness for the corresponding variable in φ, according to the partition $X^m_\exists \cup X^m_\forall$. These intuitions are fixed by the next three statements, which easily follow by the conditions on $m^\flat$ expressed above.

a) $m^\flat \trianglelefteq m$.

b) $(X,m^\flat) \models \varphi$.

c) For each marking $\tilde{m} : A \to \wp(X)$, if $\tilde{m} \trianglelefteq m^\flat$ and $(X,\tilde{m}) \models \varphi$, then $\tilde{m} = m^\flat$. ◂

1.8 Game Terminology and Parity Games

Throughout this thesis we work with automata processing trees. A very convenient way to describe a run of such automata is by means of games. In particular, since all trees are assumed leafless, a run will generally be an infinite object, that we want to model through an *infinite game*. For this purpose, we introduce some terminology and background on infinite games. All the games that we consider involve two players called *Eloise* ($\exists$) and *Abelard* ($\forall$). In some contexts we refer to player Π, meaning that we want to specify a notion for a generic player in $\{\exists,\forall\}$.

Board Games A *board game* $\mathcal{G}$ is a tuple $(G_\exists, G_\forall, E, \mathit{Win})$, where $G_\exists$ and $G_\forall$ are disjoint sets whose union $G = G_\exists \cup G_\forall$ is called *board*, $E \subseteq G \times G$ is a set of *edges*, and $\mathit{Win} \subseteq G^\omega$ is a set of G−streams. Each element $u \in G$ is a *position*. Intuitively, if u is an element of G_Π, this means that player Π is supposed to move from position u. An *initialized board game* $\mathcal{G}@u_I$ is a tuple $(G_\exists, G_\forall, u_I, E, \mathit{Win})$ where $(G_\exists, G_\forall, E, \mathit{Win})$ is a board game and $u_I \in G$ is a distinguished position that we call the *initial position* of the game.

Matches Given a board game $\mathcal{G}$, a *match* in $\mathcal{G}$ is a sequence $\pi = (u_i)_{i<\alpha}$ of positions of $\mathcal{G}$, where α is either ω or a natural number, and $(u_i,u_{i+1}) \in E$ for all i with $i+1 < \alpha$. Analogously, given an initialized board game $\mathcal{G}@u_I$, we say that π is a match in $\mathcal{G}@u_I$ if it is a match in $\mathcal{G}$ and $u_0 = u_I$. If $\alpha = \omega$, we say that π is an *infinite match*. Otherwise, $\alpha = k$ for some $k < \omega$ and π is a finite match. We refer to u_{k-1} as the *last* position in π, for which we use the notation $\mathit{last}(\pi)$. Since $\mathit{last}(\pi)$ is an element of $G = G_\exists \cup G_\forall$, then one of the two players, that we indicate with Π, is supposed to move from $\mathit{last}(\pi)$. If there is no $u \in G$ such that $(\mathit{last}(\pi),u) \in E$, we say that player Π *gets stuck* in π.

If π is infinite or π is finite with one of the two players getting stuck, we say that π is a *total match*. Otherwise π is a *partial match*. If π is a total match then it is won by some player. If π is finite, then the winner is the opponent of the player who gets stuck. Otherwise π is infinite, meaning that it is a G−stream: $\exists$ wins if π belongs to Win, and $\forall$ wins if π does not belong to Win. Given two matches $\pi = (u_i)_{i<\alpha}$ and $\pi' = (v_i)_{i<\gamma}$, we say that π' *extends* π if $\alpha \le \gamma$ and $u_i = v_i$ for all $i < \alpha$.

Strategies Given a board game $\mathcal{G}$ and a player Π, let PM_Π^G denote the set of partial matches of $\mathcal{G}$ whose last position belongs to player Π. A *strategy for* Π is a function f of type $PM_\Pi^G \to G$. A match $\pi = (u_i)_{i<\alpha}$ of $\mathcal{G}$ is f*-conform* if for each $i < \alpha$ such that $u_i \in G_\Pi$ we have that $u_{i+1} = f(u_0, \ldots, u_i)$. Given a partial match π in $Dom(f)$, the position $f(\pi)$ is *legitimate* if $(last(\pi), f(\pi))$ is in E.

Given $u \in G$, a strategy $f : PM_\Pi^G \to G$, consider the following two conditions.

1. For each f-conform partial match π of $\mathcal{G}@u$, if $last(\pi)$ is in G_Π then $f(\pi)$ is legitimate.
2. Each f-conform total match of $\mathcal{G}@u$ is won by Π.

If f respects the first condition, we say that f is a *surviving strategy* for Π in $\mathcal{G}@u$. Intuitively, if f is surviving then player Π never gets stuck in matches that are played according to f. Furthermore, if f respects both the first and the second condition, then we say that f is a *winning strategy* for Π in $\mathcal{G}@u$. If Π has a winning strategy in $\mathcal{G}@u$ then we say that u is a *winning position* for Π in $\mathcal{G}$. We denote with $Win_\Pi(\mathcal{G})$ the set of positions of $\mathcal{G}$ that are winning for Π.

Remark 1.28. As given above, a strategy for player Π is defined *for all* partial matches in PM_Π^G. However, throughout this thesis we will occasionally work with strategies which are only defined on a subset X of PM_Π^G. This is convenient for the purpose of merging several strategies $f_1, f_2, \ldots, f_k$ together, obtaining a well-defined strategy f' by the union of their graphs. In fact, we can assume that any partially defined strategy $f : X \to G$ has domain PM_Π^G, by letting $f(\pi)$ be an arbitrary position for all partial matches $\pi \in PM_\Pi^G \setminus X$.

Similarly, notice that the property of a strategy f of being surviving or winning for Π only depends on the value of f on partial matches in PM_Π^G that are f-conform. By this observation, for the purpose of showing that f is surviving or winning, we usually define it just on f-conform partial matches in PM_Π^G. ◂

Parity Games Let $\mathcal{G} = (G_\exists, G_\forall, E, Win)$ be a board game. A *parity map* is a function $\Omega : G \to \omega$ assigning a natural number to each position in G, such that $\Omega[G]$ is finite. Given an infinite match $\pi \in G^\omega$, we denote with $Inf(\pi)$ the set $\{k < \omega \mid \Omega(u) = k$ for infinitely many $u \in \pi\}$. Since π is infinite and $\Omega[G]$ is finite, then $Inf(\pi)$ is non-empty. We say that Win is a *parity set* if there exists a parity map $\Omega : G \to \omega$ such that

$$Win \quad = \quad \{\pi \in G^\omega \mid Min(Inf(\pi)) \text{ is even}\}.$$

A *parity game* is a board game $\mathcal{G} = (G_\exists, G_\forall, E, Win)$ where Win is a parity set. We can see parity games as board games where Win presents a quite regular structure. What makes them so appealing is that they enjoy a remarkable property which is called *positional determinacy*.

Positional Determinacy A strategy $f : PM_\Pi^G \to G$ is called *positional* if $f(\pi) = f(\pi')$ for each π and π' in $Dom(f)$ with $last(\pi) = last(\pi')$. Intuitively, positional strategies only depend on the last position of partial matches on which they are defined. For this reason, a positional strategy with type $PM_\Pi^G \to G$ can represented as a function of type $G_\Pi \to G$.

A board game $\mathcal{G}$ with board G is *determined* if $G = Win_\exists(\mathcal{G}) \cup Win_\forall(\mathcal{G})$, that is, each $u \in G$ is a winning position for one of the two players.

Theorem 1.29 (Positional Determinacy of Parity Games, [7], [22])**.** *For each parity game $\mathcal{G}$, there are positional strategies $f_\exists$ and $f_\forall$ respectively for player $\exists$ and $\forall$, such that for every position $u \in G$ there is a player Π such that f_Π is a winning strategy for Π in $\mathcal{G}@u$.*

Following theorem 1.29, it will be convenient to assume that each strategy we work with in parity games is positional.

1.9 Stream Automata

Many of the concepts we presented so far are related to infinite sequences, also called streams. This motivates the introduction of automata operating on streams, that will be used in Chapter 2. We assume that the reader is already familiar with elementary notions of automata theory such as run and acceptance condition, for which we refer to [11].

Definition 1.30. An *X−stream automaton* is a tuple $\mathbb{Z} = \langle Z, z_I, \Delta, Acc\rangle$ where Z is a finite set of states, $z_I \in Z$ is an initial state, $\Delta : Z \times X \to \wp(Z)$ is a transition function and $Acc \subseteq Z^\omega$ is an acceptance condition. We say that $\mathbb{Z}$ is *deterministic* if for each $(z, x) \in Z \times X$ the set $\Delta(z, x)$ is a singleton, and *non-deterministic* otherwise. We call $\mathbb{Z}$ a *parity X−stream automaton* if Acc is a parity set.

For an X−stream $(x_i)_{i<\omega}$, a *run* ρ of $\mathbb{Z}$ on $(x_i)_{i<\omega}$ is a Z−stream $(z_i)_{i<\omega}$ where $z_0 = z_I$ and $z_{i+1} \in \Delta(z_i, x_i)$ for each $i < \omega$. We say that $(x_i)_{i<\omega}$ is *accepted* by $\mathbb{Z}$ if there exists a run ρ of $\mathbb{Z}$ on $(x_i)_{i<\omega}$ such that $\rho \in Acc$. We denote with $L(\mathbb{Z})$ the set of X−streams that are accepted by $\mathbb{Z}$, also called the *language* of $\mathbb{Z}$. $\triangleleft$

Definition 1.31. For a set X, let $\mathcal{L} \subseteq X^\omega$ be a set of X−streams. Similarly to the case of trees, we refer to $\mathcal{L}$ as a *stream language*. We say that $\mathcal{L}$ is an ω-*regular language* if there is an X−stream automaton $\mathbb{Z}$ such that $L(\mathbb{Z}) = \mathcal{L}$. $\triangleleft$

Given X−stream automata $\mathbb{Z}_1$ and $\mathbb{Z}_2$, we write $\mathbb{Z}_1 \equiv \mathbb{Z}_2$ if $L(\mathbb{Z}_1) = L(\mathbb{Z}_2)$. We notice that an X−stream can also be seen as a very simple kind of tree, with a unique branch and no labeled node - automata on trees, that we introduce in the next chapter, are in fact a generalization of stream automata. By this observation, we occasionally make use of the notation introduced in this section also for automata working on trees.

Chapter 2

Automata Characterization of *MSO*

In this chapter we give an account of the expressive power of *MSO* in terms of automata working on trees. For this purpose we introduce *MSO-automata*. The underlying idea is that, for each formula $\varphi \in MSO$, we can effectively construct an *MSO*-automaton $\mathbb{A}_\varphi$ which is *equivalent* to φ, that is, $\mathbb{A}_\varphi$ has the following property:

$$\text{for any tree } \mathbb{T},\ \mathbb{T} \models \varphi \quad \textit{iff} \quad \mathbb{A}_\varphi \text{ accepts } \mathbb{T}. \tag{2.1}$$

2.1 *MSO*-Automata: Definition

Every *MSO*-automaton $\mathbb{A}$ will be given as a tuple $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$.

1. The first two components are a finite set A of states (also called *carrier*) and the initial state $a_I \in A$ of $\mathbb{A}$.
2. The automaton $\mathbb{A}$ depends on an alphabet, which is standardly given as the set $C := \wp(P)$, for P the set of propositional letters that we fixed with convention 1.1. The third component of $\mathbb{A}$ is a *transition function* Δ of type $A \times C \to FO^+(A)$. Operationally, this means that:
   - the function Δ takes as input a state $a \in A$ and the label $\sigma_C(s) \in C$ of a node s of $\mathbb{T}$;
   - the function Δ gives as output a *first-order sentence* $\Delta(a, \sigma_C(s)) \in FO^+(A)$ where states $a \in A$ of the automaton can occur positively as unary predicates.
3. The fourth component Ω is a function of type $A \to \omega$, assigning to each state $a \in A$ a natural number $\Omega(a)$.

Before giving the formal definition of *MSO*-automaton, we provide some intuitions on how the behavior of $\mathbb{A}$ is expressed in terms of Δ and Ω. For this purpose we fix a tree $\mathbb{T}$. The idea is to describe any run of $\mathbb{A}$ on $\mathbb{T}$ in terms of a game, which we call the *acceptance game* of $\mathbb{A}$ on $\mathbb{T}$. The acceptance game has two players: player $\exists$ claims that $\mathbb{T}$ should be accepted by $\mathbb{A}$, whereas player $\forall$ tries to refute this statement. A *basic position* of the game is a pair $(a, s) \in A \times T$ where a is a state of $\mathbb{A}$ and s is a node of $\mathbb{T}$. A *match* π proceeds in rounds, where each round is associated with a basic position. The interplay of the two players determines how we pass from a basic position (a_i, s_i) in round i to another basic position (a_{i+1}, s_{i+1}) in round $i+1$. Each round consists of two moves, that we can describe as follows.

- **Move of $\exists$:** from position $(a_i, s_i) \in A \times T$, player $\exists$ *provides a marking* $m : A \to \wp(\sigma_R(s_i))$ assigning sets of successors of s_i to states of $\mathbb{A}$. Then $(\sigma_R(s_i), m)$ is an *A-structure* according to definition 1.21.

  Requirement: the marking m chosen by $\exists$ must be such that the sentence $\Delta(a_i, \sigma_C(s_i))$ is *true* in $(\sigma_R(s_i), m)$.

- **Move of $\forall$:** given the marking $m : A \to \sigma_R(s_i)$, player $\forall$ chooses the next basic position $(a_{i+1}, s_{i+1}) \in A \times T$.

  Requirement: the position (a_{i+1}, s_{i+1}) chosen by $\forall$ must respect m, in the sense that s_{i+1} is in $m(a_{i+1})$.

Therefore π consists of basic positions - belonging to $\exists$ - and positions with markings - belonging to $\forall$, which occur alternated.

$$\pi \quad = \quad (a_1, s_1), m_1, (a_2, s_2), m_2, \ldots, (a_n, s_n), m_n, \ldots$$

We can assign a numeric value - which we call *parity* - to each position in π. Each basic position (a_i, s_i) is associated with parity $\Omega(a_i)$. All positions with a marking receive parity $Max(\Omega[A])$. Since every position receives some parity, acceptance games can be seen as *parity games*: winning conditions are defined accordingly.

Observe that, if π is infinite, then the minimum parity occurring infinitely often along the play is always associated with some basic position. The intuition is that positions with markings receive a conventional parity, which is not relevant for determining the winner of a match.

We are now ready to provide the formal definition of *MSO*-automata.

Definition 2.1 ([33][31]). An *MSO-automaton* on alphabet C is a tuple $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ where:

- A is a finite set of states;
- $a_I \in A$ is the *initial state* of $\mathbb{A}$;
- $\Delta : A \times C \to FO^+(A)$ is the *transition function* of $\mathbb{A}$;
- $\Omega : A \to \omega$ is the *parity map* of $\mathbb{A}$.

Given a tree $\mathbb{T}$, the *acceptance game* of $\mathbb{A}$ on $\mathbb{T}$ - notation $\mathcal{A}(\mathbb{A},\mathbb{T})$ - is a parity game defined according to the rules of table 2.1. We recall that finite matches of $\mathcal{A}(\mathbb{A},\mathbb{T})$ are lost by the player who gets stuck. An infinite match of $\mathcal{A}(\mathbb{A},\mathbb{T})$ is won by $\exists$ if and only if the *minimum* parity occurring infinitely often is even.

Position	Player	Admissible moves	Parity
$(a,s) \in A \times S$	$\exists$	$\{m : A \to \wp(\sigma_R(s)) \mid (\sigma_R(s), m) \models \Delta(a, \sigma_C(s))\}$	$\Omega(a)$
$m : A \to \wp(\sigma_R(s))$	$\forall$	$\{(b,t) \mid t \in m(b)\}$	$Max(\Omega[A])$

Table 2.1: Acceptance game for *MSO*-automata

The tree $\mathbb{T}$ is *accepted* by $\mathbb{A}$ if and only if $\exists$ has a winning strategy in $\mathcal{A}(\mathbb{A},\mathbb{T})@(a_I, s_I)$. ◁

Convention 2.2. In the sequel we assume that each *MSO*-automaton is on alphabet C, if not specified otherwise. ◂

Remark 2.3. A winning strategy f for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A},\mathbb{T})@(a_I, s_I)$ is in particular a surviving strategy for the same player in $\mathcal{G}$. Indeed, for each basic position $(a,s) \in A \times T$ that is reached in some f-conform match of $\mathcal{G}$, the marking m suggested by f makes $\Delta(a, \sigma_C(s))$ true in $\sigma_R(s)$, meaning that $\exists$ does not get stuck. The notion of surviving strategy can be conveniently restricted to subsets of T. Given $W \subseteq T$, we say that a strategy f' for $\exists$ in $\mathcal{G}$ is *surviving in* W if, for each basic position $(a,s) \in A \times W$ that is reached in some f'-conform match of $\mathcal{G}$, the marking m suggested by f' makes $\Delta(a, \sigma_C(s))$ true in $\sigma_R(s)$. ◂

Remark 2.4. Given an *MSO*-automaton $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ and a tree $\mathbb{T}$, let f be a strategy for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A},\mathbb{T})@(a_I, s_I)$. Since $\mathcal{G}$ is a parity game, we can assume f to be *positional*. Therefore it can be represented as a function from basic positions (of f-conform matches of $\mathcal{G}$) to markings:

$$f : (a,s) \quad \mapsto \quad m_{a,s},$$

where $m_{a,s} : A \to \wp(\sigma_R(s))$ is the move that f suggests to $\exists$ from position $(a,s) \in A \times T$. A very convenient way to represent the information carried by f is by displaying its graph as a tree $\mathbb{T}_f$, defined as follows:

- the carrier T_f of $\mathbb{T}_f$ consists of the basic positions in $Dom(f)$;
- the root of $\mathbb{T}_f$ is the basic position (a_I, s_I);
- the successor function $\sigma_R^f : T_f \to \wp(T_f)$ is defined by putting

  $$\sigma_R^f : (a,s) \quad \mapsto \quad \{(b,t) \mid t \in m_{a,s}(b)\}$$

  where $m_{a,s} = f(a,s)$.
- the labeling function $\sigma_C^f : T_f \to C$ is defined by putting

  $$\sigma_C^f : (a,s) \quad \mapsto \quad \sigma_C(s)$$

  where σ_C is the labeling function of $\mathbb{T}$.

An useful observation is that any f-conform partial match π of $\mathcal{A}(\mathbb{A},\mathbb{T})$ corresponds to a unique path B of $\mathbb{T}_f$, and viceversa. In order to see that, observe that B and π are both sequence of basic positions. The difference is that markings are represented explicitly in π as positions, whereas in B they determine the successor relation between basic positions. However, both presentations have the same amount of information, namely which are the sets of admissible moves for $\forall$ along the play.

We refer to $\mathbb{T}_f$ as the *tree representation* of f. It is convenient to fix a *projection function* $\pi_2^f : T_f \to T$, canonically defined by putting $\pi_2^f : (a,s) \mapsto s$. Observe that π_2^f preserves the tree structure, in the sense that, given basic positions (a,s) and (b,t) in $\mathbb{T}_f$ with (b,t) in $\sigma_R^f(a,s)$, the node $t = \pi_2^f(b,t)$ is in $\sigma_R(s) = \sigma_R(\pi_2^f(a,s))$. ◂

In the sequel we provide two basic examples of how *MSO*-automata and *MSO*-formulae can be associated, as described in (2.1).

Example 2.5. Let p and q be letters in P. We want to provide an *MSO*-automaton $\mathbb{A}_{R(p,q)}$ such that for any tree $\mathbb{T}$

$$\mathbb{A}_{R(p,q)} \text{ accepts } \mathbb{T} \quad \textit{iff} \quad \mathbb{T} \vDash R(p,q).$$

Let $\mathbb{A}_{R(p,q)} = \langle A, a_I, \Delta, \Omega \rangle$ be defined as follows.

$$\begin{array}{rcl}
A & := & \{a_0, a_1\} \\
a_I & := & a_0 \\
\Delta(a_0,c) & := & \begin{cases} \exists x\, (a_1(x) \wedge \forall y\, (y \neq x \to a_0(y))) & \text{If } p \in c \\ \forall x\, (a_0(x)) & \text{Otherwise} \end{cases} \\
\Delta(a_1,c) & := & \begin{cases} \bot & \text{If } q \notin c \\ \exists x\, (a_1(x) \wedge \forall y\, (y \neq x \to a_0(y))) & \text{If } p \in c \text{ and } q \in c \\ \forall x\, (a_0(x)) & \text{Otherwise} \end{cases} \\
\Omega(a_0) & := & 0 \\
\Omega(a_1) & := & 0
\end{array}$$

Let $\mathbb{T}$ be a tree. We provide an informal argument showing that $\mathbb{A}_{R(p,q)}$ accepts $\mathbb{T}$ if and only if every node in $\mathbb{T}$ labeled with p has a successor labeled with q.

The main observation is that, since both states of $\mathbb{A}_{R(p,q)}$ have parity even, $\exists$ is going to win all infinite matches of $\mathcal{A}(\mathbb{A}_{R(p,q)},\mathbb{T})@(a_I,s_I)$. Therefore the only chance that $\forall$ has to win is by letting $\exists$ get stuck. By definition of Δ, this happens if and only if the match arrives at some node s that is labeled with p, from which $\exists$ has to mark with a_1 some node $t \in \sigma_R(s)$ such that $q \notin \sigma_C(t)$.

If a node s with this property exists, then $\forall$ has the power of bringing the match, in finitely many rounds, to a basic position of the form (a,s) for some $a \in A$. It suffices that at each round he selects the next basic position, of the form (b,t), in such a way that $tR^\star s$. If no node s with such property exists, the match is infinite and $\exists$ is the winner. ◂

Example 2.6. Let p and q be letters in P. We want to provide an *MSO*-automaton $\mathbb{A}_{p\sqsubseteq q}$ such that for any tree $\mathbb{T}$

$$\mathbb{A}_{p\sqsubseteq q} \text{ accepts } \mathbb{T} \quad \textit{iff} \quad \mathbb{T} \vDash p \sqsubseteq q.$$

The automaton $\mathbb{A}_{p\sqsubseteq q} = \langle A, a_I, \Delta, \Omega \rangle$ is defined as follows.

$$\begin{array}{rcl}
A & := & \{a_0\} \\
a_I & := & a_0 \\
\Delta(a_0,c) & := & \begin{cases} \forall x\, a_0(x) & \text{If } q \in c \text{ or } p \notin c \\ \bot & \text{Otherwise} \end{cases} \\
\Omega(a_0) & := & 0
\end{array}$$

Similarly to example 2.6 it is straightforward to check that $\mathbb{A}_{p\sqsubseteq q}$ accepts exactly the trees where every node labeled with p is also labeled with q. ◂

2.2 Functional Strategies and Their Syntactic Characterization

The above examples give an idea of how first-order logic allows for rich and flexible specifications of the transition function of automata. As we will see in section 2.4, a remarkable consequence of this *logical perspective* is that closure properties of the tree languages definable by *MSO*-automata, such as union and complementation, are easily derivable from closure properties of the set $FO^+(A)$.

However, this flexibility is also a source of difficulties and complexity. In order to see that, let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be an *MSO*-automaton, $\mathbb{T}$ a tree and f a winning strategy for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$. Suppose that $(a,s) \in A \times T$ is a basic position occurring in an f-conform match of $\mathcal{G}$, with $\Delta(a, \sigma_C(s)) \in FO^+(A)$ defined as $\exists x\ (a_1(x) \wedge a_2(x))$ for some $a_1, a_2 \in A$. In order to make this sentence true, the marking suggested by f must assign both a_1 and a_2 to the same node $t \in \sigma_R(s)$. This means that both position (a_1,t) and (a_2,t) can be chosen by $\forall$ to continue the match. Observe that $\forall$'s move affects significantly the continuation of the match, because he does not only determine the next node - t, for instance - from which the match is played, but also the associated state - either a_1 or a_2.

It is convenient to visualize this situation by drawing the *tree representation* $\mathbb{T}_f$ of f, as in remark 2.4. If we compare $\mathbb{T}_f$ and $\mathbb{T}$ at the height of the set $\sigma_R(s)$, we see that $\mathbb{T}_f$ is more 'complex' than $\mathbb{T}$ with respect to t: both positions (a_1,t) and (a_2,t) are nodes of $\mathbb{T}_f$. Intuitively, by the sole information that the node t of $\mathbb{T}_f$ is involved in an f-conform match, we cannot determine which basic position is associated with t.

This ability of inferring the structure of $\mathbb{T}_f$ from $\mathbb{T}$ is an essential ingredient of the projection construction on *MSO*-automata, that we will introduce in section 2.4 as the automata counterpart of existential quantification in *MSO*. For this reason, we want to avoid the situation described above, by enforcing that each node in $\mathbb{T}$ is marked with *at most one* state from A along f-conform matches of $\mathcal{G}$. This corresponds to the projection function $\pi_2^f : T_f \to T$ being a *1-1 correspondence* between $\mathbb{T}_f$ and $\mathbb{T}$. In fact, we can express the same condition in terms of properties of the strategy f itself. For this purpose, we introduce the notion of a strategy for $\exists$ being *functional*.

Definition 2.7. Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be an *MSO*-automaton and $\mathbb{T}$ a tree. Let f be a strategy for $\exists$ in $\mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$. We say that f is *functional* if, for each basic position $(a,s) \in Dom(f)$, the marking $f(a,s)$ assigns to each node $t \in \sigma_R(s)$ *at most one* state $b \in A$. ◁

Our goal is to define a transformation, which allows to pass from an *MSO*-automaton $\mathbb{A}$ to an equivalent *MSO*-automaton $\mathbb{A}'$ such that, for each tree $\mathbb{T}$, a winning strategy for $\exists$ in $\mathcal{A}(\mathbb{A}', \mathbb{T})@(a_I, s_I)$ can always assumed to be functional. The idea is to tackle this question by providing a *syntactic characterization* of functionality, in terms of the first-order sentences associated with the transition function of *MSO*-automata.

For this purpose, we recall to the notion of sentence in *basic form*, as in definition 1.25. By proposition 1.26, each sentence in $FO^+(A)$ is equivalent to a disjunction of sentences in basic form. It follows that the transition function Δ of $\mathbb{A}$ can be assumed of type $A \times C \to SLatt(BF^+(A))$ instead of $A \times C \to FO^+(A)$. In the same spirit, we want to show that the codomain of Δ can be further restricted to first-order sentences having a quite specific syntactic shape, which will be associated with the property of a strategy for $\exists$ of being functional.

Definition 2.8 (Functional basic form)**.** Given a set A of unary predicates, let $\varphi \in BF^+(A)$ be a sentence in basic form depending on sequences $B_1 \ldots B_k$ and $C_1 \ldots C_j$ of subsets of A, that is,

$$\varphi \quad = \quad \exists x_1 \ldots \exists x_k\ (\mathit{diff}(\bar{x}) \wedge \bigwedge_{1 \le i \le k} \tau^+_{B_i}(x_i) \wedge \forall z\ (\mathit{diff}(\bar{x}, z) \to \bigvee_{1 \le l \le j} \tau^+_{C_l}(z))).$$

We say that φ is in *functional basic form* if each S in $\{B_1 \ldots B_k, C_1 \ldots C_j\}$ is either the empty set or a singleton. We denote with $FBF^+(A)$ the set of all sentences in $BF^+(A)$ which are in functional basic form. ◁

Definition 2.9 (Non-deterministic automata)**.** Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be an *MSO*-automaton. We say that $\mathbb{A}$ is *non-deterministic* if Δ has type $A \times C \to SLatt(FBF^+(A))$. ◁

The following statement justifies the introduction of sentences in functional basic form as the 'syntactic characterization' of functional strategies for $\exists$.

Proposition 2.10. *Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be a non-deterministic MSO-automaton. Given any tree $\mathbb{T}$, each surviving strategy for $\exists$ in $\mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$ can be assumed to be functional.*

The proof of proposition 2.10 requires some preliminary observation. In fact, it is not hard to imagine a surviving strategy f for $\exists$ in $\mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$ which is *not* functional. Given a position $(a,s) \in A \times T$, suppose that $m : A \to \wp(\sigma_R(s))$ is a marking that makes $\Delta(a, \sigma_C(s))$ true in $\sigma_R(s)$. Since $\Delta(a, \sigma_C(s))$ is an element of $FO^+(A)$, it enjoys the *monotonicity property* described in remark 1.24, meaning that each marking which extends m also makes $\Delta(a, \sigma_C(s))$ true in $\sigma_R(s)$. Among these extensions of m, we can find a marking $m' : A \to \wp(\sigma_R(s))$

assigning more than one state to some node in $\sigma_R(s)$. Then, f might suggests to $\exists$ the marking m' from position (a,s), implying that it is not a functional strategy.

In order to show proposition 2.10, the key observation is that it is in $\exists$'s interest to make the fewest number of moves available for $\forall$. Thus a rational choice for her would be to assign to each node in $\sigma_R(s)$ only the 'strictly necessary' amount of states that makes $\Delta(a,\sigma_C(s))$ true. From this point of view, m' is not a very good suggestion for $\exists$, because it generally assigns to each node more states than the marking m, which still makes $\Delta(a,\sigma_C(s))$ true in $\sigma_R(s)$. By assuming that $\exists$ plays according to this idea of rationality, we can rule out redundant markings such as m'. Following these intuitions, we introduce the notion of *minimal strategy*.

Definition 2.11. Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ be an *MSO*-automaton, $\mathbb{T}$ a tree and f a surviving strategy for $\exists$ in $\mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$. Given $a \in A$ and $s \in T$, consider the sentence $\Delta(a,\sigma_C(s))$, which we can assume to be an element of $SLatt(BF^+(A))$ by proposition 1.26. Given a disjunct $\varphi \in BF^+(A)$ of $\Delta(a,\sigma_C(s))$, let us use the notation $\mathcal{M}_{s,\varphi}$ for the set

$$\{m : A \to \wp(\sigma_R(s)) \mid (\sigma_R(s),m) \vDash \varphi\}.$$

The set $\mathcal{M}_{s,\varphi}$ can be ordered according to the relation $\trianglelefteq$ between markings, as given in definition 1.23. We say that f is *minimal for* $\mathbb{A}$ *and* $\mathbb{T}$ if, for each basic position $(a,s) \in Dom(f)$, there is a disjunct φ of $\Delta(a,\sigma_C(s))$ such that the marking $f(a,s)$ is minimal with respect to the order $\trianglelefteq$ in the set $\mathcal{M}_{s,\varphi}$, that is, there is no marking $m' \in \mathcal{M}_{s,\varphi}$ such that $m' \neq f(a,s)$ and $m' \trianglelefteq f(a,s)$. ◁

Remark 2.12. Since we work with trees where a node can have infinitely many successors, the set $\mathcal{M}_{s,\varphi}$ is generally infinite and we need some more arguing to show that it always has a minimal element with respect to the order $\trianglelefteq$. The key observation is given by remark 1.27: a marking $m : A \to \wp(\sigma_R(s))$ that makes $\varphi \in BF^+(A)$ true always has some *shrinking*. This means that there is a marking $m^\flat : A \to \wp(\sigma_R(s))$ such that $m^\flat$ is in $\mathcal{M}_{s,\varphi}$, we have that $m^\flat \trianglelefteq m$, and also there is no marking $m' \in \mathcal{M}_{s,\varphi}$ such that $m' \neq m^\flat$ and $m' \trianglelefteq m^\flat$. It follows in particular that $m^\flat$ is a minimal element of $\mathcal{M}_{s,\varphi}$. ◂

With the next proposition, we express the fact that $\exists$ can be always assumed to play according to the idea of rationality explained above.

Proposition 2.13. *Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ be an MSO-automaton and $\mathbb{T}$ a tree. The following are equivalent.*

1. *Player $\exists$ has a surviving strategy in $\mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$.*
2. *Player $\exists$ has a surviving strategy in $\mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$ which is minimal for $\mathbb{A}$ and $\mathbb{T}$.*

The same equivalence holds with 'winning' in place of 'surviving'.

Proof We confine ourselves to the case of surviving strategies. It is immediate to check that the very same argument shows the statement also for the case of winning strategies. Direction $(2 \Rightarrow 1)$ is immediate, so we focus on direction $(1 \Rightarrow 2)$. Let f be a surviving strategy for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$. We want to define a strategy $f^\flat$ which is both minimal and surviving for $\exists$ in $\mathcal{G}$. The definition of $f^\flat$ is provided for each stage of the construction of a match $\pi^\flat$ of $\mathcal{G}$, while maintaining an f-conform shadow match π of $\mathcal{G}$. For each round z that is played in $\pi^\flat$ and π, we want to keep the following condition.

The basic position occurring in the match $\pi^\flat$ also occurs in the shadow match π at the current round.	(‡)

The match $\pi^\flat$ is initialized at position (a_I,s_I). We start the construction of the shadow match π also from position (a_I,s_I), so that condition (‡) holds for the first round. Observe that, since f is a surviving strategy for $\exists$ in $\mathcal{G}$, then π is indeed (the initial part of) an f-conform match.

Inductively, consider the case of some round z_i where we are playing from the same basic position $(a,s) \in A \times T$ both in π and $\pi^\flat$. Since π is f-conform, the strategy f suggests to $\exists$ a marking $m : A \to \wp(\sigma_R(s))$ that makes $\Delta(a,\sigma_C(s))$ true. This means that there is some disjunct φ of $\Delta(a,\sigma_C(s))$ that is true in $(\sigma_R(s),m)$. By remark 2.12, there is also some marking $m^\flat : A \to \wp(\sigma_R(s))$ such that $m^\flat \trianglelefteq m$ and $m^\flat$ is a minimal element of $\mathcal{M}_{s,\varphi}$. By definition $m^\flat$ makes φ true in $\sigma_R(s)$, meaning that it also makes $\Delta(a,\sigma_C(s))$ true in $\sigma_R(s)$. Therefore $m^\flat$ is a legitimate move for $\exists$ in $\pi^\flat$ and we let it be the suggestion of the strategy $f^\flat$.

If $\forall$ gets stuck in $\pi^\flat$, then $\exists$ immediately wins the match. Otherwise, let (b,t) be the next basic position picked by $\forall$ in $\pi^\flat$. Since $m^\flat \trianglelefteq m$, then t is in $m(b)$. This means that the position (b,t) is an admissible move for $\forall$ also in the shadow match π. By letting $\forall$ choose position (b,t) in π, we can keep the same basic position in $\pi^\flat$ and π at round z_{i+1} and condition $(\ddagger)$ is maintained for one more stage of the construction.

The strategy $f^\flat$ that we just defined is minimal for $\mathbb{A}$ and $\mathbb{T}$, according to definition 2.11. Moreover, in each round that is played in the match $\pi^\flat$, the marking suggested by $f^\flat$ is a legitimate move for $\exists$, meaning that she never gets stuck in $\pi^\flat$. Since $\pi^\flat$ has been constructed as an arbitrary $f^\flat$-conform match, this suffices to show that $f^\flat$ is a surviving strategy for $\exists$. □

We are now ready to supply a proof of proposition 2.10.

Proof of proposition 2.10 Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be a non-deterministic *MSO*-automaton, $\mathbb{T}$ a tree, and suppose that $\exists$ has a surviving strategy f in $\mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$. By proposition 2.13 we can assume that f is *minimal*. Suppose by way of contradiction that f is not functional. Let $(a,s) \in Dom(f)$ and $t \in \sigma_R(s)$ be such that the marking $m = f(a,s)$ assigns two distinct states $a_1, a_2 \in A$ to t. Since f is surviving and minimal then there is some disjunct $\varphi \in FBF^+(A)$ of $\Delta(a,\sigma_C(s))$ such that $(\sigma_R(s),m) \models \varphi$ and m is minimal among the markings that make φ true in $\sigma_R(s)$.

Let $m^\flat : A \to \wp(\sigma_R(s))$ be some shrinking of m as in remark 1.27. By definition, $m^\flat \trianglelefteq m$ and $(\sigma_R(s), m^\flat) \models \varphi$. By the particular syntactic shape of φ, the node t witnesses some variable y occurring either in the existential or the universal part of φ, with associated positive A-type $\tau_S^+(y)$. By definition, $m^\flat$ assigns to t exactly the states in the set S. Since φ is in *functional* basic form, then S is either empty or a singleton, meaning that t is in $m^\flat(b)$ for *at most one* $b \in A$. Since m assigns both a_1 and a_2 to t, it follows that $m^\flat \neq m$, contradicting the assumption that m is minimal in $\mathcal{M}_{s,\varphi}$. Therefore f is a functional strategy and this completes the proof of the main statement. □

2.3 The Simulation Theorem

Our next goal is to show that every *MSO*-automaton can be assumed to be non-deterministic. This statement, which is called the *Simulation Theorem*, can be considered the main technical result on *MSO*-automata.

Theorem 2.14 (Simulation Theorem, [33]). *Given an MSO-automaton $\mathbb{A}$, there is an effectively constructible non-deterministic MSO-automaton $\mathbb{A}^{P\wp}$ such that*

$$\mathbb{A} \equiv \mathbb{A}^{P\wp}.$$

The transformation of $\mathbb{A}$ into a non-deterministic automaton $\mathbb{A}^{P\wp}$ is essentially performed in two steps.

1. First $\mathbb{A}$ is transformed into an equivalent non-deterministic automaton $\mathbb{A}^{\wp}$ with a non-parity acceptance condition.

2. Then $\mathbb{A}^{\wp}$ is transformed into an equivalent non-deterministic *MSO*-automaton $\mathbb{A}^{P\wp}$.

The conceptual core of the construction lies in the first step. As a side remark, notice that the automaton $\mathbb{A}^{\wp}$ is not 'officially' an *MSO*-automaton, because of the non-parity acceptance condition. However, it makes sense to say that such automaton is non-deterministic, this property depending only on the type of the transition function. Before introducing further technical details, we gather some intuitions underlying the construction of $\mathbb{A}^{\wp}$. For the purpose of giving the transition function of $\mathbb{A}^{\wp}$, a key observation is that each sentence $\varphi \in BF^+(A)$ can be seen as a sentence in $FBF^+(\wp(A))$, modulo a 'change of base' from A to $\wp(A)$. To be more precise, suppose that $\tau_S^+(x)$ is a positive A-type occurring in φ, for some non-empty $S \subseteq A$.

$$\tau_S^+(x) = \bigwedge_{b \in S} b(x)$$

Now we may think of $\tau_S^+(x)$ as a $\wp(A)$−type instead of an A-type.

$$\tau_S^+(x) = S(x)$$

Intuitively, what we did is to 'encapsulate' the conjunction into S. The resulting formula $S(x)$ is a positive $\wp(A)$−type. By applying this procedure for each non-empty positive A-type occurring in φ, we obtain a sentence $\varphi^{\wp}$ in $FBF^+(\wp(A))$.

Definition 2.15 (Change of base)**.** Fix a set A of unary predicates. Let $\varphi \in BF^+(A)$ be a sentence in basic form depending on sequences $B_1 \ldots B_k$ and $C_1 \ldots C_j$ of subsets of A. For each subset S in the sequence, we define the formula $\tau_S^{\wp}(x)$ as follows:

$$\tau_S^{\wp}(x) \quad := \quad \begin{cases} S(x) & \text{If } S \neq \varnothing \\ \top & \text{Otherwise} \end{cases}$$

We denote with $\varphi^{\wp}$ the sentence given as follows.

$$\varphi^{\wp} \quad = \quad \exists x_1 \ldots x_k \, (\mathit{diff}(\bar{x}) \wedge \bigwedge_{1 \leq i \leq k} \tau_{B_i}^{\wp}(x_i) \wedge \forall z \, (\mathit{diff}(\bar{x}, z) \rightarrow \bigvee_{1 \leq l \leq j} \tau_{C_l}^{\wp}(z)))$$

◁

Note that, given $\varphi \in BF^+(A)$, the sentence $\varphi^{\wp}$ is in $FBF^+(\wp(A))$. Definition 2.15 provides the tools to characterize a *powerset construction* on $\mathbb{A}$. The idea would be to construct an automaton $\mathbb{A}^{\sharp}$ with $\wp(A)$ as set of states and a transition function $\Delta^{\sharp} : \wp(A) \times C \rightarrow SLatt(FBF^+(\wp(A)))$, obtained from the transition function $\Delta : A \times C \rightarrow FO^+(A)$ of $\mathbb{A}$ by using definition 2.15. The automaton $\mathbb{A}^{\sharp}$ would be non-deterministic according to definition 2.9.

This is *almost* the construction we are going to define. In fact we still need to specify the acceptance conditions of $\mathbb{A}^{\sharp}$, in such a way that it is equivalent to the original automaton $\mathbb{A}$. It turns out that the choice of $\wp(A)$ as carrier is too coarse for this purpose.

In order to motivate this statement, let $\mathbb{T}$ be a tree. The idea is that a match $\pi^{\sharp}$ of $\mathcal{A}(\mathbb{A}^{\sharp}, \mathbb{T})$ represents a bundle of matches $\pi_1, \ldots, \pi_k$ of $\mathcal{A}(\mathbb{A}, \mathbb{T})$, which $\exists$ and $\forall$ play in parallel on the same branch of $\mathbb{T}$. We want to define the winning conditions of $\mathcal{A}(\mathbb{A}^{\sharp}, \mathbb{T})$ in such a way that $\exists$ wins $\pi^{\sharp}$ if and only if she manages to win each match $\pi_1, \ldots, \pi_k$.

The problem is how to recover the structure of the various matches $\pi_1, \ldots, \pi_k$ given the sole information of $\pi^{\sharp}$. Suppose that $(B_1, s_1), (B_2, s_2), \ldots, (B_n, s_n) \ldots$ are the basic positions visited along $\pi^{\sharp}$, with (B_i, s_i) in $\wp(A) \times T$ for each $i < \omega$. Every match $\pi_j \in \{\pi_1, \ldots, \pi_k\}$ is associated with a sequence of basic positions $(b_1, s_1), (b_2, s_2), \ldots, (b_n, s_n), \ldots$ such that b_i is in B_i for each $i < \omega$. However, we do not know *which* b_i is the element that we should pick in B_i to recover the corresponding position of π_j: more than one choice is possible.

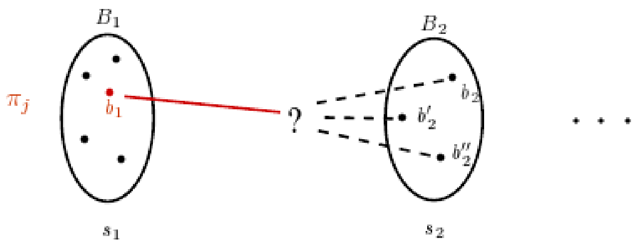

Figure 2.1: the problem of recovering the position of π_j which is associated with node s_2.

A possible way to deal with this problem is to assign a *tag* to each state $b_i \in B_i$, carrying the information of which is the 'precedent state' of $\mathbb{A}$ occurring in the corresponding match π_j. For instance, if $b_i \in B_i$ and $b_{i-1} \in B_{i-1}$ are such that $(b_{i-1}, s_{i-1}), (b_i, s_i)$ are basic positions visited in π_j at rounds $i-1$ and i, then we put the tag b_{i-1} to b_i.

This transformation turns every macro-state $B \in \wp(A)$ into a *binary relation* $R \in \wp(A \times A)$, where $(b_1, b_2) \in R$ stands for the state $b_2 \in A$ equipped with tag $b_1 \in A$. The powerset construction on $\mathbb{A}$ must be modified accordingly, leading to a *refined powerset construction* on $\mathbb{A}$, which we denote with $A^{\wp}$.

Before the formal definition of $A^{\wp}$, we need some preliminary work formalizing the intuitions given above. First, we introduce the notion of *trace*. A sequence of positions $(R_1, s_1), (R_2, s_2), \ldots, (R_n, s_n) \ldots$, with $(R_i, s_i) \in (\wp(A \times A)) \times T$, will induce a set of traces, indicating which matches of $\mathbb{A}$ on $\mathbb{T}$ are associated with the sequence.

Definition 2.16 ([31])**.** Let A be a finite set of states and let $\rho \in (\wp(A \times A))^{\omega}$ be a $\wp(A \times A)$–stream.

$$\rho \quad := \quad R_0, R_1, \ldots, R_n, \ldots$$

A *trace* $\alpha \in A^{\omega}$ through ρ is an A–stream such that $a_i R_{i+1} a_{i+1}$ for all $i < \omega$.

$$\alpha \quad := \quad a_0, a_1, \ldots, a_n, \ldots$$

◁

Definition 2.17 ([31])**.** Let A be a finite set of states and $\Omega : A \to \omega$ a parity map. We say that a trace $\alpha \in A^\omega$ is *good* if the minimum parity occurring infinitely often along α is even, and *bad* otherwise. The set $NBT_\Omega \subseteq (\wp(A \times A))^\omega$ is defined as

$$NBT_\Omega \quad := \quad \{\rho \in (\wp(A \times A))^\omega \mid \text{every trace through } \rho \text{ is good}\}.$$

◁

The last component we need to consider before giving the formal definition of $\mathbb{A}^\wp$ is the transition function. As we mentioned, the idea is to perform a 'change of base' on the transition function of the original automaton $\mathbb{A}$, passing from first-order sentences on signature A to first-order sentences on signature $\wp(A)$. Now we need to shift the same argument to the signature $\wp(A \times A)$, associated with the binary relations on A which will form the carrier of $\mathbb{A}^\wp$. For this purpose we introduce an intermediate step, transforming first-order sentences on signature A into first-order sentences on signature $A \times A$.

Definition 2.18 ([31])**.** Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ be an *MSO*-automaton. Fix $a \in A$ and $c \in C$. The sentence $\Delta^\star(a,c)$ is defined as

$$\Delta^\star(a,c) \quad := \quad \Delta(a,c)[(a,b) \setminus b \mid b \in A],$$

where $\Delta(a,c)[(a,b) \setminus b \mid b \in A]$ denotes the sentence in $FO^+(A \times A)$ obtained by replacing each occurrence of an unary predicate $b \in A$ in $\Delta(a,c)$ with the unary predicate $(a,b) \in A \times A$. ◁

The next step is to put together definition 2.15 and 2.18 to characterize a transition function ranging over sentences on signature $\wp(A \times A)$.

Definition 2.19. Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ be an *MSO*-automaton. Let $c \in C$ be a label and $R \in \wp(A \times A)$ a binary relation on A. By proposition 1.26 there is a sentence $\Psi'_{R,c} \in SLatt(BF^+(A \times A))$ such that

$$\bigwedge_{a \in Ran(R)} \Delta^\star(a,c) \quad \equiv \quad \Psi'_{R,c}.$$

We define $\Psi_{R,c}$ to be the sentence $(\Psi'_{R,c})^\wp$, where the translation $(-)^\wp$ is given as in definition 2.15. ◁

Observe that $\Psi_{R,c}$ is an element of $SLatt(FBF^+(\wp(A \times A)))$. Now we have all the ingredients to provide the definition of $\mathbb{A}^\wp$.

Definition 2.20. Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ be an *MSO*-automaton. The automaton $\mathbb{A}^\wp = \langle A^\wp, a_I^\wp, \Delta^\wp, Acc\rangle$ is defined as follows.

$$\begin{aligned} A^\wp &:= \wp(A \times A) \\ a_I^\wp &:= \{a_I, a_I\} \\ \Delta^\wp(R,c) &:= \Psi_{R,c} \\ Acc &:= NBT_\Omega \end{aligned}$$

Here $\Psi_{R,c}$ is given according to proposition 2.21 and NBT_Ω is given according to definition 2.17. The automaton $\mathbb{A}^\wp$ is called the *refined powerset construction on* $\mathbb{A}$. ◁

The next step is to show that $\mathbb{A}^\wp$ is indeed equivalent to the original automaton $\mathbb{A}$. For this purpose, it is convenient first to prove two lemmata, associating the one-step behaviors of the automata $\mathbb{A}$ and $\mathbb{A}^\wp$.

Proposition 2.21. *Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ be an MSO-automaton and $\mathbb{A}^\wp = \langle A^\wp, a_I^\wp, \Delta^\wp, Acc\rangle$ its refined powerset construction. Let $R \in A^\wp$ be a binary relation, $\mathbb{T}$ a tree and s a node of $\mathbb{T}$. Suppose that, for each $a \in Ran(R)$, there is a marking $m_a : A \to \wp(\sigma_R(s))$ such that $(\sigma_R(s), m_a) \models \Delta(a, \sigma_C(s))$. The following two statements hold.*

1. *There is a marking $m^\star : A \times A \to \wp(\sigma_R(s))$ such that*
   a) $(\sigma_R(s), m^\star) \models \bigwedge_{a \in Ran(R)} \Delta^\star(a, \sigma_C(s))$;
   b) *for each state $a \in Ran(R)$, node $t \in \sigma_R(s)$, state $b \in A$ such that $t \in m^\star(a,b)$, we have that $t \in m_a(b)$.*
2. *There is a marking $m^\wp : A^\wp \to \wp(\sigma_R(s))$ such that*
   a) $(\sigma_R(s), m^\wp) \models \Delta^\wp(R, \sigma_C(s))$;

b) *for each state* $Q \in A^{\wp}$*, node* $t \in \sigma_R(s)$ *such that* $t \in m^{\wp}(Q)$*, state* $b \in Ran(Q)$*, there is some* $a \in Ran(R)$ *such that* $t \in m_a(b)$*.*

Proof Fix some $R \in A^{\wp}$. For each $a \in Ran(R)$, let $m_a : A \to \wp(\sigma_R(s))$ be such that $(\sigma_R(s), m_a) \models \Delta(a, \sigma_C(s))$. We define a marking $m^\star : A \times A \to \wp(\sigma_R(s))$ by putting

$$m^\star(a,b) \quad := \quad m_a(b). \tag{2.2}$$

For statement 1.a, fix some $a \in Ran(R)$. It is easy to check that $(\sigma_R(s), m^\star) \models \Delta^\star(a, \sigma_C(s))$, by definition of $\Delta^\star$ and the monotonicity property of $\Delta^\star(a, \sigma_C(s)) \in FO^+(A \times A)$ as in remark 1.24. Statement 1.b is immediate by definition of $m^\star$. In order to show statement 2, we define a marking $m^{\wp} : \wp(A \times A) \to \wp(\sigma_R(s))$ by putting

$$m^{\wp}(Q) \quad := \quad \bigcap_{(a,b) \in Q} m^\star(a,b). \tag{2.3}$$

For statement 2.a, recall that $\Delta^{\wp}(R, \sigma_C(s))$ is the sentence $\Psi_{R,c} \in SLatt(FBF^+(\wp(A \times A)))$ as in definition 2.15. Then it is not hard to check that statement 2.a follows by statement 1.a and the definition of $m^\star$ according to (2.2).

For statement 2.b, the main observation is that $m^{\wp}(Q) \neq \varnothing$ implies that $Dom(Q) \subseteq Ran(R)$, for each $Q \in A^{\wp}$. In order to see that, just observe that, by (2.2), for any pair $(a,b) \in A \times A$, the marking $m^\star(a,b)$ is defined only if $a \in Ran(R)$. Then it suffices to show that, for each state $Q \in A^{\wp}$, node $t \in \sigma_R(s)$ such that $t \in m^{\wp}(Q)$, for each $(a,b) \in Q$, it holds that $t \in m_a(b)$. This is clearly the case by statement 1.b and definition of $m^{\wp}$ as in (2.3). □

Proposition 2.22. *Let* $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ *be an MSO-automaton and* $\mathbb{A}^{\wp} = \langle A^{\wp}, a_I^{\wp}, \Delta^{\wp}, Acc \rangle$ *its refined powerset construction. Let* $R \in A^{\wp}$ *be a binary relation,* $\mathbb{T}$ *a tree and* s *a node of* $\mathbb{T}$*. Suppose that there is a marking* $m^{\wp} : A^{\wp} \to \wp(\sigma_R(s))$ *such that* $(\sigma_R(s), m^{\wp}) \models \Delta^{\wp}(R, \sigma_C(s))$*. The following two statements hold.*

1. *There is a marking* $m^\star : A \times A \to \wp(\sigma_R(s))$ *such that*
   a) $(\sigma_R(s), m^\star) \models \bigwedge_{a \in Ran(R)} \Delta^\star(a, \sigma_C(s))$*;*
   b) *for each state* $a \in Ran(R)$*, node* $t \in \sigma_R(s)$*, state* $b \in A$ *such that* $t \in m^\star(a,b)$*, there is* $Q \in A^{\wp}$ *such that* $(a,b) \in Q$ *and* $t \in m^{\wp}(Q)$*.*
2. *For each* $a \in Ran(R)$*, there is a marking* $m_a : A \to \wp(\sigma_R(s))$ *such that*
   a) $(\sigma_R(s), m_a) \models \Delta(a, \sigma_C(s))$*;*
   b) *for each node* $t \in \sigma_R(s)$*, state* $b \in A$ *such that* $t \in m_a(b)$*, there is* $Q \in A^{\wp}$ *such that* $b \in Ran(Q)$ *and* $t \in m^{\wp}(Q)$*.*

Proof Fix some $R \in A^{\wp}$ and let $m^{\wp} : \wp(A \times A) \to \wp(\sigma_R(s))$ be a marking such that $(\sigma_R(s), m^{\wp}) \models \Delta^{\wp}(R, \sigma_C(s))$. We define a marking $m^\star : A \times A \to \wp(\sigma_R(s))$ by putting

$$m^\star(a,b) \quad := \quad \bigcup_{(a,b) \in Q} m^{\wp}(Q). \tag{2.4}$$

Also for each $a \in Ran(R)$ we define a marking $m_a : A \to \wp(\sigma_R(s))$ by putting

$$m_a(b) \quad := \quad m^\star(a,b). \tag{2.5}$$

The argument showing statements 1 and 2 on the base of (2.4) and (2.5) is entirely analogous to the one provided for proposition 2.21. □

Proposition 2.23. *Let* $\mathbb{A}$ *be an MSO-automaton and* $\mathbb{A}^{\wp}$ *its refined powerset construction. We have that*

$$\mathbb{A} \quad \equiv \quad \mathbb{A}^{\wp}.$$

Proof Consider an *MSO*-automaton $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ and let $\mathbb{A}^{\wp} = \langle A^{\wp}, a_I^{\wp}, \Delta^{\wp}, NBT_\Omega \rangle$ be its refined powerset construction. In order to show the two directions of the equivalence, we fix a tree $\mathbb{T}$.

($\Rightarrow$) Given a tree $\mathbb{T}$, suppose that $\exists$ has a winning strategy f in $\mathcal{G} = \mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$. Our goal is to provide a winning strategy $f^{\wp}$ to $\exists$ in $\mathcal{G}^{\wp} = \mathcal{A}(\mathbb{A}^{\wp}, \mathbb{T})@(a_I^{\wp}, s_I)$. The definition of $f^{\wp}$ is provided for each stage of the construction of a match $\pi^{\wp}$ of $\mathcal{G}^{\wp}$. While playing $\pi^{\wp}$, player $\exists$ maintains a set $\mathcal{M}$ of f-conform shadow matches. We indicate with $\mathcal{M}_i$ the set $\mathcal{M}$ at round i. Inductively, we will make sure that $\exists$ can keep the following condition for each round z_i that is played in $\pi^{\wp}$ and each match in $\mathcal{M}_i$.

The current basic position in $\pi^\wp$ is of the form $(R,s) \in A^\wp \times T$. For each $a \in Ran(R)$, there is an f-conform shadow match π_a in $\mathcal{M}_i$ at the same round z_i, such that the current basic position in π_a is $(a,s) \in A \times T$.

(‡)

Condition (‡) holds for the initial round, where we initialize the match $\pi^\wp$ at position $(a_I^\wp, s_I)$ and an f-conform match π_{a_I} of $\mathcal{G}$ from position (a_I, s_I). We set $\mathcal{M}_0 = \{\pi_{a_I}\}$.

By inductive hypothesis suppose that we have constructed (the initial part of) the match $\pi^\wp$, with rounds $z_0,\ldots,z_i$. Also, we are provided with a set $\mathcal{M}_i$, where each element of $\mathcal{M}_i$ is (the initial part of) an f-conform shadow match, with rounds $z_0,\ldots,z_i$, such that for each $j \leq i$ condition (‡) is respected by $\mathcal{M}_j$ and $\pi^\wp$. Let $(R,s) \in A^\wp \times T$ be the basic position occurring in $\pi^\wp$ at round z_i. By condition (‡), each $a \in Ran(R)$ is associated with an f-conform match π_a in $\mathcal{M}_i$ at basic position (a,s), from which the strategy f suggests a marking $m_a : A \to \wp(\sigma_R(s))$ such that $(\sigma_R(s), m_a) \vDash \Delta(a, \sigma_C(s))$. We use this assumption to obtain a marking $m^\wp : A^\wp \to \wp(\sigma_R(s))$ as in proposition 2.21. We let $\exists$ choose $m^\wp$ from position (R,s) in $\pi^\wp$.

The marking $m^\wp$ is a legitimate choice for $\exists$ by proposition 2.21, statement 2.a. If $m^\wp(Q) = \varnothing$ for all $Q \in A^\wp$, then $\forall$ gets stuck at the current round and $\exists$ wins $\pi^\wp$. Otherwise, $\forall$ is able to pick a next basic position $(Q,t) \in A^\wp \times T$. By proposition 2.21, statement 2.b, for each $b \in Ran(Q)$, there is some $a \in Ran(R)$ such that $t \in m_a(b)$. Therefore for each $b \in Ran(Q)$ we can select a match $\pi_a \in \mathcal{M}_i$ such that $t \in m_a(b)$. We define $\pi_{a,b}$ as the match π_a extended with the move given by $\forall$ choosing (b,t) as next basic position. We define $\mathcal{M}_{i+1}$ to be the collection of matches $\pi_{a,b}$ for all $b \in Ran(Q)$. In this way we are able to maintain condition (‡) also at round z_{i+1}.

If $\forall$ does not get stuck at some round, then $\pi^\wp$ is an infinite match of $\mathcal{G}^\wp$. The sequence of states visited along the play induces a $A^\wp$–stream ρ given as

$$a_I^\wp, R_1, \ldots, R_n, \ldots.$$

In order to check that $\exists$ wins $\pi^\wp$, it suffices to show that ρ is in NBT_Ω. For this purpose, let α be a trace through ρ of the form

$$a_I, a_1, \ldots, a_n, \ldots.$$

By definition of trace, we have that $a_i \in Ran(R_i)$ for each $i \leq \omega$. Then, by condition (‡), the trace α is the sequence of states visited along an f-conform match of $\mathcal{G}$. Since f is assumed to be winning, the trace α is *good*. Therefore ρ is in NBT_Ω.

(⇐) We confine ourself to a sketch, since the argument follows the same line of reasoning of the one provided for the converse direction. Suppose that $\exists$ has a winning strategy $f^\wp$ in $\mathcal{G}^\wp = \mathcal{A}(\mathbb{A}^\wp, \mathbb{T})@(a_I^\wp, s_I)$. Our goal is to provide a winning strategy f to $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$. We define f for each stage of the construction of a match π of $\mathcal{G}$, while maintaining an $f^\wp$-conform shadow match $\pi^\wp$ of $\mathcal{G}^\wp$. For each round z that is played in π and $\pi^\wp$, we want to keep the following condition.

The current basic position in π is of the form $(a,s) \in A \times T$ and the current basic position in $\pi^\wp$ is $(R,s) \in A^\wp \times T$, for some $R \in A^\wp$ with $a \in Ran(R)$.

(‡)

This condition clearly holds for the initial round. Inductively, let (a,s) and (R,s) be the positions occurring respectively in π and $\pi^\wp$ at round z_i, with $a \in Ran(R)$. Since $\pi^\wp$ is $f^\wp$-conform, the strategy $f^\wp$ suggests a marking $m^\wp : \wp(A \times A) \to \wp(\sigma_R(s))$ such that $(\sigma_R(s), m^\wp) \vDash \Delta^\wp(R, \sigma_C(s))$. We use this assumption to obtain a marking $m_a : A \to \wp(\sigma_R(s))$ as in proposition 2.22. We let $\exists$ choose m_a from position (a,s) in π.

Analogously to the converse direction, we can show that m_a is a legitimate move for $\exists$. Then either the match π ends at the current round with $\forall$ stuck and $\exists$ winning, or $\forall$ is able to pick a next basic position. Then we can show that condition (‡) can be maintained at round z_{i+1}, by using proposition 2.22. If $\forall$ never gets stuck, the match π is infinite. Let α be the A–stream induced by the sequence of states encountered along the play in π. In order to check that $\exists$ wins π, the key observation is that α is a trace through the $A^\wp$–stream associated with the $f^\wp$-conform match $\pi^\wp$. Since $f^\wp$ is winning for $\exists$ in $\mathcal{G}^\wp$, the trace α is *good*, implying that $\exists$ wins the corresponding match π. □

We observe that the non-deterministic automaton $\mathbb{A}^\wp$ is not an *MSO*-automaton: the missing component is a parity acceptance conditions replacing the condition NBT_Ω. The second part of the proof of theorem 2.14 consists in showing that we can transform $\mathbb{A}^\wp$ into an equivalent non-deterministic *MSO*-automaton $\mathbb{A}^{P\wp}$. The next proposition shows how this can be done easily, provided that NBT_Ω is an ω-regular language (definition 1.31).

Proposition 2.24 ([31])**.** *Given an automaton $\mathbb{Q} = \langle Q, q_I, \Delta_Q, Acc\rangle$ with Δ_Q of type $Q \times C \to SLatt(FBF^+(Q))$, if $Acc \subseteq Q^\omega$ is an ω-regular language then there is an effectively constructible non-deterministic MSO-automaton $\mathbb{Q}^P$ such that $\mathbb{Q} \equiv \mathbb{Q}^P$.*

Proof sketch The argument is the same given in [31], proof of proposition 6.29. We confine ourself to a sketch of the proof.

Let $\mathbb{Q} = \langle Q, q_I, \Delta_Q, Acc\rangle$ be an automaton with Δ of type $Q \times C \to SLatt(FBF^+(Q))$. By assumption Acc is an ω-regular language, meaning that there is a deterministic Q–stream automaton $\mathbb{Z} = \langle Z, z_I, \delta_Z, \Omega_Z\rangle$ with $L(\mathbb{Z}) = Acc$. We define an operation $Shift_z : FBF^+(Q) \to FBF^+(Q \times Z)$ as follows.

$$Shift_z(\varphi) := \varphi[(q, \delta_Z(z,q))/q \mid q \in Q]$$

Then we define an automaton $\mathbb{Q} \odot \mathbb{Z} := \langle Q \times Z, (q_I, z_I), \Delta, \Omega\rangle$ by putting

$$\begin{array}{rcl} \Omega(q,z) & := & \Omega_Z(z), \\ \Delta((q,z),c) & := & \bigvee\{Shift_z(\varphi) \in FBF^+(Q \times Z) \mid \varphi \text{ is a disjunct of } \Delta_Q(q,c)\}. \end{array}$$

It can be easily checked that Δ is of type $(Q \times Z) \times C \to Slatt(FBF^+(Q \times Z))$, whence $\mathbb{Q} \odot \mathbb{Z}$ is a non-deterministic *MSO*-automaton. We let $\mathbb{Q}^P$ be the automaton $\mathbb{Q} \odot \mathbb{Z}$. The proof of the main statement is concluded by showing that indeed $\mathbb{Q} \equiv \mathbb{Q}^P$. □

The following proposition shows that the assumption of proposition 2.24 holds for $\mathbb{A}^{\wp}$.

Proposition 2.25 ([33],[31])**.** *Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ be an MSO-automaton. The language NBT_Ω is ω-regular.*

Proof sketch The argument is the same as the one given in [31], proof of proposition 6.27. We confine ourself to a sketch of the proof. Let $\mathbb{A}^{\wp} = \langle A^{\wp}, a_I^{\wp}, \Delta^{\wp}, NBT_\Omega\rangle$ be the refined powerset construction on $\mathbb{A}$. In order to prove the statement it suffices to define a non-deterministic $A^{\wp}$–stream automaton $\mathbb{Z}$ accepting exactly the $A^{\wp}$–streams containing a bad trace. By closure properties of stream automata we can take a deterministic $A^{\wp}$–stream automaton $\mathbb{Z}'$ which is equivalent to $\mathbb{Z}$ and then take an automaton $\overline{\mathbb{Z}'}$ accepting the complement of $L(\mathbb{Z}')$. The automaton $\overline{\mathbb{Z}'}$ will be again a deterministic $A^{\wp}$–stream automaton and it is immediate to see that it accepts exactly the infinite $A^{\wp}$–streams where every trace is good.

For the definition of $\mathbb{Z}$, we essentially rely on the parity of the original automaton $\mathbb{A}$. More precisely, let $z_I \notin A$ be a state. We define an automaton $\mathbb{Z} = \langle Z, z_I, \Delta_Z, \Omega_Z\rangle$ by putting

$$\begin{array}{rcl} Z & := & A \cup \{z_I\} \\ \Delta_Z(a,R) & := & \begin{cases} Ran(R) & \text{If } a = z_I \\ R[a] & \text{Otherwise} \end{cases} \\ \Omega_Z(a) & := & \begin{cases} 0 & \text{If } a = z_I \\ \Omega(a)+1 & \text{Otherwise} \end{cases} . \end{array}$$

For the (easy) proof that $L(\mathbb{Z}) = \overline{L(\mathbb{A}^{\wp})}$ we refer to [31], proposition 6.27. □

Now we are ready to prove the Simulation Theorem stated at the beginning of this section.

Proof of Theorem 2.14 Let $\mathbb{A}^{\wp}$ be the refined powerset construction on $\mathbb{A}$ obtained as in 2.20. By proposition 2.25, we can apply the construction of proposition 2.24 to $\mathbb{A}^{\wp}$. We obtain a non-deterministic *MSO*-automaton $\mathbb{A}^{P\wp}$ such that the following holds.

$$\begin{array}{rcll} \mathbb{A} & \equiv & \mathbb{A}^{\wp} & \text{(proposition 2.23)} \\ & \equiv & \mathbb{A}^{P\wp} & \text{(proposition 2.24)} \end{array}$$

This concludes the proof of the theorem. □

2.4 From *MSO*-Formulae to *MSO*-Automata

The Simulation Theorem allows us to show the main result of this chapter, namely that *MSO*-automata characterize *MSO* in the sense of (2.1).

Theorem 2.26 ([33])**.** *For every $\varphi \in MSO$, there is an effectively constructible MSO-automaton $\mathbb{A}_\varphi$ such that*

for any tree $\mathbb{T}$, $\mathbb{T} \vDash \varphi$ iff $\mathbb{A}_\varphi$ accepts $\mathbb{T}$.

The proof will proceed by induction on φ. The crux of the matter is the inductive case of the existential quantifier, which is handled with a projection construction for *MSO*-automata. The hard part part of the proof, namely that every *MSO*-automaton can be assumed to be non-deterministic, has been shown in section 2.3. Once we are allowed to work under this assumption, the remaining part of the projection construction is relatively easy. The next step is to define a closure operation on tree languages, corresponding to the semantics of *MSO* existential quantification.

Definition 2.27. Given a tree $\mathbb{T}$ and a propositional letter p (not necessarily in P), we refer to the notion of *p-variant of* $\mathbb{T}$ as defined in section 1.2. Recall that, since $\mathbb{T}$ is a C-labeled tree, then any p-variant $\mathbb{T}^p$ of $\mathbb{T}$ is a $\wp(P \cup \{p\})$-labeled tree.

Let $\mathcal{L}$ be a tree language. The *projection* of $\mathcal{L}$ over p is the language $\exists p.\mathcal{L}$ defined as

$$\exists p.\mathcal{L} = \{\mathbb{T} \mid \text{ there is a } p\text{-variant } \mathbb{T}^p \text{ of } \mathbb{T} \text{ such that } \mathbb{T}^p \in \mathcal{L}\}.$$

Let $\mathcal{C}$ be a class of tree languages. $\mathcal{C}$ is *closed under projection over p* if, for any language $\mathcal{L}$ in $\mathcal{C}$, also the language $\exists p.\mathcal{L}$ is in $\mathcal{C}$. ◁

Definition 2.28 (Projection Construction)**.** Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ be a non-deterministic *MSO*-automaton on alphabet $\wp(P \cup \{p\})$. We define the automaton $\exists p.\mathbb{A} = \langle A, a_I, \Delta^{\exists}, \Omega\rangle$ *on alphabet C* by putting

$$\Delta^{\exists}(a,c) \quad := \quad \Delta(a,c) \vee \Delta(a, c \cup \{p\}).$$

The automaton $\exists p.\mathbb{A}$ is a non-deterministic *MSO*-automaton which we call the *projection construction of* $\mathbb{A}$ *over p*. ◁

Proposition 2.29. *Given a letter p and a non-deterministic MSO-automaton* $\mathbb{A}$ *on alphabet* $\wp(P \cup \{p\})$*, let* $\exists p.\mathbb{A}$ *be the projection construction of* $\mathbb{A}$ *over p, on alphabet C. The following holds.*

$$L(\exists p.\mathbb{A}) \;=\; \exists p.L(\mathbb{A}) \tag{2.6}$$

Proof Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ be an *MSO*-automaton on alphabet $\wp(P \cup \{p\})$ and $\exists p.\mathbb{A}$ be the projection construction of $\mathbb{A}$ over p. Given a tree $\mathbb{T}$, we want to show that

$$\exists p.\mathbb{A} \text{ accepts } \mathbb{T} \quad \text{iff} \quad \text{there is a } p\text{-variant } \mathbb{T}^p \text{ of } \mathbb{T} \text{ such that } \mathbb{A} \text{ accepts } \mathbb{T}^p.$$

($\Rightarrow$) Let f be a winning strategy for $\exists$ in $\mathcal{G}^{\exists} = \mathcal{A}(\exists p.\mathbb{A}, \mathbb{T})@(a_I, s_I)$. Since $\exists p.\mathbb{A}$ is non-deterministic, by proposition 2.10, we can assume f to be *functional*. Let $\mathbb{T}_f$ and $\pi_2^f : T_f \to T$ be respectively the tree representation of f and its projection function defined according to remark 2.4. Since f is functional, we can assume π_2^f to be 1-1. This means that the inverse function of π_2^f is well-defined. This is a function $(\pi_2^f)^{-1} : Ran(\pi_2^f) \to T_f$ mapping each node $s \in Ran(\pi_2^f)$ to a basic position $(a,s) \in T_f$. By definition of π_2^f, the position (a,s) is *unique* for s, in the sense that no other basic position with s as second component appears in $\mathbb{T}_f$. In other words, the fact that π_2^f is 1-1 allows us to assume the following property:

for each node $s \in T$, if s is visited along the play of some f-conform match of $\mathcal{G}^{\exists}$, then there is a unique state $a_s \in A$, such that f guarantees that s will be only visited with $\exists p.\mathbb{A}$ in state a_s.	(†)

The idea is to let f itself suggest a p-variant $\mathbb{T}^p$ of $\mathbb{T}$, by using for each node of $\mathbb{T}$ the information provided by property (†). Given a node $s \in Ran(\pi_2^f)$, let $a_s \in A$ be the first projection of $(a_s, s) = (\pi_2^f)^{-1}(s)$. We define a set X_p of nodes of $\mathbb{T}$ as follows.

$$X_p \quad := \quad \{s \in Ran(\pi_2^f) \mid (\sigma_R(s), f(a_s,s)) \vDash \Delta(a_s, \sigma_C(s) \cup \{p\})\} \tag{2.7}$$

Since the strategy f is winning, $f(a_s,s)$ is a marking such that

$$(\sigma_R(s), f(a_s,s)) \vDash \Delta(a, \sigma_C(s)) \vee \Delta(a, \sigma_C(s) \cup \{p\}).$$

Intuitively, X_p collects all nodes $s \in Ran(\pi_2^f)$ whose label $\sigma_C(s)$ is 'considered as' $\sigma_C(s) \cup \{p\}$ by f. We let $\mathbb{T}^p$ be the p-variant of $\mathbb{T}$ induced by labeling with p exactly the nodes in X_p. The proof is completed by the showing the following claim.

CLAIM 1. The strategy f is winning for $\exists$ in $\mathcal{A}(\mathbb{A}, \mathbb{T}^p)@(a_I, s_I)$.

PROOF OF CLAIM The idea is to construct an f-conform match π of $\mathcal{G} = \mathcal{A}(\mathbb{A}, \mathbb{T}^p)@(a_I, s_I)$, while maintaining an f-conform shadow match $\pi^\exists$ of $\mathcal{G}^\exists$. For each round z, we want that the same basic position $(a,s) \in A \times T$ is visited both in π and $\pi^\exists$.

Let us first show why $\exists$ is guaranteed to win π if this condition can be maintained. Since $\pi^\exists$ is f-conform and f is winning for $\exists$ in $\mathcal{G}^\exists$, either $\forall$ gets stuck in $\pi^\exists$ or $\pi^\exists$ is infinite and the minimum parity occurring infinitely often is even. If the former is the case, then it is easy to check that $\forall$ also gets stuck in π, because the same markings are suggested to $\exists$ in both games. If the latter is the case, observe that the automaton $\exists p.\mathbb{A}$ has the same carrier A and the same parity map Ω of $\mathbb{A}$, meaning that also π is infinite and the minimum parity occurring infinitely often is even. This suffices to prove that $\exists$ wins π and then f is winning for $\exists$ in $\mathcal{G}$.

It remains to show that for each round we can maintain the same basic positions in π and $\pi^\exists$. This is clearly the case for the initial round, where we initialize both matches at position (a_I, s_I). Since $(a_I, s_I) \in Win_\exists(\mathcal{G}^\exists)$, then $\pi^\exists$ is (the initial part of) an f-conform match, as requested. Inductively, suppose that we are at round z_i and both matches visit the same position (a,s). Let $m_{a,s} : A \to \wp(\sigma_R(s))$ be the marking suggested by f from position (a,s). If we can prove that $m_{a,s}$ is a legitimate move for $\exists$ in π, then $\sigma_R(s)$ is marked by $m_{a,s}$ both in π and in π', meaning that we can maintain the same positions in the two matches at round z_{i+1}.

Therefore the only thing that is left to show is that $m_{a,s}$ is a legitimate move for $\exists$ in π. Since f is winning in $\mathcal{G}^\exists$, then we have that

$$(\sigma_R(s), m_{a,s}) \models \Delta(a, \sigma_C(s)) \vee \Delta(a, \sigma_C(s) \cup \{p\}), \tag{2.8}$$

where $\sigma_C : T \to C$ is the labeling function of $\mathbb{T}$. What we need to show is that

$$(\sigma_R(s), m_{a,s}) \models \Delta(a, \sigma_C^p(s)), \tag{2.9}$$

where $\sigma_C^p : T \to \wp(P \cup \{p\})$ is the labeling function of $\mathbb{T}^p$. For this purpose, we distinguish two cases.

1. First suppose that $p \in \sigma_C^p(s)$. By definition of $\mathbb{T}^p$ we have that $s \in X_p$, where X_p is defined according to (2.7). By definition of X_p the following holds:

$$(\sigma_R(s), m_{a,s}) \models \Delta(a, \sigma_C(s) \cup \{p\}). \tag{2.10}$$

   By definition of p-variant we have that σ_C^p agrees with σ_C on all propositional letters in P but p, meaning that $\sigma_C^p(s) = \sigma_C(s) \cup \{p\}$. Hence $\Delta(a, \sigma_C^p(s)) = \Delta(a, \sigma_C(s) \cup \{p\})$ and (2.9) just follows from (2.10).

2. In the remaining case we have that $p \notin \sigma_C^p(s)$. By definition of $\mathbb{T}^p$ this means that $s \notin X_p$. By inductive hypothesis f is defined on (a,s), meaning that (a,s) is a node of $\mathbb{T}_f$ and therefore $s \in Ran(\pi_f^2)$. This means that the reason why s is not in X_p according to (2.7) is that

$$(\sigma_R(s), m_{a,s}) \not\models \Delta(a, \sigma_C(s) \cup \{p\}). \tag{2.11}$$

   However, the marking $m_{a,s}$ suggested by f must be legitimate in $\mathcal{G}^\exists$, meaning that it makes the other disjunct of $\Delta^\exists(a, \sigma_C(s))$ true. Therefore

$$(\sigma_R(s), m_{a,s}) \models \Delta(a, \sigma_C(s)). \tag{2.12}$$

   Since we are considering the case in which $p \notin \sigma_C^p(s)$, by definition of p-variant we have that that $\sigma_C^p(s) = \sigma_C(s)$. Hence $\Delta(a, \sigma_C^p(s)) = \Delta(a, \sigma_C(s))$ and (2.9) just follows from (2.12).

For each case we reached the conclusion that the marking $m_{a,s}$ is a legitimate choice for $\exists$ in G. This concludes the proof of claim 1 and of direction $(\Rightarrow)$. ■

$(\Leftarrow)$ Let $\mathbb{T}^p$ be a p-variant of $\mathbb{T}$ such that $\exists$ has a winning strategy f in $\mathcal{A}(\mathbb{A}, \mathbb{T}^p)@(a_I, s_I)$. Using an argument analogous to the one showing claim 1, it is easy to check that f is also winning for $\exists$ in $\mathcal{A}(\exists p.\mathbb{A}, \mathbb{T})@(a_I, s_I)$.

□

In the sequel we present two further constructions for *MSO*-automata. We will use them to handle respectively the inductive case of disjunction and the inductive case of negation in the proof of theorem 2.26.

Proposition 2.30. *Given MSO-automata $\mathbb{A}_1$ and $\mathbb{A}_2$, there is an effectively constructible MSO-automaton $\mathbb{A}_{1,2}$ such that*

$$L(\mathbb{A}_{1,2}) \quad = \quad L(\mathbb{A}_1) \cup L(\mathbb{A}_2).$$

Proof sketch The definition of $\mathbb{A}_{1,2}$ as the *sum* of $\mathbb{A}_1$ and $\mathbb{A}_2$ is given by a very standard construction of automata theory. We refer to [11], lemma 16.5 for a more detailed proof and we confine ourselves to a sketch. Let $\mathbb{A}_1 = \langle A_1, a_I^1, \Delta_1, \Omega_1 \rangle$ and $\mathbb{A}_2 = \langle A_2, a_I^2, \Delta_2, \Omega_2 \rangle$ be two *MSO*-automata. Let $a_I \notin A_1 \cup A_2$ be a state. We define the *MSO*-automaton $\mathbb{A}_{1,2} = \langle A_{1,2}, a_I, \Delta_{1,2}, \Omega_{1,2} \rangle$ by putting

$$
\begin{aligned}
A_{1,2} &:= A_1 \cup A_2 \cup \{a_I\} \\
\Delta(a,c) &:= \begin{cases} \Delta_1(a,c) & \text{if } a \in A_1 \\ \Delta_2(a,c) & \text{if } a \in A_2 \\ \Delta_1(a_I^1,c) \vee \Delta_2(a_I^2,c) & \text{if } a = a_I \end{cases} \\
\Omega_{1,2}(a) &:= \begin{cases} \Omega_1(a) & \text{if } a \in A_1 \\ \Omega_2(a) & \text{if } a \in A_2 \\ 0 & \text{if } a = a_I. \end{cases} .
\end{aligned}
$$

It is easy to check that, for each tree $\mathbb{T}$,

$$\mathbb{A}_{1,2} \text{ accepts } \mathbb{T} \quad \textit{iff} \quad \mathbb{A}_1 \text{ accepts } \mathbb{T} \text{ or } \mathbb{A}_2 \text{ accepts } \mathbb{T}.$$

□

Proposition 2.31. *Given an MSO-automaton $\mathbb{A}$, there is an effectively constructible MSO-automaton $\overline{\mathbb{A}}$ such that*

$$L(\overline{\mathbb{A}}) = \overline{L(\mathbb{A})}.$$

Proof We refer to Appendix A for a proof of this statement. □

We have now all the ingredients to prove the main statement of this section.

Proof of Theorem 2.26 The proof is by induction on φ.

- If φ is of the form $p \sqsubseteq q$, for p and q in P, then the automaton $\mathbb{A}_\varphi$ is provided by example 2.6.
- If φ is of the form $R(p,q)$, for p and q in P, then the automaton $\mathbb{A}_\varphi$ is provided by example 2.5.
- If φ is of the form $\neg\psi$, by inductive hypothesis we have a weak *MSO*-automaton $\mathbb{A}_\psi$ that is equivalent to ψ. Let $\mathbb{A}_{\neg\psi}$ be the *MSO*-automaton obtained by applying proposition 2.31 to $\mathbb{A}_\psi$. The following derivation shows that $\mathbb{A}_{\neg\psi}$ is equivalent to $\neg\psi$.

$$
\begin{array}{lcll}
\mathbb{A}_{\neg\psi} \text{ accepts } \mathbb{T} & \Leftrightarrow & \mathbb{A}_\psi \text{ does not accept } \mathbb{T} & \text{(proposition 2.31)} \\
& \Leftrightarrow & \mathbb{T} \nvDash \psi & \text{(inductive hypothesis)} \\
& \Leftrightarrow & \mathbb{T} \vDash \neg\psi & \text{(semantics of } \mathit{MSO}\text{)}
\end{array}
$$

- If φ is of the form $\psi_1 \vee \psi_2$, by inductive hypothesis we have *MSO*-automata $\mathbb{A}_{\psi_1}$ and $\mathbb{A}_{\psi_2}$ that are equivalent respectively to ψ_1 and ψ_2. Let $\mathbb{A}_\varphi$ be the *MSO*-automaton obtained by applying proposition 2.30 to $\mathbb{A}_{\psi_1}$ and $\mathbb{A}_{\psi_2}$. The following derivation shows that $\mathbb{A}_\varphi$ is equivalent to $\psi_1 \vee \psi_2$.

$$
\begin{array}{lcll}
\mathbb{A}_\varphi \text{ accepts } \mathbb{T} & \Leftrightarrow & \mathbb{A}_{\psi_1} \text{ accepts } \mathbb{T} \text{ or } \mathbb{A}_{\psi_2} \text{ accepts } \mathbb{T} & \text{(proposition 2.30)} \\
& \Leftrightarrow & \mathbb{T} \vDash \psi_1 \text{ or } \mathbb{T} \vDash \psi_2 & \text{(inductive hypothesis)} \\
& \Leftrightarrow & \mathbb{T} \vDash \psi_1 \vee \psi_2 & \text{(semantics of } \mathit{MSO}\text{)}
\end{array}
$$

- If $\varphi = \exists p.\psi$, by inductive hypothesis we have an *MSO*-automaton $\mathbb{A}_\psi$ that is equivalent to ψ. By theorem 2.14 there is a non-deterministic *MSO*-automaton $\mathbb{A}_\psi^{P\wp}$ which is equivalent to $\mathbb{A}_\psi$. Let $\exists p.\mathbb{A}_\psi$ be the *MSO*-automaton obtained from $\mathbb{A}_\psi^{P\wp}$ by proposition 2.29. The following derivation shows that $\exists p.\mathbb{A}_\psi$ is equivalent to $\exists p.\psi$.

$$
\begin{array}{lcll}
\exists p.\mathbb{A}_\psi \text{ accepts } \mathbb{T} & \Leftrightarrow & \text{there is } X_p \subseteq \mathbb{T} \text{ such that } \mathbb{A}_\psi^{P\wp} \text{ accepts } \mathbb{T}[p \mapsto X_p] & \text{(proposition 2.29)} \\
& \Leftrightarrow & \text{there is } X_p \subseteq \mathbb{T} \text{ such that } \mathbb{A}_\psi \text{ accepts } \mathbb{T}[p \mapsto X_p] & \text{(theorem 2.14)} \\
& \Leftrightarrow & \text{there is } X_p \subseteq \mathbb{T} \text{ such that } \mathbb{T}[p \mapsto X_p] \vDash \psi & \text{(inductive hypothesis)} \\
& \Leftrightarrow & \mathbb{T} \vDash \exists p.\psi & \text{(semantics of } \mathit{MSO}\text{)}
\end{array}
$$

□

Remark 2.32. The converse of theorem 2.26 is also provable: for every *MSO*-automaton $\mathbb{A}$, there is a formula $\varphi_{\mathbb{A}} \in \mathit{MSO}$ that is true exactly in the trees that are accepted by $\mathbb{A}$. We refer to [33] for a proof of this statement. ◂

2.5 Coda: Normal Form for Non-Deterministic *MSO*-Automata

We conclude this chapter with a normal form theorem, putting non-deterministic *MSO*-automata in a very convenient shape for the theory that is developed in the sequel of this thesis. In order to motivate our statement, let $\mathbb{A}$ be a non-deterministic *MSO*-automaton and $\mathbb{T}$ a tree. By proposition 2.10 we know that each surviving strategy f for $\exists$ in $\mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$ can be assumed to be *functional*: given a position $(a,s) \in Dom(f)$, the marking m suggested by f assigns *at most one* state $b \in A$ to each node $t \in \sigma_R(s)$. In this section we want to strengthen this condition, by showing that m can be assumed to assign *exactly one* state $b \in A$ to each node $t \in \sigma_R(s)$. As before, the idea is to enforce this property by refining the set of first-order sentences associated with the transition function Δ of $\mathbb{A}$. For this purpose, we introduce the notion of *special* basic form.

Definition 2.33 (Special basic form). Given a set A of unary predicates, let $a_1 \dots a_k$ and $c_1 \dots c_j$ be sequences of *elements* of A, with $j \neq 0$. A sentence $\varphi \in FO^+(A)$ is in *special basic form* if it is of shape

$$\varphi \quad = \quad \exists x_1 \dots \exists x_k \ (diff(\bar{x}) \wedge \bigwedge_{1 \leq i \leq k} a_i(x_i) \wedge \forall z\ (diff(\bar{x},z) \to \bigvee_{1 \leq l \leq j} c_l(z)))$$

We denote with $SBF^+(A)$ the set of all sentences in $FO^+(A)$ which are in special basic form. ◁

Sentences in special basic form are a particular case of sentences in functional basic form. In order to see that, fix some φ in $FBF^+(A)$. By definition, for each positive A-type $\tau_S^+(x)$ occurring in φ, the set S is either empty or a singleton. By saying that φ is in *special* basic form, we rule out the case in which S is empty. In other words, φ does not have 'holes': it follows that, in each A-structure (X,m) where φ is true, each element of X is marked with *at least one* unary predicate from A according to m.

From the perspective of acceptance games, we can associate this situation with the property of a strategy for $\exists$ of being *full*.

Definition 2.34. Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be an *MSO*-automaton and $\mathbb{T}$ a tree. Let f be a strategy for $\exists$ in $\mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$. We say that f is *full* if, for each basic position $(a,s) \in Dom(f)$, the marking $f(a,s)$ assigns to each node $t \in \sigma_R(s)$ *at least one* state $b \in A$. ◁

Remark 2.35. Given a full surviving strategy f for $\exists$ in $\mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$, we observe that all f-conform matches are infinite. Indeed, player $\exists$ never gets stuck because f is surviving. Moreover, player $\forall$ does not get stuck neither: since f is full, from each basic position $(a,s) \in A \times T$ that is visited along the play, the suggested marking m assigns some state to each successor of s. Since $\mathbb{T}$ is a leafless tree, this marking always induces some pair $(b,t) \in A \times T$ that is an admissible choice for player $\forall$.

We can visualize this property in terms of the tree representation $\mathbb{T}_f$ of f. By the fact that f is full and surviving, it is easy to see that each node of $\mathbb{T}$ is visited in some f-conform match. It follows that the projection function $\pi_2^f : T_f \to T$ is *onto*. As we observed before, if f is also functional, then π_2^f is also a 1-1 correspondence between $\mathbb{T}_f$ and $\mathbb{T}$. This means that the property of f of being both full and functional corresponds to the property of π_2^f of being a *bijection* between $\mathbb{T}_f$ and $\mathbb{T}$. ◂

The following statement fixes the relation between sentences in special basic form and full and functional strategies.

Proposition 2.36. *Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be a non-deterministic MSO-automaton with Δ of type $A \times C \to SLatt(SBF^+(A))$. Given a tree $\mathbb{T}$, we can assume that each surviving strategy for $\exists$ in $\mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$ is full and functional.*

Proof Let f be a surviving strategy for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$. Since $\mathbb{A}$ is non-deterministic, then by proposition 2.10 we can assume f to be functional. In order to see that f is also full, let $(a,s) \in Dom(f)$ be a basic position of $\mathcal{G}$ from which f suggests a marking $m : A \to \wp(\sigma_R(s))$ that makes $\Delta(a,\sigma_C(s))$ true in $\sigma_R(s)$. By definition $\Delta(a,\sigma_C(s))$ is a disjunction of sentences in special basic form, implying that there is some disjunct $\varphi \in SBF^+(A)$ which is true in $(\sigma_R(s),m)$. By the particular syntactic shape of φ, each node t in $(\sigma_R(s),m)$ witnesses some variable y, with $b(y)$ a subformula occurring either in the existential or the universal part of φ, for some $b \in A$. This means that m assigns the state b to the node t. Since t was an arbitrary node of $\sigma_R(s)$, it follows that each node of $\sigma_R(s)$ is marked with some state according to m. Since (a,s) was an arbitrary position in $Dom(f)$, it follows that f is indeed a full strategy. □

The observations provided in remark 2.35 make full and functional strategies quite appealing. By proposition 2.36, we can assume each strategy for $\exists$ to be full and functional by bringing non-deterministic *MSO*-automata in a specific normal form, as shown with the next statement.

Proposition 2.37. *For each non-deterministic MSO-automaton $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$, there is an equivalent non-deterministic MSO-automaton $\mathbb{A}^F = \langle A^F, a_I^F, \Delta^F, \Omega^F \rangle$ with Δ^F of type $A^F \times C \to SLatt(SBF^+(A^F))$.*

In order to prove proposition 2.37, the idea is to let $\mathbb{A}^F$ be an automaton based on the same carrier A of $\mathbb{A}$, with the addition of a state $a_\top \notin A$. Roughly, the transition function Δ^F of $\mathbb{A}^F$ will be given by the sentences in functional basic form which are associated with Δ, where each 'hole' has been filled with the new state $a_\top$. We want that $a_\top$ plays the role of a placeholder, in the sense that its presence is irrelevant for the acceptance game of $\mathbb{A}^F$. By this observation we will be able to show $\mathbb{A}$ and $\mathbb{A}^F$ are in fact equivalent.

As a preliminary step towards the construction of $\mathbb{A}^F$, we introduce a procedure transforming a sentence $\varphi \in FBF^+(A)$ in functional basic form into a sentence in special basic form. This translation will depend on a given unary predicate b. The idea is that each first-order variable y that is bound in φ is brought into the scope of at least one unary predicate. If y is a variable that was not in the scope of any unary predicate in φ, then the subformula $b(y)$ is placed in the translation of φ.

Definition 2.38 (b-translation)**.** Fix a set A of unary predicates and an unary predicate b (not necessarily in A). Let $\varphi \in FBF^+(A)$ be a sentence in functional basic form, depending on sequences $B_1 \dots B_k$ and $C_1 \dots C_j$ of subsets of A. For each subset S in the sequence, we define the formula $\tau_S^b(x)$ as follows:

$$\tau_S^b(x) \quad := \quad \begin{cases} \tau_S(x) & \text{If } S \neq \varnothing \\ b(x) & \text{Otherwise.} \end{cases}$$

The sentence $\varphi^b \in SBF^+(A \cup \{b\})$ is defined as follows.

$$\varphi^b \quad := \quad \exists x_1 \dots x_k \,(\mathit{diff}(\bar{x}) \wedge \bigwedge_{1 \le i \le k} \tau_{B_i}^b(x_i) \wedge \forall z\, (\mathit{diff}(\bar{x}, z) \to \bigvee_{1 \le l \le j} \tau_{C_l}^b(z)))$$

In the particular case in which $j = 0$, the subformula $\bigvee_{1 \le l \le j} \tau_{C_l}^b(z)$ is replaced with $b(z)$. ◁

Definition 2.39. Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be a non-deterministic *MSO*-automaton and $a_\top$ be a state that is not in A. We define an automaton $\mathbb{A}^F = \langle A^F, a_I, \Delta^F, \Omega^F \rangle$ by putting

$$A^F \quad := \quad A \cup \{a_\top\}$$

$$\Delta^F(a,c) \quad := \quad \begin{cases} \forall x\, a_\top(x) & \text{If } a = a_\top \\ \bigvee \{\varphi^{a_\top} \in SBF^+(A^F) \mid \varphi \text{ is a disjunct of } \Delta(a,c)\} & \text{Otherwise} \end{cases}$$

$$\Omega^F(a) \quad := \quad \begin{cases} 0 & \text{If } a = a_\top \\ \Omega(a) & \text{Otherwise} \end{cases} .$$

For each disjunct $\varphi \in FBF^+(A)$ of $\Delta(a,c)$, the sentence $\varphi^{a_\top}$ is the $a_\top$-translation of φ according to definition 2.38. It follows that the transition function Δ^F has type $A^F \times C \to SLatt(SBF^+(A^F))$. The automaton $\mathbb{A}^F$ is called the *completion* of $\mathbb{A}$. ◁

Proposition 2.40. *Let $\mathbb{A}$ be a non-deterministic MSO-automaton and $\mathbb{A}^F$ its completion. The following holds.*

$$\mathbb{A} \quad \equiv \quad \mathbb{A}^F$$

Proof In order to show the two directions of the equivalence, we fix a tree $\mathbb{T}$.

($\Rightarrow$) Let f be a winning strategy for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$. By proposition 2.13 we can assume f to be *minimal*. We want to define a winning strategy f^F for $\exists$ in the game $\mathcal{G}^F = \mathcal{A}(\mathbb{A}^F, \mathbb{T})@(a_I, s_I)$. The idea is that f^F will be constructed as a sort of 'saturation' of f: from a basic position (a, s), the marking m^F suggested by f^F will extend the marking suggested by f by assigning the state $a_\top$ to every node of $\sigma_R(s)$ that was left unmarked according to f. By minimality of f, this means that there is a disjunct φ of $\Delta(a, \sigma_R(s))$, such that every unmarked node corresponds to a 'hole', i.e. an A-type $\tau_S^+(x)$ with $S = \varnothing$, in φ. By this observation,it can be checked that m^F makes $\varphi^{a_\top}$ true in $\sigma_R(s)$ and then it is a legitimate move for $\exists$ in $\mathcal{G}^F$. For the next basic position, we can always assume that $\forall$ does not choose a position of the form $(a_\top, t)$, because this is winning for $\exists$ by definition of $\mathbb{A}^F$. It follows that we only need to consider f^F-conform matches having the same positions of some f-conform match.

Now we proceed with the formal part of the proof. The strategy f^F will be defined for each stage of the construction of a match π^F in $\mathcal{G}^F$. While playing π^F, player $\exists$ maintains an f-conform shadow match π. Inductively, we will make sure that either infinitely many basic positions associated with state $a_\top$ occur in π^F or $\exists$ can keep the same basic positions for each round in π^F and π.

At the initial stage the match π^F only consists of the position (a_I, s_I). We initialize a shadow match π from the same position (a_I, s_I). By assumption the position (a_I, s_I) is winning for $\exists$ in $\mathcal{G}$, meaning that π is (the initial part of) an f-conform match. By inductive hypothesis suppose that we have constructed (the initial part of) the match π^F, with rounds $z_0, \dots, z_i$, and (the initial part of) an f-conform shadow match π, with rounds $z_0, \dots, z_i$, such that for each round $j \leq i$ the two matches have the same positions. Let (a,s) the basic position occurring in π^F at round z_i. By inductive hypothesis (a,s) also occurs in π at the same round. Let $m : A \to \wp(\sigma_R(s))$ be the suggestion of f from position (a,s). We denote with $\mathcal{U}_m \subseteq \sigma_R(s)$ be the set defined as

$$\mathcal{U}_m \;:=\; \{s \in \sigma_R(s) \mid s \notin m(a) \text{ for all } a \in A\}.$$

We define a marking $m^F : A^F \to \wp(\sigma_R(s))$ by putting

$$m^F(a) \quad := \quad \begin{cases} \mathcal{U}_m & \text{If } a = a_\top \\ m(a) & \text{Otherwise} \end{cases} .$$

In the sequel we show two claims on f^F.

CLAIM 2. The marking m^F suggested by f^F is a legitimate choice for $\exists$.

PROOF OF CLAIM By assumption m makes $\Delta(a, \sigma_C(s))$ true in $\sigma_R(s)$. By minimality of f, there is some disjunct $\varphi \in FBF^+(A)$ of $\Delta(a, \sigma_C(s))$ such that m is minimal among the markings that make φ true in $\sigma_R(s)$. In order to show that m^F is a legitimate choice for $\exists$, it suffices to show that m^F makes $\varphi^{a_\top} \in SBF^+(A^F)$ true. This can be easily checked by minimality of m, definition of $a_\top$-translation and m^F. ■

CLAIM 3. Either the next basic position in π^F is winning for $\exists$ in $\mathcal{G}^F$ or we can maintain the same basic positions in π^F and π at round z_{i+1}.

PROOF OF CLAIM Suppose that $\forall$ picks (b,t) as next position in π^F. If $b = a_\top$, then $\forall$ is doomed to lose the match, because by definition of $\Delta^F(a_\top, \sigma_C(t))$ all next rounds will be associated with basic positions with parity $\Omega(a_\top) = 0$. Therefore (b,t) is a winning position for $\exists$ in $\mathcal{G}^F$.

Otherwise, $b \neq a_\top$ is a state of A and by definition of m^F we have that $t \in m(b)$. This means that (b,t) is a legitimate choice for $\forall$ in π and we can keep the same basic positions one round further in the two matches. ■

By claim 2 and 3, the strategy f^F is surviving for $\exists$. By proposition 2.36, f^F is full and then the match π^F is infinite. By claim 3 it follows that either a winning position for $\exists$ is visited at some round in π^F, or the same parities occur infinitely often in π^F and the f-conform shadow match π. Since f is a winning strategy for $\exists$ in $\mathcal{G}$, also in the latter case the match π^F is won by $\exists$.

($\Leftarrow$) Let f^F be a winning strategy for $\exists$ in $\mathcal{G}^F = \mathcal{A}(\mathbb{A}^F, \mathbb{T})@(a_I, s_I)$. We want to define a winning strategy f for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$. Suppose that $(a,s) \in A \times T$ is a basic position on which f^F is defined, providing a marking $m^F : A^F \to \wp(\sigma_R(s))$. The idea is to let the suggestion of f be the restriction of m^F to a marking $m : A \to \wp(\sigma_R(s))$. Along the same line of reasoning used for direction ($\Rightarrow$), it is easy to check that, while playing an f-conform match π of $\mathcal{G}$, we can maintain an f^F-conform shadow match $\pi^\wp$ of $\mathcal{G}^F$, such that π and $\pi^\wp$ have the same basic positions and $\exists$ never gets stuck in π. □

We are now ready to supply the proof of our normal form theorem for non-deterministic *MSO*-automata.

Proof of proposition 2.37 Let $\mathbb{A}$ be a non-deterministic *MSO*-automaton. By proposition 2.40 the automaton $\mathbb{A}$ is equivalent to its completion $\mathbb{A}^F$. By definition, $\mathbb{A}^F$ is a non-deterministic *MSO*-automaton, with transition function of type $A^F \times C \to SLatt(SBF^+(A^F))$. □

Historical notes

Automata characterizing *MSO* on trees have been first introduced by Rabin [27]. Rabin's results is restricted to binary trees, but it can be easily extended to trees of k-bounded branching degree, for k a positive integer [23].

Automata characterizing *MSO* on trees of arbitrary branching degree have been introduced by Walukiewicz [33]. He works with automata that are in fact equivalent to *MSO*-automata, providing results that are analogous to the Simulation Theorem and the closure properties of section 2.4. Our presentation of *MSO*-automata - in particular, the focus on the Simulation Theorem - follows the perspective of Venema [31]. We refer to [31], theorem 6.17, for a result on μ-automata which is analogous to the Simulation Theorem. The idea of the refined powerset construction also comes from the analysis of μ-automata given in [31], but the same motivation is already behind the *Safra construction* for stream automata [29].

Chapter 3

Automata Characterization of *WFMSO*

In this chapter we work with a restricted class of *MSO*-automata, which we call *weak MSO*-automata. Intuitively, an *MSO*-automaton is weak if we can impose a quasi-order on its states that is 'respected' by the transition function and the parity map.

Definition 3.1. A *weak MSO-automaton* on alphabet C is a tuple $\mathbb{A} = \langle A, \leq, a_I, \Delta, \Omega \rangle$ where:

- $\langle A, a_I, \Delta, \Omega \rangle$ is an *MSO*-automaton;
- the relation $\leq\, \subseteq A \times A$ is a quasi-order on A;
- $\Delta(a,c)$ is in $FO^+(\{b \in A \mid a \leq b\})$ for every $a \in A$ and $c \in C$;
- for every $a, b \in A$ if $a \leq b$ and $b \leq a$ then $\Omega(a) = \Omega(b)$.

Let $\mathbb{T}$ be a tree. The *acceptance game* of $\mathbb{A}$ on $\mathbb{T}$ - notation $\mathcal{A}(\mathbb{A}, \mathbb{T})$ - and its winning conditions are defined as for *MSO*-automata, according to definition 2.1.

Remark 3.2. By definition, for each weak *MSO*-automaton $\mathbb{A} = \langle A, \leq, a_I, \Delta, \Omega \rangle$, the *MSO*-automaton $\mathbb{A}' = \langle A, a_I, \Delta, \Omega \rangle$ is equivalent to $\mathbb{A}$. ◂

Remark 3.3 ([25])**.** The weakness constraint simplifies the structure of the acceptance game for weak *MSO*-automata. Let $\mathbb{A} = \langle A, \leq, a_I, \Delta, \Omega \rangle$ be a weak *MSO*-automaton and $\mathbb{T}$ a tree. Since $\leq$ is a quasi-order, we can partition A into equivalence classes by quotienting the symmetric closure of $\leq$, which we denote with $\equiv_{\leq}$. Let π be an infinite match of $\mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$, associated with basic positions

$$(a_I, s_I), (a_1, s_1), \ldots, (a_n, s_n), \ldots.$$

By definition of Δ, for each $i < \omega$ we have that $a_i \leq a_{i+1}$. Since A is finite, there is some $k < \omega$ in which π stabilizes, in the sense that for all $j \geq k$ we have that $a_j \leq a_{j+1}$ and $a_{j+1} \leq a_j$. In other words, the match π is eventually trapped into an equivalence class E of $\equiv_{\leq}$. According to definition 3.1, all states in E have the same parity. This means that there is a *unique* parity $n \in Ran(\Omega)$ occurring infinitely often in π.

By this observation it follows that we can restrict Ω to a function $\Omega' : A \to \{0,1\}$ without loss of generality. The parity map Ω' is defined by putting $\Omega'(a) = 0$ if $\Omega(a)$ is even and $\Omega'(a) = 1$ otherwise, for each $a \in A$. It is easy to check that $\mathbb{A} = \langle A, \leq, a_I, \Delta, \Omega \rangle$ and $\mathbb{A}' = \langle A, \leq, a_I, \Delta, \Omega' \rangle$ are equivalent. ◂

The main goal of this chapter is to characterize the expressive power of *WFMSO* in terms of weak *MSO*-automata, in analogy with the case of *MSO* and *MSO*-automata. We want to show that for every formula $\varphi \in WFMSO$ we can effectively construct a weak *MSO*-automaton which is equivalent to φ. The argument proceeds by induction on φ. We focus on the inductive case of *WFMSO* existential quantification, which is the non-trivial part of the proof. For this purpose, we define a closure operation on tree languages corresponding to the semantics of *WFMSO* existential quantification.

Definition 3.4. Let $\mathbb{T}$ be a tree and p a propositional letter (not necessarily in P). We refer to the notion of *well-closed p-variant of* $\mathbb{T}$ as defined in section 1.2. Recall that, since $\mathbb{T}$ is a C-labeled tree, any well-closed p-variant $\mathbb{T}^p$ of $\mathbb{T}$ is a $\wp(P \cup \{p\})$-labeled tree.

Let $\mathcal{L}$ be a tree language. The *well-closed projection* of $\mathcal{L}$ over p is the language $\exists_W p.\mathcal{L}$ defined as

$$\exists_W p.\mathcal{L} = \{\mathbb{T} \mid \text{ there is a well-closed } p\text{-variant } \mathbb{T}^p \text{ of } \mathbb{T} \text{ such that } \mathbb{T}^p \in \mathcal{L}\}.$$

Let $\mathcal{C}$ be a class of tree languages. $\mathcal{C}$ is *closed under well-closed projection over p* if, for any language $\mathcal{L}$ in $\mathcal{C}$, also the language $\exists_W p.\mathcal{L}$ is in $\mathcal{C}$. ◃

3.1 The Two-Sorted Construction

Our goal is to provide a *projection construction* that, given a weak *MSO*-automaton $\mathbb{A}$, provides a weak *MSO*-automaton $\exists_W p.\mathbb{A}$ recognizing $\exists_W p.L(\mathbb{A})$.

The idea is to proceed by analogy with the proof showing that the tree languages recognized by *MSO*-automata are closed under projection. In the case of *MSO*-automata, we proved that the construction is correct by using in an essential way the assumption that the starting automaton is non-deterministic. In the case of weak *MSO*-automata, this passage requires a finer analysis.

Remark 3.5. As shown by theorem 2.14, every *MSO*-automaton can be assumed to be non-deterministic. However, this is not the case once we restrict to weak *MSO*-automata. In order to see that, let $\mathbb{A}$ be an *MSO*-automaton. We recapitulate the steps leading to the non-deterministic version of $\mathbb{A}$. This is achieved by performing a refined version of the powerset construction. A non-deterministic automaton $\mathbb{A}^{\wp}$ equivalent to $\mathbb{A}$ is first constructed with a non-parity acceptance condition. Through some further transformations we obtain an equivalent non-deterministic automaton $\mathbb{A}^{P\wp}$ with a parity acceptance condition.

Now suppose that the starting automaton $\mathbb{A}$ is weak with quasi-order $\leq$. The weakness affects essentially the parity acceptance condition of $\mathbb{A}$, which however is not carried to the definition of $\mathbb{A}^{\wp}$. Even if we define a quasi-order $\leq^{P\wp}$ on the macro-states of $\mathbb{A}^{P\wp}$ on the base of $\leq$, there is no way to guarantee that the parity acceptance condition of $\mathbb{A}^{P\wp}$ respects $\leq^{P\wp}$. In other words, even if $\mathbb{A}$ is weak, the equivalent non-deterministic *MSO* automaton $\mathbb{A}^{P\wp}$ is generally not weak[1]. ◂

Because of remark 3.5 we cannot use the full power of non-determinism in the projection construction for weak *MSO*-automata. However, in the sequel we show how a restricted version of non-determinism fits the best for our purposes.

Let $\mathbb{Q}$ be a weak *MSO*-automaton, $\mathbb{T}$ a tree and f a winning strategy for $\exists$ in $\mathcal{G}_Q = \mathcal{A}(\mathbb{Q},\mathbb{T})@(a_I,s_I)$. By proposition 2.10 non-determinism corresponds to the strategy f being *functional*. The main idea is that we require f to be functional only for a finite initial segment (i.e. a partial match) π_F of each match f-conform π of $\mathcal{G}_Q$. This amounts to say that $\mathbb{Q}$ behaves as a non-deterministic automaton as far as the match is played along π_F. We call this behavior the *non-deterministic mode* of $\mathbb{Q}$. Through the remaining part of the match there is no requirement on f being functional or $\mathbb{Q}$ behaving as a non-deterministic automaton. We say that in this part $\mathbb{Q}$ has entered the *alternating mode*.

As we did in the previous chapter, it is convenient to draw the *tree representation* $\mathbb{T}_f$ of f, which comes with a projection function $\pi_2^f : T_f \to T$ according to remark 2.4. Each f-conform match of $\mathcal{G}_Q$ is associated with some backwards closed path of $\mathbb{T}_f$. Since f is not assumed to be functional, π_2^f is generally not a 1-1 correspondence between $\mathbb{T}_f$ and $\mathbb{T}$: for some node $s \in T$, there can possibly be more than one state a of $\mathbb{Q}$ such that (a,s) is a node of $\mathbb{T}_f$. However, each f-conform match π has an initial segment π_F where f is functional. The set of all π_Fs forms a *prefix* $\mathbb{W}_f$ of $\mathbb{T}_f$. The restriction of π_2^f to $\mathbb{W}_f$ is a 1-1 correspondence between $\mathbb{W}_f$ and some well-founded subtree $\mathbb{W}$ of $\mathbb{T}$.

The intuition behind the projection construction is to build a p-variant of $\mathbb{T}$ by allowing nodes labeled with p to be only in $\mathbb{W}$. The resulting p-variant will be well-closed according to definition 3.4.

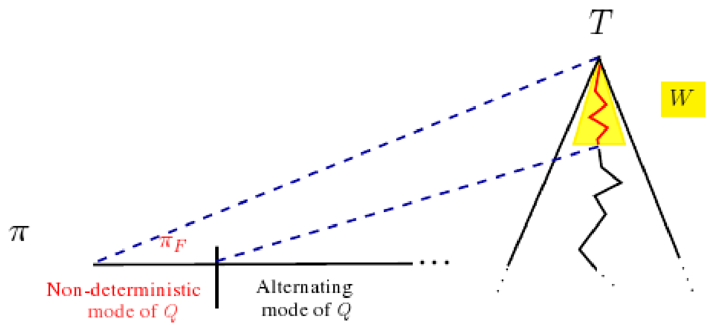

Figure 3.1: the initial segment π_F of an f-conform match π induces a branch in the well-founded subtree $\mathbb{W}$ of $\mathbb{T}$.

[1]The same phenomenon has been observed in a slightly different setting in [14], remark on theorem 4.1.

Definition 3.6. Let $\mathbb{Q} = \langle Q, q_I, \Delta, \Omega \rangle$ be an *MSO*-automaton such that $Q = B_1 \uplus B_2$ for two finite sets B_1 and B_2. We say that $\mathbb{Q}$ is *non-deterministic in* B_1 if the following condition holds.

- Let $\mathbb{T}$ be a tree and f a winning strategy for $\exists$ in $\mathcal{A}(\mathbb{Q}, \mathbb{T})@(q_I, s_I)$. Let $\mathbb{T}_f$ and $\pi_f^2 : T_f \to T$ be respectively the tree representation of f and its projection function as in remark 2.4. There exists a prefix W_f of $\mathbb{T}_f$ such that:

  1. all nodes in W_f are of the form (q_1, s) with $q_1 \in B_1$;
  2. all nodes in $T_f \setminus W_f$ are of the form (q_2, s) with $q_2 \in B_2$;
  3. the restriction of the projection function π_2^f to W_f is 1-1.

◁

In analogy with the case of *MSO*-automata, we want to show that for every weak *MSO*-automaton $\mathbb{A}$ we can canonically construct an equivalent weak *MSO*-automaton that is non-deterministic in some non-empty set of states B. The idea is to take as set B the carrier $A^{\wp}$ of the refined powerset construction on $\mathbb{A}$, as in definition 2.20. The resulting automaton $\mathbb{A}^{2S}$ is called *two-sorted* because it roughly consists of a copy of $\mathbb{A}^{\wp}$ and a copy of $\mathbb{A}$. Given a tree $\mathbb{T}$, we want that any match of $\mathcal{A}(\mathbb{A}^{2S}, \mathbb{T})$ is split in two parts:

1. for finitely many steps $\mathbb{A}^{2S}$ plays a match of the acceptance game of $\mathbb{A}^{\wp}$ on $\mathbb{T}$;
2. at a certain stage $\mathbb{A}^{2S}$ abandons the first match and plays a match of the acceptance game of $\mathbb{A}$ on $\mathbb{T}$.

The first part corresponds to $\mathbb{A}^{2S}$ being in the non-deterministic mode, whereas the second part corresponds to $\mathbb{A}^{2S}$ being in the alternating mode. In fact $\mathbb{A}^{\wp}$ is a non-deterministic automaton, whereas $\mathbb{A}$ is generally not. The formal definition of $\mathbb{A}^{2S}$ will guarantee the correctness of this construction, in the sense that $\mathbb{A}^{2S}$ turns out to be equivalent to the original automaton $\mathbb{A}$. Before going into the details of the construction of $\mathbb{A}^{2S}$, we informally explain its components.

- The set of states of $\mathbb{A}^{2S}$ includes both the original states of $\mathbb{A}$ and the macro-states of $\mathbb{A}^{\wp}$.
- The quasi-order in $\mathbb{A}^{2S}$ of the states of $\mathbb{A}$ is the same as in the original automaton $\mathbb{A}$. In order to guarantee that we can switch from the non-deterministic mode to the alternating mode in any stage of the simulation of $\mathbb{A}^{\wp}$, we put every state of $\mathbb{A}$ 'above' every macro-state, with respect to the quasi-order. The quasi-order between macro-states themselves is defined as the flat order, because the transition of $\mathbb{A}^{\wp}$ does not originally have any order to respect.
- The initial state of $\mathbb{A}^{2S}$ is the initial state of $\mathbb{A}^{\wp}$ because we start every run with $\mathbb{A}^{2S}$ in the non-deterministic mode.
- The transition function of $\mathbb{A}^{2S}$ is just the same of the original automaton $\mathbb{A}$ on the states of $\mathbb{A}$. On the macro-states of $\mathbb{A}^{\wp}$ the transition function of $\mathbb{A}^{2S}$ allows for continuing in the non-deterministic mode or switching to the alternating mode and enter the simulation of $\mathbb{A}$.

$$\begin{array}{rcl} \Delta^{2S}(a,c) & := & \Delta(a,c) \\ \Delta^{2S}(R,c) & := & \Delta^{\wp}(R,c) \vee \bigwedge_{a \in Ran(R)} \Delta(a,c) \end{array}$$

  Observe that the match in the game for $\mathbb{A}$ always starts from a state $a \in Ran(R)$, which is a winning position if R was a winning position in the game for $\mathbb{A}^{\wp}$, as shown in the proof of theorem 2.14. Intuitively, this property guarantees that switching from the non-deterministic to the alternating mode does not alter the winning conditions of a match.
- The parity map for $\mathbb{A}^{2S}$ has the same value as the map of $\mathbb{A}$ on the states of the original automaton $\mathbb{A}$. All the macro-states of $\mathbb{A}^{\wp}$ receive an odd parity. In this way we guarantee that every match that is won by $\exists$ eventually makes $\mathbb{A}^{2S}$ switch from the non-deterministic mode to the alternating mode.

We are now ready to provide the formal definition of $\mathbb{A}^{2S}$.

Definition 3.7. Let $\mathbb{A} = \langle A, \leq, a_I, \Delta, \Omega \rangle$ be a weak *MSO*-automaton and $\mathbb{A}^{\wp} = \langle A^{\wp}, a_I^{\wp}, \Delta^{\wp}, NBT_{\Omega} \rangle$ its refined powerset construction as in definition 2.20. The weak *MSO*-automaton $\mathbb{A}^{2S} = \langle A^{2S}, \leq^{2S}, a_I^{2S}, \Delta^{2S}, \Omega^{2S} \rangle$ is defined as follows.

$$
\begin{array}{rcl}
A^{2S} & := & A \cup A^{\wp} \\
\leq^{2S} & := & \leq \cup (A^{\wp} \times A) \cup (A^{\wp} \times A^{\wp}) \\
a_I^{2S} & := & a_I^{\wp} \\
\Delta^{2S}(a,c) & := & \Delta(a,c) \\
\Delta^{2S}(R,c) & := & \Delta^{\wp}(R,c) \vee \bigwedge_{a \in Ran(R)} \Delta(a,c) \\
\Omega^{2S}(a) & := & \Omega(a) \\
\Omega^{2S}(R) & := & 1
\end{array}
$$

Here a stands for any state in A and R stands for any state in $A^{\wp}$. The automaton $\mathbb{A}^{2S}$ is called the *two-sorted construction over* $\mathbb{A}$. ◁

Proposition 3.8. *Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be a weak MSO-automaton and $\mathbb{A}^{2S}$ the two sorted construction on $\mathbb{A}$. The automaton $\mathbb{A}^{2S}$ is non-deterministic in $A^{\wp}$.*

Proof Let $\mathbb{T}$ be a tree and f a winning strategy for $\exists$ in $\mathcal{G}^{2S} = \mathcal{A}(\mathbb{A}^{2S}, \mathbb{T})@(q_I, s_I)$. By proposition 2.13 we can assume f to be *minimal*. Let $\mathbb{T}_f = \langle T_f, (q_I, s_I), R_f, V_f \rangle$ and $\pi_2^f : T_f \to T$ be respectively the tree representation of f and the projection function for T_f as in remark 2.4. We need to show that there is a prefix W_f of $\mathbb{T}_f$ with the properties described as in definition 3.6.

In order to define W_f, it suffices to provide a frontier B_W of $\mathbb{T}_f$. For this purpose, consider an arbitrary branch E in $\mathbb{T}_f$. By definition of $\mathbb{T}_f$, the branch E corresponds to the sequence of basic positions visited in an f-conform match π_E of $\mathcal{G}^{2S}$. The key observation is that, since f is winning and minimal, we can split E into an initial segment where $\mathbb{A}^{2S}$ is in the non-deterministic mode and the remaining part of E where $\mathbb{A}^{2S}$ is in the alternating mode.

On the base of this observation, it is easy to see that there is a node $(R, s) \in A^{\wp} \times T$ along E with the following properties:

- for each node $(q, t) \in A^{2S} \times T$ in E such that $(q, t) R_f^{\star} (R, s)$, we have that $q \in A^{\wp}$;
- for each node $(q, t) \in A^{2S} \times T$ in E such that $(R, s) R_f^{+} (q, t)$, we have that $q \in A$.

We put $E \cap B_W := \{(R, s)\}$. Since E was an arbitrary branch of $\mathbb{T}_f$, this suffices to define a frontier B_W of $\mathbb{T}_f$. We let W_f be the prefix induced by the frontier B_W, i.e.

$$
W_f \;\; := \;\; \{(Q, t) \in T_f \mid (Q, t) R_f^{\star} (Q', t') \text{ for some } (Q', t') \in B_W\}.
$$

In order to prove the main statement, it remains to show that the restriction $\pi^f_{2\upharpoonright W_f} : W_f \to T$ of π_2^f to W_f is 1-1. Let $W \subseteq T$ defined by putting $W := \pi_2^f[W_f]$. Let $f_{\upharpoonright W}$ denote the restriction of f to a strategy for $\exists$ in partial matches of $\mathcal{G}^{2S}$ which are played along nodes in W, that is, $Dom(f_{\upharpoonright W}) = Dom(f) \cap (A^{2S} \times W)$. By definition of $\mathbb{T}_f$ and π_2^f, in order to show that $\pi^f_{2\upharpoonright W_f}$ is 1-1, it suffices to show that $f_{\upharpoonright W}$ is a functional strategy.

For this purpose, let π be an $f_{\upharpoonright W}$-conform partial match of $\mathcal{G}^{2S}$ which is played along nodes in W, with basic positions $(a_I^{\wp}, s_I), (R_1, s_1), \ldots, (R_k, s_k)$. The key observation is that π can be also seen as an $f_{\upharpoonright W}$-conform partial match of $\mathcal{G}^{\wp} = \mathcal{A}(\mathbb{A}^{\wp}, \mathbb{T})@(a_I^{\wp}, s_I)$. In order to see that, consider any round z_i in π with $i < k$, associated with basic position (R_i, s_i). It suffices to show that the marking m_i suggested by $f_{\upharpoonright W}$ makes $\Delta^{\wp}(R_i, \sigma_C(s_i))$ true in $\sigma_R(s_i)$. Since f is surviving, then $f_{\upharpoonright W}$ is surviving in W and m_i makes $\Delta^{2S}(R_i, \sigma_C(s_i))$ true in $\sigma_R(s_i)$. By definition of Δ^{2S}, this means that m_i makes either $\Delta^{\wp}(R_i, \sigma_C(s_i))$ or $\bigwedge_{a \in Ran(R)} \Delta(a, c)$ true in $\sigma_R(s_i)$. By assumption we know that m_i assigns $R_{i+1} \in A^{\wp}$ to the node s_{i+1}, because (R_{i+1}, s_{i+1}) is the next basic position in π. Since $f_{\upharpoonright W}$ is minimal, this means that m_i makes $\Delta^{\wp}(R_i, \sigma_C(s_i))$ true in $\sigma_R(s_i)$.

Therefore $f_{\upharpoonright W}$ is a strategy for $\exists$ in $\mathcal{G}^{\wp}$ which is surviving in W. By proposition 2.10, it follows that $f_{\upharpoonright W}$ is a functional strategy, whence the function $\pi^f_{2\upharpoonright W_f}$ is 1-1. □

In the sequel we show that the two-sorted construction produces an automaton $\mathbb{A}^{2S}$ which is equivalent to the starting automaton $\mathbb{A}$.

Proposition 3.9. *Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be a weak MSO-automaton and $\mathbb{A}^{2S}$ the two sorted construction on $\mathbb{A}$. For each tree $\mathbb{T}$, the automaton $\mathbb{A}^{2S}$ accepts $\mathbb{T}$ if and only if $\mathbb{A}$ accepts $\mathbb{T}$.*

Proof **($\Rightarrow$)** Let $\mathbb{T}$ be a tree, f^{2S} a winning strategy of $\exists$ in $\mathcal{G}^{2S} = \mathcal{A}(\mathbb{A}^{2S}, \mathbb{T})@(a_I^{\wp}, s_I)$. We want to define a strategy f that is winning for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$. The key observation underlying the construction of f is that, for each f^{2S}-conform match π^{2S} of $\mathcal{G}^{2S}$, there is a round z in which $\mathbb{A}^{2S}$ enters the alternating mode: from that moment on, $\mathbb{A}^{2S}$ 'behaves' as $\mathbb{A}$, and π^{2S} becomes in fact a match of $\mathcal{G}$. Then we can simply let f be defined as f^{2S} for all rounds after z. For the rounds preceding round z, $\mathbb{A}^{2S}$ is in non-deterministic mode and 'behaves' as $\mathbb{A}^{\wp}$. This means that in this phase π^{2S} looks as match of the game $\mathcal{A}(\mathbb{A}^{\wp}, \mathbb{T})@(a_I^{\wp}, s_I)$. However, the equivalence between $\mathbb{A}^{\wp}$ and $\mathbb{A}$ provides us with a canonical way to construct the strategy f for those rounds before z, in terms of a strategy for $\exists$ in $\mathcal{A}(\mathbb{A}^{\wp}, \mathbb{T})@(a_I^{\wp}, s_I)$.

Now we proceed with the formal part of the proof. The strategy f will be defined for each stage of the construction of a match π in $\mathcal{G}$. While playing π, player $\exists$ maintains an f^{2S}-conform shadow match π^{2S}. Inductively, we will make sure that $\exists$ can keep the following condition for each round z in π and π^{2S}.

Either case 1 or case 2 holds for the current round z.

1. **Case 1** The current basic position in π is of the form $(a, s) \in A \times T$ and the current basic position in π^{2S} is of the form $(R, s) \in A^{\wp} \times T$, with $a \in Ran(R)$.
2. **Case 2** Both π and π^{2S} are at the same basic position of the form $(a, s) \in A \times T$.

($\ddagger$)

Let us first show why $\exists$ is guaranteed to win π if she never gets stuck and condition ($\ddagger$) is maintained for each round that is played in π and π^{2S}. The case in which finitely many rounds are played corresponds to player $\forall$ getting stuck in π. In the remaining case, suppose that infinitely many rounds are played in π and π^{2S}. We argue that there is some round z in which a basic position of the form $(b, t) \in A \times T$ occurs in the f^{2S}-conform shadow match π^{2S}. If this was not the case, by definition of Δ^{2S}, infinitely many positions of the form $(R, t) \in A^{\wp} \times T$ would occur in π^{2S}. Then the unique parity $n < \omega$ occurring infinitely often along π^{2S} would be associated with a position of the form $(R, t) \in A^{\wp} \times T$. By definition of Ω^{2S}, the parity n is odd, contradicting the fact that f^{2S} is a winning strategy for $\exists$.

Therefore there is a round $z < \omega$ in which a basic position of the form $(b, t) \in A \times T$ occurs in π^{2S}. By definition of Δ^{2S}, for all successive rounds $z_{i+1}, z_{i+2}, \ldots$ in which we maintain the match π^{2S}, only positions of the form $(d, r) \in A \times T$ can occur in π^{2S}. Since the strategy f^{2S} is winning, then the unique parity $m < \omega$ occurring infinitely often along the play associated with rounds $z_i, z_{i+1}, z_{i+2}, \ldots$ is even. By condition ($\ddagger$), rounds $z_i, z_{i+1}, z_{i+2}, \ldots$ in π have the same basic positions of rounds $z_i, z_{i+1}, z_{i+2}, \ldots$ in π^{2S}. It follows that m is also the unique parity occurring infinitely often along π, so $\exists$ wins π.

Now the goal is to define a strategy f for $\exists$ in π, so that the move suggested by f is always legitimate and $\exists$ can maintain condition ($\ddagger$) for each inductive step of the construction of π and π^{2S}. At the initial round the match π consists only of the position (a_I, s_I). We start the construction of the shadow match π^{2S} from position $(a_I^{\wp}, s_I)$. By assumption $(a_I^{\wp}, s_I) \in Win_{\exists}(\mathcal{G}^{2S})$, so π^{2S} is in fact (the initial part of) an f^{2S}-conform shadow match. Since a_I is in $Ran(a_I^{\wp})$, condition ($\ddagger$) holds for the first stage of the construction.

Inductively, suppose that we have constructed (the initial part of) the match π, with rounds $z_0, \ldots, z_i$, and (the initial part of) an f^{2S}-conform shadow match π^{2S}, with rounds $z_0, \ldots, z_i$, such that for each $j \leq i$ condition ($\ddagger$) is respected. Let (a, s) the basic position occurring in π at round z_i. In order to define the value of f on (a, s), we distinguish two cases.

1. First, suppose that position (a, s) occurs also in π^{2S} at round z_i. Let $m^{2S}_{a,s}$ be the marking suggested by f^{2S} to $\exists$ from position (a, s) in π^{2S}. We let $m^{2S}_{a,s}$ be also the suggestion for $\exists$ from position (a, s) in π.

2. Otherwise, suppose that at round z_i a position (R, s) occurs in π^{2S} with $a \in Ran(R)$. Let $m^{2S}_{R,s}$ be the marking suggested by f^{2S} to $\exists$ from position (R, s) in π^{2S}. We distinguish two further cases.

   a) If $(\sigma_R(s), m^{2S}_{R,s}) \models \bigwedge_{b \in Ran(R)} \Delta(b, \sigma_C(s))$, then $\exists$ picks the marking $m^{2S}_{R,s}$.

   b) If $(\sigma_R(s), m^{2S}_{R,s}) \models \Delta^{\wp}(R, \sigma_C(s))$, since a is in $Ran(R)$ we can apply proposition 2.22 to get a marking $m_a : A \to \wp(\sigma_R(s))$. We let m_a be the choice of $\exists$ in π.

Observe that cases 2.a and 2.b are exhaustive because f^{2S} is winning, meaning that

$$(\sigma_R(s), m^{2S}_{R,s}) \vDash \Delta^{\wp}(R, \sigma_C(s)) \vee \bigwedge_{b \in Ran(R)} \Delta(b, \sigma_C(s)).$$

Also cases 1 and 2 are exhaustive because condition (‡) holds by inductive hypothesis. In order to complete the proof, we need to show that the moves for $\exists$ prescribed by f are legitimate and allow to maintain condition (‡) for one more round.

CLAIM 4. The move suggested by f from position (a,s) is legitimate.

PROOF OF CLAIM This clearly holds for case 1 of the definition of f. In order to prove the same for case 2.a, the key observation is that by assumption $a \in Ran(R)$. It follows that $\Delta(a, \sigma_C(s))$ is one of the conjuncts of $\bigwedge_{b \in Ran(R)} \Delta(b, \sigma_C(s))$, therefore

$$(\sigma_R(s), m^{2S}_{R,s}) \vDash \Delta(a, \sigma_C(s)).$$

Concerning case 2.b, proposition 2.22 guarantees that $(\sigma_R(s), m_a) \vDash \Delta(a,s)$. ■

CLAIM 5. If $\forall$ does not get stuck at round z_i, then player $\exists$ can maintain condition (‡) at round z_{i+1}.

PROOF OF CLAIM If f was defined according to case 1, then the set $\sigma_R(s)$ is marked according to marking $m^{2S}_{a,s}$ both in π and π^{2S}. Therefore any next basic position (b,t) chosen by $\forall$ in π^{2S} is also available in the shadow match π^{2S}. By letting $\forall$ choose (b,t) in π^{2S}, we have the same basic positions in the two matches at round z_{i+1}, so that condition (‡) is maintained.

Otherwise, suppose that f was defined according to case 2.a. The same argument provided for case 1 applies. In the remaining case, f suggested a marking m to $\exists$ according to case 2.b. Let (b,t) be the next basic position chosen by $\forall$ in π. Proposition 2.22 guarantees that there is $Q \in A^{\wp}$ such that $b \in Ran(Q)$ and $t \in m^{\wp}(Q)$. Therefore the basic position (Q,t) is an available choice for $\forall$ in the shadow match π^{2S}. By letting $\forall$ choose (Q,t) in π^{2S}, we are able to maintain condition (‡). ■

The proof of the two claims completes the proof of the main statement.

($\Leftarrow$) Given a winning strategy f for $\exists$ in $\mathcal{G}$, we want to construct a winning strategy f^{2S} for $\exists$ in $\mathcal{G}^{2S}$. The definition of f is provided for each stage of the construction of a match π^{2S} of $\mathcal{G}^{2S}$, while maintaining an f-conform shadow match π of $\mathcal{G}$.

The idea of the definition is to let A^{2S} enter immediately the alternating mode. This means that we make $\exists$ play the very same strategy f in π^{2S}. From position $(a^{\wp}_I, s_I)$ in π^{2S}, the marking m suggested by f makes $\Delta(a_I, \sigma_C(s_I))$ true, which is just the formula $\bigwedge_{a \in Ran(a^{\wp}_I)} \Delta(a, \sigma_C(s_I))$. Therefore $(\sigma_R(s_I), m) \vDash \Delta^{2S}(a^{\wp}_I, \sigma_C(s_I))$. It is straightforward to check that from this round on the two matches have the same basic positions.

□

3.2 From *WFMSO* to Weak *MSO*-Automata

We are now ready to show the main result of this section: the class of tree languages recognized by weak *MSO*-automata is closed under well-closed projection. The argument is analogous to the one showing that *MSO*-automata are closed under projection, but we use the two-sorted construction instead of the refined powerset construction. The p-variant induced by the projection automaton will be guaranteed to be well-closed because all nodes labeled with p are visited when the automaton is in non-deterministic mode.

Definition 3.10. Let $\mathbb{A} = \langle A, \leq, a_I, \Delta, \Omega \rangle$ be a weak *MSO*-automaton on alphabet $\wp(P \cup \{p\})$. Let $\mathbb{A}^{2S}$ denote its two-sorted construction. We define the automaton $\exists_W p.\mathbb{A} = \langle A^{2S}, \leq^{2S}, a^{2S}_I, \tilde{\Delta}, \Omega^{2S} \rangle$ on alphabet C by putting

$$\begin{array}{rcl} \tilde{\Delta}(a,c) & := & \Delta^{2S}(a,c) \\ \tilde{\Delta}(R,c) & := & \Delta^{2S}(R,c) \vee \Delta^{2S}(R, c \cup \{p\}). \end{array}$$

The automaton $\exists_W p.\mathbb{A}$ is called the *two-sorted projection construction of* $\mathbb{A}$ *over* p.

Remark 3.11. Given $\mathbb{A}$ and $\exists_W p.\mathbb{A}$ as above, we observe that $\exists_W p.\mathbb{A}$ is a weak *MSO*-automaton which is non-deterministic in $A^{\wp}$ according to definition 3.6. In order to see that, the key observation is that $\exists_W p.\mathbb{A}$ is defined almost as $\mathbb{A}^{2S}$. Then the same argument provided in proposition 3.8 to show that $\mathbb{A}^{2S}$ is non-deterministic in $A^{\wp}$ also holds for the case of $\exists_W p.\mathbb{A}$. ◂

Proposition 3.12. *For each weak MSO-automaton $\mathbb{A}$ on alphabet $\wp(P \cup \{p\})$, let $\exists_W p.\mathbb{A}$ be the two-sorted projection construction of $\mathbb{A}$ over p, on alphabet C. The following holds.*

$$L(\exists_W p.\mathbb{A}) = \exists_W p.L(\mathbb{A}) \tag{3.1}$$

Proof Let $\mathbb{A} = \langle A, \leq, a_I, \Delta, \Omega\rangle$ be a weak *MSO*-automaton on alphabet $\wp(P \cup \{p\})$ and $\exists_W p.\mathbb{A} = \langle A^{2S}, \leq^{2S}, a_I^{2S}, \tilde{\Delta}, \Omega^{2S}\rangle$ the two-sorted projection construction of $\mathbb{A}$ over p. We want to prove that, for any tree $\mathbb{T}$:

$\exists_W p.\mathbb{A}$ accepts $\mathbb{T}$ iff there is a well-closed p-variant $\mathbb{T}^p$ of $\mathbb{T}$ such that $\mathbb{A}$ accepts $\mathbb{T}^p$.

($\Rightarrow$) Let $\mathbb{T}$ be a tree and $\tilde{f}$ a winning strategy of $\exists$ in $\tilde{\mathcal{G}} = \mathcal{A}(\exists_W p.\mathbb{A}, \mathbb{T})@(a_I^{\wp}, s_I)$. Our goal is to provide a well-closed p-variant $\mathbb{T}^p$ of the tree $\mathbb{T}$ and a winning strategy for $\exists$ in $\mathcal{G}^{2S} = \mathcal{A}(\mathbb{A}^{2S}, \mathbb{T}^p)@(a_I^{\wp}, s_I)$. Then the proof is completed by applying proposition 3.9.

The intuitions underlying the construction of a p-variant $\mathbb{T}^p$ are similar to the ones provided for the case of *MSO*-automata - proposition 2.29. The essential difference is that we cannot assume the automaton $\exists_W p.\mathbb{A}$ to be non-deterministic. However, we can assume $\exists_W p.\mathbb{A}$ to be non-deterministic in $A^{\wp}$, and this is exactly what is needed.

By definition of $\exists_W p.\mathbb{A}$ in terms of the two-sorted construction $\mathbb{A}^{2S}$, there is a well-founded subtree $\mathbb{W}$ of $\mathbb{T}$ where $\exists_W p.\mathbb{A}$ behaves as a non-deterministic automaton. Then we can use the information suggested by $\tilde{f}$ to label with p the nodes in $\mathbb{W}$, just as in the proof of proposition 2.29. Since $\mathbb{W}$ is well-founded, we are guaranteed that the resulting p-variant $\mathbb{T}^p$ is well-closed. Since $\mathbb{T}^p$ has been defined on the base of the information suggested by $\tilde{f}$, it is not difficult to check that $\tilde{f}$ serves also as a winning strategy for $\exists$ in $\mathcal{G}^{2S}$.

In the sequel we give a proof of direction ($\Rightarrow$) following the intuitions described above. Let $\mathbb{T}_{\tilde{f}}$ be the tree representation of $\tilde{f}$ and $\pi_2^{\tilde{f}} : T_{\tilde{f}} \to T$ the associated projection function. By remark 3.11, the automaton $\exists_W p.\mathbb{A}$ is non-deterministic in $A^{\wp}$, meaning that there is a prefix $W_{\tilde{f}}$ of $\mathbb{T}_{\tilde{f}}$ with the properties described as in definition 3.6. Just as we observed in the proof of proposition 2.29, since $\pi_2^{\tilde{f}}$ is 1-1 on $W_{\tilde{f}}$ then for each node $s \in \pi_2^{\tilde{f}}[W_{\tilde{f}}]$ there is a unique $q \in A^{2S}$ such that $\pi_2^{\tilde{f}}(q,s) = s$. By definition of $W_{\tilde{f}}$ we know that $q = R$ for some $R \in A^{\wp}$. Since such R is unique for s, we use the notation R_s to indicate R. We use the map $\pi_2^{\tilde{f}}$ to define a subset X_p of $\mathbb{T}$ as follows.

$$X_p \quad := \quad \{s \in \pi_2^{\tilde{f}}[W_{\tilde{f}}] \mid (\sigma_R(s), \tilde{f}(R_s, s)) \models \Delta^{2S}(R_s, \sigma_C(s) \cup \{p\})\} \tag{3.2}$$

Observe that by definition X_p is a subset of $\pi_2^{\tilde{f}}[W_{\tilde{f}}]$. Since $W_{\tilde{f}}$ is a well-founded subtree of $\mathbb{T}_f$, then also $\pi_2^{\tilde{f}}[W_{\tilde{f}}]$ is a well-founded subtree of $\mathbb{T}$. It follows that X_p is in $WC(\mathbb{T})$.

We denote with $\mathbb{T}^p$ the p-variant of $\mathbb{T}$ induced by X_p. By the fact that X_p is in $WC(\mathbb{T})$, the tree $\mathbb{T}^p$ is in fact a *well-closed* p-variant of $\mathbb{T}$. In order to complete the proof of direction ($\Rightarrow$), we need to provide a winning strategy for $\exists$ in $\mathcal{G}^{2S}$. In fact we claim that such strategy can be taken to be $\tilde{f}$ itself.

CLAIM 6. The strategy $\tilde{f}$ is winning for $\exists$ in $\mathcal{G}^{2S}$.

PROOF OF CLAIM In order to prove the claim, we construct an $\tilde{f}$-conform match π^{2S} of $\mathcal{G}^{2S}$, while maintaining an $\tilde{f}$-conform shadow match $\tilde{\pi}$ of $\tilde{\mathcal{G}}$. Inductively, we will make sure that $\exists$ never gets stuck in π^{2S} and she can keep the same basic positions for each round z_i that is played in $\tilde{\pi}$ and π^{2S}. We refer to this as condition (‡). Since $\exists_W p.\mathbb{A}$ and $\mathbb{A}^{2S}$ have the same set of states and the same parity map, it is immediate to check that maintaining condition (‡) along the whole play is enough to show that $\exists$ wins the match π^{2S}.

At the initial round the match π^{2S} consists only of the position $(a_I^{\wp}, s_I)$. We start the construction of an $\tilde{f}$-conform shadow match $\tilde{\pi}$ from the same position $(a_I^{\wp}, s_I)$. Inductively, let $(q,s) \in A^{2S} \times T$ be the basic position occurring both in $\tilde{\pi}$ and π^{2S} at round z_i. By inductive hypothesis $\tilde{\pi}$ is $\tilde{f}$-conform, meaning that $\tilde{f}$ suggests a marking $m : A^{2S} \to \wp(\sigma_R(s))$ that makes $\tilde{\Delta}(q, \sigma_C(s))$ true in $\sigma_R(s)$. In order to keep π^{2S} conform to the strategy $\tilde{f}$, we suggest the same marking to $\exists$ in π^{2S}. Then we need to check the following two statements:

1. the marking m is a legitimate choice for $\exists$ in π^{2S};

2. condition (‡) can be maintained in the next round z_{i+1}.

In order to show the first statement, we distinguish two cases.

a) If q is of the form $a \in A$, then the first property of $W_{\tilde{f}}$ provided by definition 3.6 excludes that s is in $W_{\tilde{f}}$. Therefore s is also not in X_p, and the label of s in $\mathbb{T}^p$ does not contain p. More formally, this means that

$$\sigma_C^p(s) = \sigma_C(s), \tag{3.3}$$

where $\sigma_C^p : T \to \wp(P \cup \{p\})$ is the labeling function of $\mathbb{T}^p$ and $\sigma_C : T \to C$ the labeling function of $\mathbb{T}$.

By assumption m makes $\Delta^{2S}(a, \sigma_C(s))$ true in $\sigma_R(s)$. By (3.3) the marking m also makes $\Delta^{2S}(a, \sigma_C^p(s))$ true in $\sigma_R(s)$ and then it is an admissible choice for $\exists$ in π^{2S}.

b) Otherwise, suppose that q is of the form $R \in A^{\wp}$. By the second property of $W_{\tilde{f}}$ provided by definition 3.6 we know that (R,s) is a node in $W_{\tilde{f}}$. Since $\pi_2^{\tilde{f}}$ is 1-1 on $W_{\tilde{f}}$ we also know that R is the *unique* state such that $(R,s) \in W_{\tilde{f}}$. At this point, we distinguish two further cases.

a) If the label of s in $\mathbb{T}^p$ does contain p, by definition s is in X_p. This means that

$$(\sigma_R(s), m) \models \Delta^{2S}(R, \sigma_C(s) \cup \{p\}). \tag{3.4}$$

By definition of σ_C^p the label $\sigma_C(s) \cup \{p\}$ is equal to $\sigma_C^p(s)$. Therefore m makes $\Delta^{2S}(R, \sigma_C^p(s))$ true in $(\sigma_R(s), m)$ and it is an admissible choice for $\exists$ in π^{2S}.

b) In the remaining case, the label of s in $\mathbb{T}^p$ does not contain p. This means that

$$(\sigma_R(s), m) \not\models \Delta^{2S}(R, \sigma_C(s) \cup \{p\}). \tag{3.5}$$

Since m is a winning position for $\exists$ in $\tilde{\pi}$, it follows that m makes the other disjunct of $\tilde{\Delta}(R, \sigma_C(s))$ true, i.e.

$$(\sigma_R(s), m) \models \Delta^{2S}(R, \sigma_C(s)). \tag{3.6}$$

Since by assumption the label of s in $\mathbb{T}^p$ does not contain p, we are in the situation depicted in (3.3). Therefore m makes $\Delta^{2S}(R, \sigma_C^p(s))$ true in $(\sigma_R(s), m)$ and it is an admissible choice for $\exists$ in π^{2S}.

It should be clear that the cases considered above exhaust the possibilities for the position (q,s). Therefore m is an admissible choice for $\exists$ in π^{2S}. It remains to show the second statement, namely that condition (‡) can be maintained in the next round z_{i+1}. This is clear because the same marking is suggested to $\exists$ in π^{2S} and $\tilde{\pi}$. ■

The proof of the claim together with proposition 3.9 completes the proof of direction ($\Rightarrow$).

($\Leftarrow$) Let $\mathbb{T}$ be a tree and $\mathbb{T}^p$ a well-closed p-variant of $\mathbb{T}$ that is accepted by $\mathbb{A}$. Let f be a winning strategy for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A}, \mathbb{T}^p)@(a_I, s_I)$. We want to provide a winning strategy $\tilde{f}$ to $\exists$ in $\tilde{\mathcal{G}} = \mathcal{A}(\exists_W p.\mathbb{A}, \mathbb{T})@(a_I^{\wp}, s_I)$.

As a preliminary remark, observe that the automaton $\exists_W p.\mathbb{A}$ is almost defined as $\mathbb{A}^{2S}$. Thus the argument we are going to use to construct $\tilde{f}$ is somehow similar to the one provided for showing that $\mathbb{A}^{2S}$ and $\mathbb{A}$ are equivalent, in proposition 3.9.

As in direction ($\Leftarrow$) of proposition 3.9, we could be tempted to define $\tilde{f}$ simply as f. Intuitively, this would have the effect of making $\exists_W p.\mathbb{A}$ enter the alternating mode already at the first round of any $\tilde{f}$-conform match of $\tilde{\mathcal{G}}$. Then, if the moves suggested by $\tilde{f}$ are always legitimate, any $\tilde{f}$-conform match of $\tilde{\mathcal{G}}$ would have the same basic positions of an f-conform match of $\mathcal{G}$.

The problem with this argument is indeed that the moves suggested by f are not generally legitimate. In order to see that, suppose that s is a node whose label includes p in $\mathbb{T}^p$. If $(a,s) \in Win_{\exists}(\mathcal{G})$ is a winning position for $\exists$ in $\mathcal{G}$, then f would provide a marking $m_{a,s}$ such that $(\sigma_R(s), m_{a,s}) \models \Delta(a, \sigma_C^p(s))$. However, since by assumption $\sigma_C^p(s) = \sigma_C(s) \cup \{p\}$ is different from $\sigma_C(s)$, then generally $(\sigma_R(s), m_{a,s}) \not\models \Delta(a, \sigma_C(s))$. If the same position (a,s) is visited in a match of $\tilde{\mathcal{G}}$, player $\exists$ should come up with a marking that makes $\Delta(a, \sigma_C(s))$ true. This means that $m_{a,s}$ may not be a legitimate choice for her.

We are going to refine this argument such that it gives a correct proof of the statement. The underlying idea of our construction is that $\tilde{f}$ is going to be defined as f only from positions of the form $(a,s) \in A \times T$, where we are guaranteed that the label of s in $\mathbb{T}^p$ does not include p, so that we can avoid the situation described above. In fact this condition holds for any match of $\tilde{\mathcal{G}}$ after finitely many rounds. The reason is that the p-variant $\mathbb{T}^p$ is well-closed, meaning that all nodes with label including p are confined into a well-founded subtree X_p of $\mathbb{T}$.

Therefore we are going to put $\tilde{f}$ equal to f as soon as a match of $\tilde{\mathcal{G}}$ 'crosses' the frontier of X_p. Intuitively, this moment coincides with $\exists_W p.\mathbb{A}$ entering the alternating mode. For all the precedent rounds, we keep $\exists_W p.\mathbb{A}$ in the non-deterministic mode. All positions visited in this initial part will be of the form $(R,s) \in A^{\wp} \times T$. A legitimate move for $\exists$ from such position can make either $\Delta^{2S}(R, \sigma_C(s))$ or $\Delta^{2S}(R, \sigma_C(s) \cup \{p\})$ true: intuitively, this means that we are able to deal both with the case $p \in \sigma^p(s)$ and $p \notin \sigma^p(s)$. The marking for $\exists$ from position (R,s) will be suggested by f, passing through the equivalence between $\mathbb{A}$ and $\mathbb{A}^{\wp}$ as we already observed in proposition 3.9.

On the base of these intuitions, we proceed with the formal details of the proof. In the sequel we say that a node $s \in T$ is *p-free* if $\mathbb{T}.s$ does not contain any node t such that $p \in \sigma_C^p(t)$. The strategy $\tilde{f}$ will be defined for each stage of the construction of a match $\tilde{\pi}$ in $\tilde{\mathcal{G}}$. While playing $\tilde{\pi}$, player $\exists$ maintains a set $\mathcal{M}$ of f-conform shadow matches. We indicate with $\mathcal{M}_i$ the set $\mathcal{M}$ at round i. Inductively, we will make sure that $\exists$ can keep the following condition for each round z_i in $\tilde{\pi}$ and each match in $\mathcal{M}_i$.

Either case 1 or case 2 holds for the current round z_i.

1. **Case 1** The current basic position in $\tilde{\pi}$ is of the form $(R,s) \in A^{\wp} \times T$. For each $a \in Ran(R)$ there is an f-conform shadow match π_a in $\mathcal{M}_i$ at the same round z_i, such that the current basic position in π_a is $(a,s) \in A \times T$. Either the node s is not p-free or s has a sibling t which is not p-free.

2. **Case 2** The current basic position in $\tilde{\pi}$ is of the form $(a,s) \in A \times T$, with s a p-free node. The set $\mathcal{M}_i$ consists only of one f-conform shadow match π, whose current basic position is (a,s).

(‡)

Let us first show why $\exists$ is guaranteed to win $\tilde{\pi}$ if she never gets stuck in $\tilde{\pi}$ and she can keep condition (‡) for each round that is played in $\tilde{\pi}$ and the shadow matches in $\mathcal{M}$. If $\forall$ gets stuck in $\tilde{\pi}$, then $\exists$ immediately wins the match. Otherwise, $\tilde{\pi}$ is infinite and we argue that in finitely many steps a basic position of the form $(b,t) \in A \times T$ occurs in $\tilde{\pi}$. If this was not the case, consider the branch B of $\mathbb{T}$ along which the match $\tilde{\pi}$ is played. By condition (‡), there would be infinitely many nodes r in $\mathbb{T}^p$ such that $\sigma_R(r)$ intersects B and some node in $\sigma_R(r)$ is labeled with p. This contradicts the assumption that $\mathbb{T}^p$ is a well-closed p-variant.

Therefore the shadow match $\tilde{\pi}$ arrives to a round z_i, associated with basic position (b,t), where t is p-free. By condition (‡) at the same round z_i there is a unique f-conform shadow match π which is at the same basic position (b,t). By definition of $\tilde{\Delta}$, for all successive rounds $z_{i+1}, z_{i+2}, \ldots$ in which we maintain the match $\tilde{\pi}$, only positions of the form $(d,r) \in A \times T$ can occur in $\tilde{\pi}$. By condition (‡) the same positions occur in π in the same rounds. Observe that all nodes visited in rounds $z_{i+1}, z_{i+2}, \ldots$ are p-free since they are all elements of $\mathbb{T}_t$ and t is p-free. Since the strategy f is winning, then the unique parity $m < \omega$ occurring infinitely often along the play associated with rounds $z_i, z_{i+1}, z_{i+2}, \ldots$ is even. It follows that m is also the unique parity occurring infinitely often along the match $\tilde{\pi}$, meaning that it is won by $\exists$.

Now the goal is to define the strategy $\tilde{f}$ for $\exists$ in $\tilde{\pi}$, in such a way that the marking suggested by $\tilde{f}$ is always legitimate and she can maintain condition (‡) for each round that is played in $\tilde{\pi}$ and the shadow matches in $\mathcal{M}$.

First we consider the trivial case in which the p-variant $\mathbb{T}^p$ does not have any node whose label includes p. Then $\mathbb{T}^p$ is just defined as $\mathbb{T}$ and it is straightforward to check that by putting $\tilde{f}$ equal to f player $\exists$ is going to win the match $\tilde{\pi}$. The argument showing that follows the lines of the proof of direction ($\Leftarrow$) in proposition 3.9.

Therefore in the sequel we can assume that $\mathbb{T}^p$ contains at least one node whose label includes p, implying that the node s_I is not p-free. At the initial round, the match $\tilde{\pi}$ consists only of the position $(a_I^{\wp}, s_I)$. We start the construction of a shadow match π_{a_I} from position (a_I, s_I). By assumption $(a_I, s_I) \in Win_{\exists}(\mathcal{G})$, so π_{a_I} is in fact (the initial part of) an f-conform shadow match. We initialize $\mathcal{M}_1 = \{\pi_{a_I}\}$.

Since $a_I \in Ran(a_I^{\wp})$ and s_I is not p-free then case 1 of condition (‡) holds for the first stage of the construction. Inductively, suppose that we have constructed (the initial part of) the match $\tilde{\pi}$, with rounds $z_0, \ldots, z_i$. Also, we are provided with a set $\mathcal{M}_i$, where each element of M_i is (the initial part of) an f-conform shadow match, with rounds $z_0, \ldots, z_i$, such that for each $j \leq i$ condition (‡) is respected by $\mathcal{M}_j$ and $\tilde{\pi}$. Let $(q,s) \in A^{2S} \times T$ be the basic position occurring in $\tilde{\pi}$ at round z_i. In order to define the value of $\tilde{f}$ on (q,s), we distinguish two cases.

1. First, suppose that $q = a$ for some $a \in A$. By case 2 of condition (‡) position (q,s) occurs also in π at round z_i, where π is the only f-conform shadow match we are maintaining in $\mathcal{M}_i$. Let $m_{a,s}$ be the marking suggested by f to $\exists$ from position (a,s) in π. We let $m_{a,s}$ be also the suggestion for $\exists$ from position (a,s) in $\tilde{\pi}$.

2. Otherwise, we have that $q = R$ for some $R \in A^{\wp}$. By case 1 of condition (‡) the set $\mathcal{M}_i$ contains one f-conform shadow match π_a at position (a,s) in the same round z_i, for each $a \in Ran(R)$. Let $m_{a,s}$ be the marking suggested by f to $\exists$ from position (a,s) in π_a. We distinguish two further cases.

a) If every $t \in \sigma_R(s)$ is p-free, then we define $\tilde{m} : A \to \wp(\sigma_R(s))$ by putting

$$\tilde{m} : b \quad \mapsto \quad \bigcup_{a \in Ran(R)} m_a(b). \tag{3.7}$$

We let $\tilde{m}$ be the choice of $\exists$ in $\tilde{\pi}$.

b) Otherwise, some $t \in \sigma_R(s)$ is not p-free. In this case we define $\tilde{m} : A^{\wp} \to \wp(\sigma_R(s))$ as the marking obtained by the set of markings $\{m_{a,s} \mid a \in Ran(R)\}$ according to proposition 2.21. We let $\tilde{m}$ be the choice of $\exists$ in $\tilde{\pi}$.

By definition of A^{2S} the two cases considered above are exhaustive for $q \in A^{2S}$.

In order to complete the proof, we need to show that the move suggested by $\tilde{f}$ is legitimate and allows to maintain condition (‡) for one more round.

CLAIM 7. The move suggested by $\tilde{f}$ from position (q,s) is legitimate.

PROOF OF CLAIM If $\tilde{f}$ is defined according to case 1, then $q = a$ for some $a \in A$ and by assumption s is p-free. This means that $\sigma_C(s) = \sigma_C^p(s)$. Since f is winning we know that

$$(\sigma_R(s), m_{a,s}) \vDash \Delta(a, \sigma_C^p(s)).$$

Then we also have that

$$(\sigma_R(s), m_{a,s}) \vDash \Delta(a, \sigma_C(s)),$$

meaning that $m_{a,s}$ is a legitimate move for $\exists$.

Otherwise, suppose that $\tilde{f}$ is defined according to case 2.a. Then $q = R$ for some $R \in A^{\wp}$. As a preliminary step, we observe that for each $a \in Ran(R)$ the sentence $\Delta(a, \sigma_C(s)) \in SLatt(BF^+(A))$ is a disjunction of sentences in basic form. Therefore it enjoys the *monotonicity property* given in remark 1.24. By definition for each $a \in Ran(R)$ the marking $\tilde{m}$ as in (3.7) extends $m_{a,s}$. Since by assumption

$$(\sigma_R(s), m_{a,s}) \vDash \Delta(a, \sigma_C^p(s)),$$

then we also have that

$$(\sigma_R(s), \tilde{m}) \vDash \Delta(a, \sigma_C^p(s)).$$

Therefore

$$(\sigma_R(s), \tilde{m}) \vDash \bigwedge_{a \in Ran(R)} \Delta(a, \sigma_C^p(s)). \tag{3.8}$$

By assumption we cannot guarantee that the node s is p-free. Thus we distinguish two cases.

a) If $p \in \sigma_C^p(s)$, then by definition of p-variant $\sigma_C^p(s) = \sigma_C(s) \cup \{p\}$. This means that $\Delta(a, \sigma_C^p(s)) = \Delta(a, \sigma_C(s) \cup \{p\})$, for each $a \in Ran(R)$.

b) If $p \notin \sigma_C^p(s)$, then by definition of p-variant $\sigma_C^p(s) = \sigma_C(s)$. This means that $\Delta(a, \sigma_C^p(s)) = \Delta(a, \sigma_C(s))$, for each $a \in Ran(R)$.

If we are in the first case, then

$$(\sigma_R(s), \tilde{m}) \vDash \bigwedge_{a \in Ran(R)} \Delta(a, \sigma_C(s) \cup \{p\}).$$

Otherwise we are in the second case and

$$(\sigma_R(s), \tilde{m}) \vDash \bigwedge_{a \in Ran(R)} \Delta(a, \sigma_C(s)).$$

By definition of $\tilde{\Delta}(R, \sigma_C(s))$ this is sufficient to show that $\tilde{m}$ is a legitimate choice for $\exists$.

The last case we need to consider is when $\tilde{f}$ is defined according to case 2.b. Also in this case $q = R$ for some $R \in A^{\wp}$. By the fact that f is winning we know that for each $a \in Ran(R)$ the marking $m_{a,s}$ makes $\Delta(a, \sigma_C^p(s))$ true in $\sigma_R(s)$. Therefore proposition 2.21 guarantees that

$$(\sigma_R(s), \tilde{m}) \vDash \Delta^{\wp}(R, \sigma_C^p(s)). \tag{3.9}$$

Just as in case 2.a, by definition of $\tilde{\Delta}(R, \sigma_C(s))$ this suffices to show that $\tilde{m}$ is a legitimate choice for $\exists$.

■

CLAIM 8. By playing according to $\tilde{f}$, either player $\exists$ makes $\forall$ get stuck at round z or she can maintain condition ($\ddagger$) in round z_{i+1}.

PROOF OF CLAIM If f was defined according to case 1, then the set $\sigma_R(s)$ is marked according to marking $m_{a,s}$ both in π and $\tilde{\pi}$, therefore any next basic position (b,t) chosen by $\forall$ in $\tilde{\pi}$ is also available in the shadow match π. By letting $\forall$ choose (b,t) in π we have the same basic positions in the two matches at round z_{i+1}. By definition, by the fact that s is p-free it follows that also $t \in \sigma_R(s)$ is p-free. Therefore case 2 of condition ($\ddagger$) is respected.

Otherwise, suppose that f was defined according to case 2.a. By definition of $\tilde{m}$ for any next basic position (b,t) chosen by $\forall$ in $\tilde{\pi}$ the same position is available for $\forall$ in the shadow match π_a, for some $a \in Ran(R)$. We select one such π_a and we dismiss all others shadow matches in $\mathcal{M}_i$, so that $M_{i+1} := \{\pi_a\}$. We let $\forall$ choose (b,t) in π_a. Then $\tilde{\pi}$ and π_a have the same basic position (b,t) occurring at round z_{i+1}. By assumption of case 2.a the node t is p-free. Therefore case 2 of condition ($\ddagger$) is respected.

In the remaining case, f was defined according to case 2.b. Let $(Q,t) \in A^{\wp} \times T$ be the choice of $\forall$ in $\tilde{\pi}$ as next basic position. Proposition 2.21 guarantees that for each $b \in Ran(Q)$ there is some $a \in Ran(R)$ such that $t \in m_a(b)$. Therefore for each $b \in Ran(Q)$ we can select a match $\pi_a \in \mathcal{M}_i$ such that $t \in m_a(b)$. We define $\pi_{a,b}$ as the match π_a extended with the movement given by $\forall$ choosing (b,t) as next basic position. We define $\mathcal{M}_{i+1}$ to be the collection of matches $\pi_{a,b}$ for all $b \in Ran(Q)$.

Since we are in case 2.b of the definition of f, there is some node $r \in \sigma_R(s)$ such that r is not p-free. Therefore either t is not p-free or it has a sibling which is not p-free. Then it is immediate to check that (Q,t) and $\mathcal{M}_{i+1}$ respect case 1 of condition ($\ddagger$). ∎

The proof of the two claims complete the proof of direction ($\Leftarrow$). □

As a corollary we obtain a characterization of the expressive power of *WFMSO* in terms of weak *MSO*-automata.

Theorem 3.13. *For every $\varphi \in WFMSO$, there is an effectively constructible weak MSO-automaton $\mathbb{A}_\varphi$ such that*

$$\text{for any tree } \mathbb{T},\ \mathbb{T} \models \varphi \quad \text{iff} \quad \mathbb{A}_\varphi \text{ accepts } \mathbb{T}.$$

Proof The proof is by induction on φ. Atomic and boolean cases are easily handled with the same argument supplied for *MSO* and *MSO*-automata, reflecting the fact that the semantics of *MSO* and *WFMSO* coincides on these cases. For the case of existential quantifier, we use instead proposition 3.12.

- For the atomic case, we have to consider formulae $R(p,q)$ and $p \sqsubseteq q$. The corresponding automata are $\mathbb{A}_{R(p,q)}$ and $\mathbb{A}_{p\sqsubseteq q}$, defined respectively as in example 2.5 and 2.6. It is easy to see that these two automata are in fact weak *MSO*-automata, where the quasi-order is trivially the cartesian product of states $A \times A$.

- If φ is of the form $\neg\psi$, by inductive hypothesis we have a weak *MSO*-automaton $\mathbb{A}_\psi$ that is equivalent to ψ. Let $\mathbb{A}_{\neg\psi}$ be the *MSO*-automaton obtained by applying proposition 2.31 to $\mathbb{A}_\psi$. By imposing the same quasi-order of $\mathbb{A}_\psi$ also on $\mathbb{A}_{\neg\psi}$, we obtain a weak *MSO*-automaton accepting the same language of $\mathbb{A}_{\neg\psi}$, that is the complement of the language of $\mathbb{A}_\psi$ by proposition 2.31.

- If $\varphi = \psi_1 \vee \psi_2$, let $\mathbb{A}_{\psi_1}$ and $\mathbb{A}_{\psi_2}$ be weak *MSO*-automata which are equivalent respectively to ψ_1 and ψ_2. Let $\mathbb{A}_\varphi$ be the *MSO*-automaton obtained by applying proposition 2.30 to $\mathbb{A}_{\psi_1}$ and $\mathbb{A}_{\psi_2}$. It should be clear that $\mathbb{A}_\varphi$ can be turned into a weak *MSO*-automaton: in order to see that, recall that by proposition 2.30 the carrier of $\mathbb{A}_\varphi$ is $A_1 \cup A_2 \cup \{a_I\}$, where A_i is the carrier of $\mathbb{A}_{\psi_i}$, for $i \in \{1,2\}$. Let $\leq_i$ denote the quasi-order of $\mathbb{A}_{\psi_i}$. We define the quasi-order $\leq$ on $\mathbb{A}_\varphi$ by putting

$$\leq \quad := \quad \leq_1 \cup \leq_2 \cup (\{a_I\} \times (A_1 \cup A_2)).$$

  It is immediate to check that the addition of the quasi-order $\leq$ yields a well-defined weak *MSO*-automaton accepting the same language of $\mathbb{A}_\varphi$, which is, by proposition 2.30, the union of languages of $\mathbb{A}_{\psi_1}$ and $\mathbb{A}_{\psi_2}$.

- If $\varphi = \exists p.\psi$, let $\mathbb{A}_\psi$ be the weak *MSO*-automaton that is equivalent to ψ. Let $\exists_W p.\mathbb{A}_\psi$ be the weak *MSO*-automaton obtained from $\mathbb{A}_\psi$ by proposition 3.12. The following derivation shows that $\exists_W p.\mathbb{A}_\psi$ is equivalent to $\exists p.\psi$.

$$\begin{array}{lcll} \exists_W p.\mathbb{A}_\psi \text{ accepts } \mathbb{T} & \Leftrightarrow & \text{there is } X_p \in WC(\mathbb{T}) \text{ such that } \mathbb{A}_\psi \text{ accepts } \mathbb{T}[p \mapsto X_p] & \text{(proposition 3.12)} \\ & \Leftrightarrow & \text{there is } X_p \in WC(\mathbb{T}) \text{ such that } \mathbb{T}[p \mapsto X_p] \models \psi & \text{(inductive hypothesis)} \\ & \Leftrightarrow & \mathbb{T} \models \exists p.\psi & \text{(semantics of } WFMSO\text{)} \end{array}$$

□

Historical Notes

The version of weak *MSO*-automata working on binary trees are often called *weak alternating tree automata* in the literature. They have been first introduced in [24] as the automata characterizing the expressive power of *WMSO* on binary trees. The idea of an automaton with an 'alternating mode' and a 'non-deterministic mode', which underlies the two-sorted construction, also comes from [24], proof of Lemma 1.

Our definition of weak *MSO*-automata is equivalent to the one provided in [14], definition 4.1. We refer to [25] for an overview on the different characterizations of the weakness constraint for automata working on trees.

Chapter 4

Logical Characterization of Weak *MSO*-Automata

In this chapter we show the converse direction of theorem 3.13, namely that, for every weak *MSO*-automaton $\mathbb{A}$, there is a formula $\varphi_{\mathbb{A}} \in WFMSO$ which is equivalent to $\mathbb{A}$.

Let $\mathbb{T}$ be a tree. The idea is to construct $\varphi_{\mathbb{A}}$ in such a way that it expresses the existence of a winning strategy f for $\exists$ in $\mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$. This encoding goes smoothly if we can assume that f marks each node of $\mathbb{T}$ with *exactly one* state of $\mathbb{A}$.

For this purpose, by theorem 2.14 we can construct a non-deterministic *MSO*-automaton $\mathbb{A}^{P\wp}$ which is equivalent to $\mathbb{A}$. However, as noticed in remark 3.5, the automaton $\mathbb{A}^{P\wp}$ is not generally a weak *MSO*-automaton. This means that different parities can occur infinitely often in the same match of $\mathcal{A}(\mathbb{A}^{P\wp},\mathbb{T})$. As we will see in the sequel, we cannot give an account of the winning conditions of this acceptance game by referring only to well-founded subtrees of $\mathbb{T}$. The quantification of *WFMSO* is too restrictive and we need instead the full generality of *MSO* quantifiers.

In the sequel we overcome this difficulty by showing that, because it is weak, $\mathbb{A}$ can be turned into an equivalent non-deterministic *MSO*-automaton $\mathbb{B}$ where the parity map $\Omega_B : B \to \omega$ can be assumed to range only over $\{0,1\}$. The acceptance game associated with $\mathbb{B}$ turns out to be essentially simpler than the one for arbitrary *MSO*-automata. The states of $\mathbb{B}$ can be divided into *accepting states* - the ones with parity 0 - and *rejecting states* - the ones with parity 1. It should be clear that $\exists$ wins a match if and only if at least one accepting state occurs infinitely often along the play. This resembles a Büchi acceptance condition, and in fact there is an equivalent description of $\mathbb{B}$ as an automaton where the acceptance condition is given by a set F of accepting states, instead of a parity map Ω. It turns out that Büchi acceptance conditions can be described in terms of well-founded trees, so that we can express them by means of *WFMSO*-formulae without requiring the full expressiveness of *MSO* quantifiers. This is the key observation leading to the logical characterization of non-deterministic Büchi automata and weak *MSO*-automata.

4.1 From Weak *MSO*-Automata to Non-Deterministic Büchi Automata

Definition 4.1. A *non-deterministic Büchi automaton* (abbreviated *NDB*) is a non-deterministic *MSO*-automaton $\mathbb{B} = \langle B, b_I, \Delta, \Omega\rangle$ where $Ran(\Omega)$ is a subset of $\{0,1\}$. ◁

Remark 4.2. The name 'Büchi' for *MSO*-automata as in definition 4.1 is motivated by the fact that we can reformulate their acceptance condition as a standard Büchi condition. More precisely, suppose that $\mathbb{B} = \langle B, b_I, \Delta, F\rangle$ is a non-deterministic automaton where F is a subset of B. For a tree $\mathbb{T}$, we define the winning conditions of $\mathcal{A}(\mathbb{B},\mathbb{T})$ as follows:

- ⋆ player $\exists$ wins an infinite match in $\mathcal{A}(\mathbb{B},\mathbb{T})$ if and only if there is a state $b \in F$ occurring infinitely often along the play.

Now let $\Omega_F : B \to \{0,1\}$ be the characteristic function of F, given by putting $\Omega_F(b) := 0$ if $b \in F$ and $\Omega_F(b) := 1$ otherwise, for each $b \in B$. It should be clear that $\mathbb{B}' = \langle B, b_I, \Delta, \Omega_F\rangle$ is a *NDB* automaton which is equivalent to $\mathbb{B}$.

In the sequel, it will be convenient to adopt the presentation of *NDB* automata with an accepting set F, keeping in mind that these are just *MSO*-automata of a particular kind. ◂

Remark 4.3. We observe that the normal form theorem for non-deterministic *MSO*-automata (proposition 2.37) applies to *NDB* automata as well. Indeed, it is immediate to check that the completion of an *NDB* automaton is also an *NBD* automaton. By this fact, we can always assume to work with *NDB* automata where the transition function Δ ranges over disjunctions of sentences in *special* basic form. ◂

Definition 4.4 (Büchi powerset construction)**.** Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ be a weak *MSO*-automaton. By remark 3.3 we can assume that $Ran(\Omega)$ is a subset of $\{0,1\}$. Let $\mathbb{A}^{\wp} = \langle A^{\wp}, a_I^{\wp}, \Delta^{\wp}, NBT_{\Omega}\rangle$ be the refined powerset construction over $\mathbb{A}$ according to definition 2.20. We define an *NDB* automaton $\mathbb{A}^B = \langle A^{\wp}, a_I^{\wp}, \Delta^{\wp}, F_{\Omega}\rangle$ by putting

$$F_{\Omega} \quad := \quad \{R \in A^{\wp} \mid \Omega(a) = 0 \text{ for all } a \in Ran(R)\}.$$

We say that $\mathbb{A}^B$ is the *Büchi powerset construction* over $\mathbb{A}$. ◁

Proposition 4.5. *Let* $\mathbb{A} = \langle A, a_I, \Delta, \Omega\rangle$ *be a weak MSO-automaton and* $\mathbb{A}^B = \langle A^{\wp}, a_I^{\wp}, \Delta^{\wp}, F_{\Omega}\rangle$ *the Büchi powerset construction over* $\mathbb{A}$*. We have that*

$$\mathbb{A} \equiv \mathbb{A}^B.$$

Proof The main observation of the proof is that $\mathbb{A}^B$ is almost defined as $\mathbb{A}^{\wp}$: the equivalence between $\mathbb{A}^{\wp}$ and $\mathbb{A}^B$ is easier to prove and in fact it suffices to show the claim, because $\mathbb{A}^{\wp}$ is equivalent to $\mathbb{A}$ by proposition 2.23. In order to show the two directions of $\mathbb{A}^{\wp} \equiv \mathbb{A}^B$, we fix a tree $\mathbb{T}$.

($\Rightarrow$) Let $f^{\wp}$ be a winning strategy for $\exists$ in $\mathcal{G}^{\wp} = \mathcal{A}(\mathbb{A}^{\wp}, \mathbb{T})@(a_I^{\wp}, s_I)$. We claim that $f^{\wp}$ is also winning for $\exists$ in $\mathcal{G}^B = \mathcal{A}(\mathbb{A}^B, \mathbb{T})@(a_I^{\wp}, s_I)$. In order to show that, we first observe the following.

CLAIM 9. Every $f^{\wp}$-conform match of $\mathcal{G}^B$ has the same basic positions of an $f^{\wp}$-conform match of $\mathcal{G}^{\wp}$.

PROOF OF CLAIM This is clear by the fact that $\mathbb{A}^B$ and $\mathbb{A}^{\wp}$ are based on the same set of states, the same initial state and the same transition function. ■

The proof that $f^{\wp}$ is winning for $\exists$ in $\mathcal{G}^B$ is completed by showing the following claim.

CLAIM 10. Let π^B be an $f^{\wp}$-conform infinite match of $\mathcal{G}^B$ and ρ the infinite sequence $a_I^{\wp}, R_1, \ldots, R_n, \ldots$ of states of $\mathbb{A}^B$ visited along the play. There is some $R \in A^{\wp}$ occurring infinitely often in ρ such that R is in F_{Ω}.

PROOF OF CLAIM By claim 9 we have an $f^{\wp}$-conform infinite match $\pi^{\wp}$ of $\mathcal{G}^{\wp}$ with the same sequence of states $\rho = R_0, R_1, \ldots, R_n, \ldots$ visited along the play. Since $f^{\wp}$ is winning for $\exists$ in $\mathcal{G}^{\wp}$, all traces through ρ are *good* according to definition 2.17. This means that every trace through ρ corresponds to a match of $\mathcal{G} = \mathcal{A}(\mathbb{A}, \mathbb{T})@(a_I, s_I)$ that is won by $\exists$. Let π be one such match associated with a trace through ρ. Since the automaton $\mathbb{A}$ is weak, exactly *one* parity $n \in Ran(\Omega)$ occurs infinitely often along π, as explained in remark 3.3. By the fact that $\exists$ is the winner and $Ran(\Omega)$ is a subset of $\{0,1\}$, this unique parity n must be 0.

Since π corresponds to an arbitrary trace through ρ, we have that after finitely many steps only positions with parity 0 occur on each trace through ρ. This means that, for some $k < \omega$, there is a state $R_k \in A^{\wp}$ occurring in ρ after which all the states occurring in ρ belong to the set

$$F = \{R \in A^{\wp} \mid \Omega(a) = 0 \text{ for all } a \in Ran(R)\}.$$

Since F_{Ω} is a finite set, there is at least one state $Q \in F_{\Omega}$ occurring infinitely often along ρ. ■

The proof of the claim completes the proof of direction ($\Rightarrow$).

($\Leftarrow$) The argument showing this direction is completely analogous to the one showing direction ($\Rightarrow$), so we just sketch the main steps. Given a winning strategy f^B for $\exists$ in $\mathcal{G}^B$, the same strategy f^B can be shown to be winning for $\exists$ in $\mathcal{G}^{\wp}$. The key observation is that, since f^B is winning for $\exists$ in $\mathcal{G}^B$, for each infinite f^B-conform match $\pi^{\wp}$ of $\mathcal{G}^{\wp}$, there is some $R \in F_{\Omega}$ occurring infinitely often along the play. Let ρ be the sequence of states $R_0, R_1, \ldots, R_n, \ldots$ visited along the play in $\pi^{\wp}$. Every trace α through ρ encounters some $a \in Ran(R)$ infinitely often. By definition of F_{Ω}, this means that basic positions with parity 0 occur infinitely often in the match π_{α} of $\mathcal{G}$ associated with α. It follows that π_{α} is won by $\exists$. Since α was an arbitrary trace through ρ, then every trace through ρ is *good* and $\exists$ also wins $\pi^{\wp}$. □

4.2 The Bounded Information Property

In this section we formalize two essential intuitions about non-deterministic Büchi automata:

1. checking whether a non-deterministic Büchi automaton $\mathbb{B}$ accepts a tree $\mathbb{T}$ reduces to check a condition on prefixes of $\mathbb{T}$ (proposition 4.7);

2. checking whether the intersection of the languages of two non-deterministic Büchi automata is not empty reduces to the construction of a finite sequence of well-founded trees with certain properties (proposition 4.9).

The idea is that a run of a non-deterministic Büchi automaton on a tree $\mathbb{T}$ can be split into several tasks concerning well-founded subtrees (and prefixes, which are just a particular kind of well-founded subtrees) of $\mathbb{T}$, and there is never the need to consider $\mathbb{T}$ as a whole. We informally refer to this as the *Bounded Information Property* of non-deterministic Büchi automata. Intuitively, this is the key property allowing to describe winning strategies for non-deterministic Büchi automata by means of *WFMSO*-formulae - under certain conditions that we will see in the sequel.

Definition 4.6. Let $\mathbb{B} = \langle B, b_I, \Delta, F\rangle$ be a non-deterministic Büchi automaton and $\mathbb{T}$ a tree. Let f be a surviving strategy for $\exists$ in $\mathcal{A}(\mathbb{B},\mathbb{T})@(b_I,s_I)$. Let $\gamma \leq \omega$ be an ordinal. *A γ-accepting sequence for f over $\mathbb{B}$ and $\mathbb{T}$* is a sequence $(E_i)_{i<\gamma}$ such that, for all $i<\gamma$:

1. E_i is a prefix of $\mathbb{T}$;
2. $Ft(E_i) < Ft(E_{i+1})$;
3. for all s in the frontier of E_i, there is a unique $a \in A$ such that $(a,s) \in Dom(f)$; in addition, a is in F.

◁

Intuitively, for $k<\omega$, a k-accepting sequence for a surviving strategy f witnesses the fact that f 'behaves as' a winning strategy for $\exists$ in the prefix E_k of $\mathbb{T}$. For each prefix E_i in the sequence, the condition that each $s \in Ft(E_i)$ is associated with a *unique* accepting state is motivated by the fact that f can be assumed to be full and functional, $\mathbb{B}$ being non-deterministic.

Proposition 4.7. *Let $\mathbb{B} = \langle B, b_I, \Delta, F\rangle$ be a non-deterministic Büchi automaton and $\mathbb{T}$ a tree. The following are equivalent.*

- *Player $\exists$ has a winning strategy in $\mathcal{A}(\mathbb{B},\mathbb{T})@(b_I,s_I)$.*
- *Player $\exists$ has a surviving strategy f in $\mathcal{A}(\mathbb{B},\mathbb{T})@(b_I,s_I)$ and there is an ω-accepting sequence for f over $\mathbb{B}$ and $\mathbb{T}$.*

Proof ($\Rightarrow$) Let f be a winning strategy for $\exists$ in $\mathcal{A}(\mathbb{B},\mathbb{T})@(b_I,s_I)$. By proposition 2.36 and remark 4.3, we can assume f to be full and functional. Let $\mathbb{T}_f$ the tree representation of f.

The idea of the proof is that the sequence $(E_i)_{i<\omega}$ is easily definable on $\mathbb{T}_f$: since a branch of $\mathbb{T}_f$ corresponds to the sequence of basic positions visited along an f-conform match π, there are infinitely many positions occurring in π of the form (b,t) with $b \in F$. We intersect π with each frontier in one of these positions. Then it is sufficient to project this construction on the tree $\mathbb{T}$ to obtain an analogous construction as in the statement.

We provide the formal details following these intuitions. Given a branch S of $\mathbb{T}$, since f is full there is an f-conform match π that is played along S. Since f is winning, there are infinitely many positions $(a_0,s_0),(a_1,s_1),(a_2,s_2),\ldots$ occurring along π such that each a_i is in F. We can order these positions in a sequence $(a_i,s_i)_{i<\omega}$ according to the round of π in which they occur, such that (a_i,s_i) is visited in π before (a_{i+1},s_{i+1}). An infinite sequence $(s_i)_{i<\omega}$ of nodes is induced, such that $s_i R^+ s_{i+1}$ and all nodes in the sequence are in the branch S. For each $i<\omega$, we let E_i be the prefix of $\mathbb{T}$ induced by defining its frontier $Ft(E_i)$ as follows:

⋆ for each branch S of $\mathbb{T}$, let $(s_i)_{i<\omega}$ be the associated sequence of nodes of S, given as above. We put $Ft(E_i) \cap S = \{s_i\}$.

By construction of $(s_i)_{i<\omega}$ it is straightforward to check that $(E_i)_{i<\omega}$ is an ω-accepting sequence for f over $\mathbb{B}$ and $\mathbb{T}$.

($\Leftarrow$) Let f be a surviving strategy for $\exists$ in $\mathcal{G}_{\mathbb{B}} = \mathcal{A}(\mathbb{B},\mathbb{T})@(b_I,s_I)$ and $(E_i)_{i<\omega}$ an ω-accepting sequence for f over $\mathbb{B}$ and $\mathbb{T}$. We claim that f is in fact a winning strategy for $\exists$ in $\mathcal{A}(\mathbb{B},\mathbb{T})@(b_I,s_I)$.

For this purpose, let π be an infinite f-conform match of $\mathcal{G}_{\mathbb{B}}$. Let S be the branch of $\mathbb{T}$ on which π is played. The branch S intersects the frontier $Ft(E_i)$ of each well-founded subtree E_i from the sequence $(E_i)_{i<\omega}$, in a node $s_i \in S \cap Ft(E_i)$. This induces a sequence $(s_i)_{i<\omega}$ of nodes in S. By the fact that $Ft(E_i) < Ft(E_{i+1})$ for each $i<\omega$, all nodes in $(s_i)_{i<\omega}$ are distinct, and furthermore for each $i<\omega$ there is a unique $b \in B$ such that $(b,s_i) \in Dom(f)$ and $b \in F$. This means that the match π played on S visits infinitely many basic positions associated with accepting states of $\mathbb{B}$. Therefore $\exists$ wins π. This suffices to show that f is a winning strategy for $\exists$ in $\mathcal{G}_{\mathbb{B}}$. □

For non-deterministic Büchi automata $\mathbb{B}_1$ and $\mathbb{B}_2$ and a tree $\mathbb{T} \in L(\mathbb{B}_1) \cap L(\mathbb{B}_2)$, let $(G_i^1)_{i<\omega}$ and $(G_i^2)_{i<\omega}$ be ω-accepting sequences respectively for $\mathbb{B}_1$ and $\mathbb{B}_2$ on $\mathbb{T}$. We introduce the notion of *k-trap* for $\mathbb{B}_1$ and $\mathbb{B}_2$. The idea is that a k-trap is a finite sequence $(E_i)_{i\le k}$ witnessing the interleaving of sequences $(G_i^1)_{i<\omega}$ and $(G_i^2)_{i<\omega}$ up to k.

Definition 4.8. Let $\mathbb{B}_1 = \langle B_1, b_I^1, \Delta_1, F_1\rangle$ and $\mathbb{B}_2 = \langle B_2, b_I^2, \Delta_2, F_2\rangle$ be *NDB* automata and let $\mathbb{T}$ be a tree. Given some fixed $k < \omega$, let $(E_i)_{i\le k}$ be a sequence of prefixes of $\mathbb{T}$ such that $E_0 = \{s_I\}$ and $E_i \subsetneq E_{i+1}$ for all $i \le k$.

We say that $\mathbb{T}$ and $(E_i)_{i\le k}$ constitute a *k-trap for* $\mathbb{B}_1$ *and* $\mathbb{B}_2$ if there exist

1. a strategy f_1 for $\exists$ in $\mathcal{A}(\mathbb{B}_1,\mathbb{T})@(b_I^1, s_I)$ which is surviving in E_k,
2. a strategy f_2 for $\exists$ in $\mathcal{A}(\mathbb{B}_2,\mathbb{T})@(b_I^2, s_I)$ which is surviving in E_k,
3. a k-accepting sequence $(G_i^1)_{i\le k}$ for f_1 over $\mathbb{B}^1$ and $\mathbb{T}$,
4. a k-accepting sequence $(G_i^2)_{i\le k}$ for f_2 over $\mathbb{B}^2$ and $\mathbb{T}$,

such that, for all $i < k$, the following conditions hold:

- $Ft(E_i) \le Ft(G_i^1) < Ft(E_{i+1})$;
- $Ft(E_i) \le Ft(G_i^2) < Ft(E_{i+1})$.

We say that the strategies f_1 and f_2 *witness* the k-trap for $\mathbb{B}_1$ and $\mathbb{B}_2$.

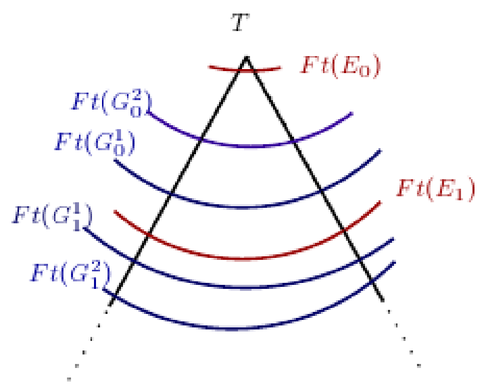

Figure 4.1: initial part of a k-trap

Proposition 4.9 ([28])**.** *Let $\mathbb{B}_1$ and $\mathbb{B}_2$ be NDB automata and let m be the product of the cardinalities of their carriers. If there exists an m-trap for $\mathbb{B}_1$ and $\mathbb{B}_2$ then $L(\mathbb{B}_1) \cap L(\mathbb{B}_2) \neq \varnothing$.*

Proof sketch A detailed proof of this result can be found in [28], proof of theorem 27. In the sequel we confine ourselves to a sketch giving the idea of the argument.

Let $\mathbb{B}_1 = \langle B_1, b_I^1, \Delta_1, F_1\rangle$ and $\mathbb{B}_2 = \langle B_2, b_I^2, \Delta_2, F_2\rangle$ be *NDB* automata. Our initial assumption is that an m-trap for $\mathbb{B}_1$ and $\mathbb{B}_2$ exists. Let $\mathbb{T}$ be the tree associated with the m-trap as in definition 4.8. The idea is that, for each pair of states (b_1, b_2) in $B_1 \times B_2$, the m-trap gives to $\exists$:

- a subtree $\mathbb{T}_{b_1,b_2}$ of $\mathbb{T}$ with root s_I and a prefix E_{b_1,b_2} of $\mathbb{T}_{b_1,b_2}$;
- the information on how to play $\mathcal{G}^1 = \mathcal{A}(\mathbb{B}_1,\mathbb{T}_{b_1,b_2})@(b_1,s_I)$ and $\mathcal{G}^2 = \mathcal{A}(\mathbb{B}_2,\mathbb{T}_{b_1,b_2})@(b_2,s_I)$, in such a way that along each match of $\mathcal{G}^n$ which is played in E_{b_1,b_2} she never gets stuck and a basic position with an accepting state $b \in F^n$ occurs, for each $n \in \{1,2\}$.

Putting together infinitely many copies of those subtrees we can construct a tree $\mathbb{T}_\omega$, whose membership in $L(\mathbb{B}_1) \cap L(\mathbb{B}_2)$ is witnessed by patching together the strategies suggested by the m-trap.

For this purpose, we define a sequence of relations $(H_i)_{i<\omega}$, where each H_i is a subset of $B_1 \times B_2$. For the base case, we set $H_0 = B_1 \times B_2$. For the inductive step, we put $(b_1,b_2) \in H_{i+1}$ if and only if the following conditions hold.

1. $(b_1, b_2) \in H_i$.

2. There is a tree $\mathbb{T}_{b_1,b_2}$ with root s_I, prefixes E_{b_1,b_2}, G_{b_1} and G_{b_2} of $\mathbb{T}_{b_1,b_2}$, strategies f_{b_1} and f_{b_2} for $\exists$ respectively in $\mathcal{A}(\mathbb{B}_1,\mathbb{T}_{b_1,b_2})@(b_1,s_I)$ and $\mathcal{A}(\mathbb{B}_2,\mathbb{T}_{b_1,b_2})@(b_2,s_I)$, such that the following holds.

    a) The strategy f_{b_1} is surviving for $\exists$ in E_{b_1,b_2}.
    b) The strategy f_{b_2} is surviving for $\exists$ in E_{b_1,b_2}.
    c) $Ft(G_{b_1}) < Ft(E_{b_1,b_2})$.
    d) $Ft(G_{b_2}) < Ft(E_{b_1,b_2})$.
    e) For all $s \in Ft(G_{b_1})$, there is a unique position $(b'_1,s) \in B_1 \times T_{b_1,b_2}$ such that (b'_1,s) is a node of the tree representation $\mathbb{T}_{f_{b_1}}$ of f_{b_1}; in addition, b'_1 is in F_1.
    f) For all $s \in Ft(G_{b_2})$, there is a unique position $(b'_2,s) \in B_2 \times T_{b_1,b_2}$ such that (b'_2,s) is a node of the tree representation $\mathbb{T}_{f_{b_2}}$ of f_{b_2}; in addition b'_2 is in F_2.
    g) For all $s \in Ft(E_{b_1,b_2})$, there is a unique pair (b'_1,b'_2) such that (b'_1,s) and (b'_2,s) are nodes respectively of $\mathbb{T}_{f_{b_1}}$ and $\mathbb{T}_{f_{b_2}}$; in addition, (b'_1,b'_2) is in H_i.

Provided this construction, the proof of the main statement is reduced to the proof of the following two facts.

1. If there exists an m-trap for $\mathbb{B}_1$ and $\mathbb{B}_2$ then (b^1_I,b^2_I) is in H_m.
2. If (b^1_I,b^2_I) is in H_m then $L(\mathbb{B}_1) \cap L(\mathbb{B}_2) \neq \varnothing$.

We refer to [28] for a proof of the first fact, showing how the components witnessing $(b^1_I,b^2_I) \in H_m$ are already provided by our m-trap. Instead we focus on the second fact, which is proved by constructing a tree $\mathbb{T}_\omega \in L(\mathbb{B}_1) \cap L(\mathbb{B}_2)$, from the assumption that $(b^1_I,b^2_I) \in H_m$. The key observation is that by definition the sequence $(H_n)_{n<\omega}$ stabilizes at m. This means that $H_m = H_{m+k}$ for all $k < \omega$, as can be shown by a simple combinatorial argument, using the fact that m is the cardinality of $B_1 \times B_2$. Roughly, the argument goes as follows: given $(b_1,b_2) \in H_m$, there is a sequence $(b^i_1,b^i_2)_{i\leq m}$ of pairs, with $(b^i_1,b^i_2) \in H_i$ for each $i \leq m$, which is determined by the property of point 2.g of the definition of $(H_i)_{i<\omega}$. Since $(b^i_1,b^i_2)_{i\leq m}$ has $m+1$ elements, there is at least one pair which is repeated, that is, $(b^l_1,b^l_2) = (b^j_1,b^j_2)$, for some l,j with $l < j < m+1$. This means that we can suitably 'plug' $(b^n_1,b^n_2)_{n<j}$ in place of $(b^n_1,b^n_2)_{n<l}$, to expand the sequence $(b^i_1,b^i_2)_{i\leq m}$ to length $m+1+(j-l)$. This witnesses that (b_1,b_2) was in fact a member of $H_{m+(j-l)}$, and then we can repeat the argument.

Now we proceed with the construction of the tree $\mathbb{T}_\omega \in L(\mathbb{B}_1) \cap L(\mathbb{B}_2)$. The idea is to provide $\mathbb{T}_\omega$ as the limit of a sequence $(\mathbb{T}_i)_{i<\omega}$ of well-founded trees, with T_i a prefix of T_{i+1} for each $i < \omega$. Given $n \in \{1,2\}$, we also construct the graph of a strategy f^n_ω for $\exists$ in $\mathcal{A}(\mathbb{B}_n,\mathbb{T}_\omega)$ as the limit of a sequence $(f^n_i)_{i<\omega}$, with the graph of f^n_i strictly contained in the graph of f^n_{i+1}, for each $i < \omega$. In the sequel we sketch the inductive construction of $(\mathbb{T}_i)_{i<\omega}$ and $(f^n_i)_{i<\omega}$.

- For the first element $\mathbb{T}_0$ in the sequence, we use the assumption that (b^1_I,b^2_I) is in $H_m = H_{m+1}$ to get a tree $\mathbb{T}_{b^1_I,b^2_I}$ with root s_I, prefixes $E_{b^1_I,b^2_I}$, $G_{b^1_I}$ and $G_{b^2_I}$ of $\mathbb{T}_{b^1_I,b^2_I}$, strategies $f_{b^1_I}$ and $f_{b^2_I}$ for $\exists$ respectively in $\mathcal{A}(\mathbb{B}_1,\mathbb{T}_{b^1_I,b^2_I})@(b^1_I,s_I)$ and $\mathcal{A}(\mathbb{B}_2,\mathbb{T}_{b^1_I,b^2_I})@(b^2_I,s_I)$ with the properties given by definition of H_{m+1}. Then we put $\mathbb{T}_0 := E_{b^1_I,b^2_I}$.

  For $n \in \{1,2\}$, the strategy f^n_1 is defined to be the restriction of $f_{b^n_I}$ to basic positions from $B_n \times (E_{b^1_I,b^2_I} \setminus Ft(E_{b^1_I,b^2_I}))$.

- The tree $\mathbb{T}_{i+1}$ will be given as a well-founded extension of $\mathbb{T}_i$. By inductive hypothesis each $t \in Ft(T_i)$ is associated with a pair $(b^1_t,b^2_t) \in H_m$. This also means that (b^1_t,b^2_t) is in H_{m+1}, and we can repeat the same argument that we used for the base case to get a tree $\mathbb{T}_{b^1_t,b^2_t}$ with root t, prefixes $E_{b^1_t,b^2_t}$, $G_{b^1_t}$ and $G_{b^2_t}$ of $\mathbb{T}_{b^1_t,b^2_t}$, strategies $f_{b^1_t}$ and $f_{b^2_t}$ for $\exists$ respectively in $\mathcal{A}(\mathbb{B}_1,\mathbb{T}_{b^1_t,b^2_t})@(b^1_t,t)$ and $\mathcal{A}(\mathbb{B}_2,\mathbb{T}_{b^1_t,b^2_t})@(b^2_t,t)$ with the properties given by definition of H_{m+1}. We define T_{i+1} by putting

$$T_{i+1} \quad = \quad (T_i \setminus Ft(T_i)) \cup \bigcup_{t\in Ft(T_i)} E_{b^1_t,b^2_t}.$$

  By construction T_{i+1} yields a tree structure $\mathbb{T}_{i+1}$, which is induced by $\mathbb{T}_i$ and $\mathbb{T}_{b^1_t,b^2_t}$.

  For each $n \in \{1,2\}$, and $t \in Ft(T_i)$, let $f^E_{b^n_t}$ denote the restriction of $f_{b^n_t}$ to basic positions from $B_n \times E_{b^1_t,b^2_t}$. the strategy f^n_{i+1} is defined by putting

$$f^n_{i+1} \quad = \quad f^n_i \cup \bigcup_{t\in Ft(T_i)} f^E_{b^n_t},$$

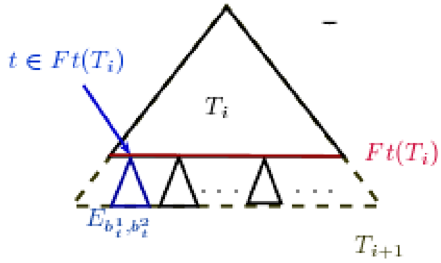

Figure 4.2: construction of $\mathbb{T}_{i+1}$.

where $\cup$ is the union of graphs of functions. In order to check that f_{i+1} is indeed a function, observe that by inductive hypothesis f_i^n is a function with domain $B_n \times (T_i \setminus Ft(T_i))$ and each $f^E_{b_t^n}$ has domain $B_n \times E_{b_t^1,b_t^2}$. All these strategies have disjoint domains by construction. It follows that also f_{i+1} is uniquely defined on each basic position in its domain, which is just $B_n \times T_{i+1}$ by definition of T_{i+1}.

Let r be a node in $Ft(T_{i+1})$. By construction we have that r is in $Ft(E_{b_t^1,b_t^2})$ for some $t \in Ft(T_i)$. Since t is associated with the pair $(b_t^1, b_t^2) \in H_{m+1}$, then by definition of H_{m+1} the node r is associated with a pair $(b_r^1, b_r^2) \in H_m$. This suffices to maintain the inductive hypothesis in the next stage $i+2$.

The proof is concluded by checking that for each $n \in \{1,2\}$ the strategy f_ω^n is winning for $\exists$ in $\mathcal{G}_\omega := \mathcal{A}(\mathbb{B}_n, \mathbb{T}_\omega)@(b_I^n, e)$. For this purpose, the key observation is that an f_ω^n-conform match π_ω of $\mathcal{G}_\omega$ can be seen as an infinite sequence $\pi_1, \pi_2, \ldots, \pi_n$ of partial matches of $\mathcal{G}_\omega$, where each π_i is an f_i^n-conform match played along the well-founded subtree $\mathbb{T}_i$ of $\mathbb{T}$. By definition of f_i^n, along π_i a basic position of the form $(b,s) \in B_n \times T_i$ is visited with $b \in F_n$. This means that some $b \in F_n$ is visited infinitely often in π_ω, which is then won by $\exists$. We refer to [28] for further details. □

4.3 *WFMSO*-Formulae for Büchi Acceptance Conditions

In this section we introduce some auxiliary definitions of *WFMSO*-formulae, as a preparation for the logical characterization of weak *MSO*-automata and *NDB* automata. For the sake of readability, we use the two-sorted presentation of the monadic second-order language, as in remark 1.11. Recall that we indicate with x, y and z variables for nodes, and with p and q variables for sets of nodes.

Definition 4.10 (y-parametrization)**.** Let y be a first-order variable, B a set of unary predicates and $\varphi \in SBF^+(B)$ a sentence in special basic form. We recall that by definition 2.33 the sentence φ is of shape

$$\varphi \quad = \quad \exists x_1 \ldots \exists x_k \; (\mathit{diff}(\bar{x}) \wedge \bigwedge_{1 \le i \le k} a_i(x_i) \wedge \forall z \; (\mathit{diff}(\bar{x}, z) \rightarrow \bigvee_{1 \le l \le j} b_l(z)))$$

where each a_i and b_j is in B. We can assume without loss of generality that y does not occur in φ. The formula $\varphi_{y,B}$ is given as follows.

$$\varphi_{y,B} \quad = \quad \exists x_1 \ldots \exists x_k \; (\bigwedge_{1 \le i \le k} yRx_i \wedge \mathit{diff}(\bar{x}) \wedge \bigwedge_{i \le k} a_i(x_i) \wedge \forall z \; ((yRz \wedge \mathit{diff}(\bar{x}, z)) \rightarrow \bigvee_{1 \le l \le j} b_l(z)))$$

◁

The intuitive idea behind definition 4.10 is to bound the quantifiers of φ to the variable y. The formula $\varphi_{y,B}$ depends on the free variable y and unary predicates from B. We can see each unary predicate occurring in $\varphi_{y,B}$ as a set variable. Therefore $\varphi_{y,B}$ is a well-formed formula of *WFMSO* according to definition 1.9.

Definition 4.11. Let $\mathbb{B} = \langle B, b_I, \Delta, F \rangle$ be a *NDB* automaton and y a first-order variable. Fix some $b \in B$ and $c \in C$. By definition $\Delta(b,c) \in SLatt(SBF^+(B))$ is a disjunction of sentences in special basic form. We let $\Delta_{y,B}(b,c)$ be the *WFMSO*-formula defined by putting

$$\Delta_{y,B}(b,c) \quad := \quad \bigvee \{\varphi_{y,B} \in WFMSO \mid \varphi \text{ is a disjunct of } \Delta(b,c)\},$$

where, for each φ, the sentence $\varphi_{y,B}$ is given as its y-parametrization. ◁

Notation 4.12. Let p be a propositional letter and $\mathbb{T}$ a tree. We indicate with $\|p\|_{\mathbb{T}}$ the set of nodes of $\mathbb{T}$ on which p is true, according to the labeling function $\sigma_C : T \to C$ of $\mathbb{T}$. For the sake of readability, we omit the subscript when the tree $\mathbb{T}$ is clear from the context.

Definition 4.13. Let $B = \{b_1, \dots, b_k\}$ and $P = \{p_1, \dots, p_j\}$ be two finite collection of set variables, representing respectively the states of an *NDB* automaton and the propositional letters forming the labels of a *C*-labeled tree.

Let $\Delta : B \times C \to SLatt(SBF^+(B))$ be a function. Intuitively, given $(a,c) \in B \times C$, the formula $\Delta(a,c) \in SLatt(SBF^+(B))$ represents the value of the transition function of an *NDB* automaton on the state a and a node with label c.

We define a series of formulae of *WFMSO*, providing an informal explanation of their meaning. All valuations are referred to a fixed tree $\mathbb{T}$.

$$\begin{aligned} p \subseteq q \quad &:= \quad \forall x\,(x \in p \to x \in q) \\ Unique(p,q) \quad &:= \quad \exists x\,(x \in p \wedge x \in q \wedge \forall y\,(y \in p \wedge y \in q \to y \approx x)) \end{aligned}$$

The formula $p \subseteq q$ simply states that the set of nodes $\|p\|$ is included in the set of nodes $\|q\|$. The formula $Unique(p,q)$ is true if the intersection of $\|p\|$ and $\|q\|$ is a singleton.

$$\begin{aligned} InCl(p) \quad &:= \quad \forall x\,\forall y\,((xRy \wedge y \in p) \to x \in p) \\ FoCl(p) \quad &:= \quad \forall x\,(x \in p \to (\exists y\,(xRy \wedge y \in p))) \end{aligned}$$

Intuitively, the formula $InCl(p)$ holds for a set of nodes $\|p\|$ if it is backwards closed. Analogously, $FoCl(P)$ holds if $\|p\|$ is frontwards closed.

$$\begin{aligned} xR^\star y \quad &:= \quad \forall p\,((InCl(p) \wedge y \in p) \to x \in p) \\ Root(x) \quad &:= \quad \forall y\,(xR^\star y) \end{aligned}$$

The relation $R^\star$ has the semantics of the reflexive and transitive closure of the successor relation between nodes. Indeed, $xR^\star y$ holds if every backwards closed set $\|p\|$ that includes $\|y\|$ also includes $\|x\|$.

The formula $Root(x)$ holds for any $\|x\|$ which is an ancestor of any other node. For any tree $\mathbb{T}$, the formula $Root(x)$ is true if and only if we take $\|x\|$ to be the root s_I.

$$Path(p) \quad := \quad \forall x\,\forall y\,((x \in p \wedge y \in p \to (xR^\star y \wedge \forall z\,(xR^\star z \wedge zR^\star y \to z \in p)) \vee (yR^\star x \wedge \forall z\,(yR^\star z \wedge zR^\star x \to z \in p))$$

Intuitively, $Path(p)$ holds if $\|p\|$ is a path. The condition we impose is that, for every two nodes $\|x\|$ and $\|y\|$ in $\|p\|$, either $\|x\|$ is an ancestor of $\|y\|$ or $\|y\|$ is an ancestor of $\|x\|$, and all nodes in between are also in $\|p\|$.

$$\begin{aligned} Branch(p) \quad &:= \quad Path(p) \wedge InCl(p) \wedge FoCl(p) \\ Front(p) \quad &:= \quad \forall q\,(Branch(q) \to Unique(p,q)) \end{aligned}$$

The formula $Branch(p)$ holds if and only if the set of nodes $\|p\|$ is a branch. Similarly, $Front(p)$ expresses the fact that $\|p\|$ is a frontier.

$$Pfix_z(p) \quad := \quad \exists q\,(Front(q) \wedge \forall x\,(x \in p \leftrightarrow (zR^\star x \wedge \exists y\,(y \in q \wedge xR^\star y))))$$

The formula $Pfix_z(p)$ depends on a variable z. It expresses the fact that $\|p\|$ is a prefix of $\mathbb{T}.\|z\|$ induced by $\|q\|$, where $\|q\|$ is a frontier of $\mathbb{T}$.

$$\begin{aligned} State_{a,B}(x) \quad &:= \quad x \in a \wedge \bigwedge_{b \in B \setminus \{a\}} \neg(x \in b) \\ Part_B(p) \quad &:= \quad \forall x\,(x \in p \to \bigvee_{a \in B} State_{a,B}(x)) \end{aligned}$$

The formula $State_{a,B}(x)$ holds if a is the only set-variable among the ones in the collection B such that $\|x\| \in \|a\|$.

The formula $Part_B(p)$ holds if the set $\|p\|$ is partitioned according to the set-variables in B.

$$Label_{c,C}(x) \quad := \quad \bigwedge_{p_i \in c} x \in p_i \wedge \bigwedge_{p_i \notin c} \neg(x \in p_i)$$

The formula $Label_{c,C}(x)$ depends on a subset c of P and a first-order variable x. Intuitively, it expresses the fact that, for every propositional letter $p_i \in c$, the node $\|x\|$ is in $\|p_i\|$, i.e. c is the label of $\|x\|$.

$$1Trans_{B,C}(x) \quad := \quad \bigvee_{a \in B} \bigvee_{c \in C} ((State_{a,B}(x) \wedge Label_{c,C}(x)) \rightarrow \Delta_{x,B}(a,c))$$

The formula $1Trans_{B,C}(x)$ depends on B, C and a first-order variable x. We provide an automata-theoretic reading of its meaning. Given a set variable $a \in B$ and a finite collection $c \in C$ of set variables from P, suppose that the node $\|x\|$ has label c and is marked with $\|a\|$. Then the formula expresses that player $\exists$ would not get stuck at the basic position $(a, \|x\|)$, because the successors of $\|x\|$ are marked with states in B in such a way that $\Delta(a,c)$, seen as a *WFMSO*-formula, is true in $\sigma_R(\|x\|)$. The 'relativization' of $\Delta(a,c)$ to $\sigma_R(\|x\|)$ is performed by transforming $\Delta(a,c) \in SLatt(SBF^+(B))$ into $\Delta_{x,B}(a,c) \in WFMSO$ as described in definition 4.11.

$$Trans_{B,C}(p) \quad := \quad \forall x \, (x \in p \rightarrow \ (1Trans_{B,C}(x)))$$

The formula $1Trans_{B,C}(x)$ described a 'good situation' for $\exists$ in the specific case of a node $\|x\|$. The formula $Trans_{B,C}(p)$ just expresses the fact that this holds for all nodes in $\|p\|$.

$$Surv_{B,C}(p) \quad := \quad Part_B(p) \wedge Trans_{B,C}(p)$$

Intuitively, for a suitable $\|p\|$, if $Surv_{B,P}(p)$ holds then $\exists$ is guaranteed to have a legitimate move available from every node in $\|p\|$. Moreover, $Part_B(p)$ expresses that the surviving strategy for $\exists$ in $\|p\|$ assigns a unique state $b \in B$ to each node of $\|p\|$. ◁

4.4 From Non-Deterministic Büchi Automata to *WFMSO*

We are now ready to prove the main result of this chapter.

Theorem 4.14. *For any weak MSO-automaton $\mathbb{A}$, there is a formula $\varphi \in WFMSO$ such that*

for any tree $\mathbb{T}$, $\mathbb{T} \vDash \varphi$ iff $\mathbb{A}$ accepts $\mathbb{T}$.

Before going into details, we gather some intuitions on the argument showing proposition 4.14. Let $\mathbb{A}$ be a weak *MSO*-automaton and $\mathbb{B}$ an *NDB* automaton which is equivalent to $\mathbb{A}$ as in proposition 4.4. Since weak *MSO*-automata are closed under complementation, we are also provided with a weak *MSO*-automaton $\overline{\mathbb{A}}$ recognizing the complement of $L(\mathbb{A})$, and consequently an *NDB* automaton $\overline{\mathbb{B}}$ which is equivalent to $\overline{\mathbb{A}}$.

The idea is to define a formula $\varphi_{\mathbb{B},\overline{\mathbb{B}}} \in WFMSO$ that is true in a tree $\mathbb{T}$ if and only if $\mathbb{B}$ accepts $\mathbb{T}$. Since $\mathbb{B}$ is equivalent to $\mathbb{A}$, this suffices to show that also the formula $\varphi_{\mathbb{B},\overline{\mathbb{B}}}$ is equivalent to $\mathbb{A}$.

Let m be the product of the cardinalities of B and $\overline{B}$. The formula $\varphi_{\mathbb{B},\overline{\mathbb{B}}} \in WFMSO$ will express the existence of a strategy f for $\exists$ and an m-accepting sequence $(E_i)_{i \le m}$ such that f is full, functional and surviving in E_m. The key observation is that the encoding of an m-accepting sequence $(E_i)_{i \le m}$ and the associated surviving strategy into a formula only needs variables for well-closed sets of nodes. This means that the expressive power of *WFMSO* will suffice.

Proposition 4.7 will help showing one direction of the equivalence, namely that, given a tree $\mathbb{T}$ and a winning strategy f for $\exists$ in $\mathcal{A}(\mathbb{B},\mathbb{T})$, the formula $\varphi_{\mathbb{B},\overline{\mathbb{B}}}$ is true in $\mathbb{T}$.

For the converse direction, we use the automaton $\overline{\mathbb{B}}$ accepting the complement of the language of $\mathbb{B}$. The idea is to suppose by way of contradiction that $\overline{\mathbb{B}}$ accepts a tree $\mathbb{T}$ where $\varphi_{\mathbb{B},\overline{\mathbb{B}}}$ is true. Then by proposition 4.7 there is an ω-accepting sequence $(E_i^\delta)_{i<\omega}$ witnessing the fact that $\mathbb{T}$ is in $L(\overline{\mathbb{B}})$. The ω-accepting sequence $(E_i^\delta)_{i<\omega}$ contains an m-accepting sequence $(E_i^\delta)_{i \le m}$. By the fact that $\varphi_{\mathbb{B},\overline{\mathbb{B}}}$ is true, we also have an m-accepting sequence $(E_i)_{i \le m}$. Then we can show that the two sequences witness a trap for $\mathbb{B}$ and $\overline{\mathbb{B}}$ as in definition 4.8. But by proposition 4.9 this means that the intersection of $L(\mathbb{B})$ and $L(\overline{\mathbb{B}})$ is non-empty, contradicting the fact that $\overline{\mathbb{B}}$ accepts the complement of $L(\mathbb{B})$.

In the sequel we formalize the intuitions given above. The first step is to define the formula $\varphi_{\mathbb{B},\overline{\mathbb{B}}}$.

Definition 4.15. Let $\mathbb{B}$ and $\overline{\mathbb{B}}$ be *NDB*-automata, with $B = \{b_1,\ldots,b_k\}$ and $F \subseteq B$ respectively the set of states and of accepting states of $\mathbb{B}$. For each $b \in B$, we define by induction a sequence of formulae $K_i^b(x)$. Put $K_0^b(x) := \top$. The formula $K_{i+1}^b(x)$ is given as follows.

$$K_{i+1}^b(x) \quad := \quad \forall p\ \exists p'\ \exists b_1 \ldots \exists b_k \Bigg(Pfix_x(p) \to \Bigg(p \subseteq p' \wedge Pfix_x(p') \wedge Surv_{B,C}(p') \wedge$$

$$State_{b,B}(x) \wedge \Big(\forall y\ \big(y \in Front(p') \to (\bigvee_{b' \in F} (State_{b',B}(y) \wedge K_i^{b'}(y)))\big)\Big)\Bigg)\Bigg)$$

Let m be the product of the cardinalities of the carriers of $\mathbb{B}$ and $\overline{\mathbb{B}}$. The formula $\varphi_{\mathbb{B},\overline{\mathbb{B}}} \in WFMSO$ is defined by putting

$$\varphi_{\mathbb{B},\overline{\mathbb{B}}} := \exists y\ (Root(y) \wedge K_m^{b_I}(y)).$$

◁

Observe that, for any $k < \omega$, we have that $K_k^b(x) \in WFMSO$. Given a tree $\mathbb{T}$, we supply an intuitive reading of the semantics of $K_{i+1}^b(x)$:

- for each prefix $\|p\|$ of $\mathbb{T}.\|x\|$, there is a prefix $\|p'\|$ of $\mathbb{T}.\|x\|$ including $\|p\|$, and a function $m_p : B \to \wp(\|p\|)$, such that $\exists$ has a full and functional strategy f in $\mathcal{A}(\mathbb{B},\mathbb{T}.\|x\|)@(\|x\|,a)$, which is surviving in $\|p'\|$ and has the following properties:
  - from each basic position $(b_s,s) \in \mathbb{T}_f$, the strategy f suggests to $\exists$ the restriction of m_p to a marking $m_{p,s} : B \to \wp(\sigma_R(s))$;
  - for each node $\|y\|$ on the frontier of $\|p'\|$, the unique $b' \in B$ such that $(b',\|y\|) \in \mathbb{T}_f$ is an accepting state in F, and the formula $K_i^{b'}(y)$ is true in $\mathbb{T}$.

The next is the key lemma in the proof of theorem 4.14.

Proposition 4.16. *Let $\mathbb{B} = \langle B, b_I, \Delta, F\rangle$ and $\overline{\mathbb{B}} = \langle \overline{B}, \overline{b}_I, \overline{\Delta}, \overline{F}\rangle$ be NDB automata such that $L(\overline{\mathbb{B}}) = \overline{L(\mathbb{B})}$. Let $\varphi_{\mathbb{B},\overline{\mathbb{B}}} \in WFMSO$ be constructed from $\mathbb{B}$ and $\overline{\mathbb{B}}$ as in definition 4.15. For any C-labeled tree $\mathbb{T}$, we have that*

$$\mathbb{B} \textit{ accepts } \mathbb{T} \quad \textit{iff} \quad \mathbb{T} \models \varphi_{\mathbb{B},\overline{\mathbb{B}}}.$$

Proof We fix $\mathbb{B}$, $\overline{\mathbb{B}}$ and $\mathbb{T}$ as in the statement. **(⇒)** Let f be a winning strategy for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{B},\mathbb{T})@(b_I,s_I)$. By proposition 2.36 and remark 4.3, we can assume f to be full and functional. Let $\mathbb{T}_f$ and $\pi_2^f : T_f \to T$ be respectively the tree representation of f and the associated projection function. Since f is winning then we are provided with an ω-accepting sequence $(E_i)_{i<\omega}$ for f over $\mathbb{B}$ and $\mathbb{T}$, according to proposition 4.7. Our goal is to show that $\mathbb{T} \models \varphi_{\mathbb{B},\overline{\mathbb{B}}}$. In fact, it suffices to show the following statement.

CLAIM 11. For each $i < \omega$, for each $(b,s) \in B \times T$, if (b,s) is a winning position for $\exists$ in $\mathcal{G}$, then $\mathbb{T} \models K_i^b(x)$, with $\|x\| = s$.

PROOF OF CLAIM We proceed by induction on i. Since $K_0^b(x) = \top$, the base case is trivial. Inductively, let (b,s) be a winning position for $\exists$ in $\mathcal{G}$. We put $\|x\| = s$ and we claim that $\mathbb{T} \models K_{i+1}^b(x)$. Following the syntactic shape of $K_{i+1}^b(x)$, we let $\|p\|$ be an arbitrary prefix E of $\mathbb{T}.s$. By definition of the sequence $(E_i)_{i<\omega}$, for each $i < \omega$ we have that $Ft(E_i) < Ft(E_{i+1})$, implying that there is some prefix E_n in the sequence such that $E \subseteq E_n$. We pick $E_n \cap T.s$ as the witness for the set-variable p' in $K_{i+1}^b(x)$.

We still need to provide witnesses for set-variables $b_1,\ldots,b_k$ occurring in $K_{i+1}^b(x)$. The idea is to let them be suggested by the strategy f. Since f is full and functional, then the projection function $\pi_2^f : T_f \to T$ associated with its tree representation $\mathbb{T}_f$ is 1-1 and onto. For each b_j in $\{b_1,\ldots,b_k\}$, we define its valuation by putting

$$\|b_j\| \quad := \quad \{s \in (E_n \cap T.s) \mid b_j = b_s\}, \tag{4.1}$$

where, for each $s \in E_n$, the state b_s is the first member of the pair $(b_s,s) = (\pi_2^f)^{-1}(s)$. Since $E_n \cap T.s$ is well-founded then $\|b_j\|$ is well-closed, so that it is a legitimate witness for b_j according to the semantics of *WFMSO*.

The subformula $Surv_{B,C}(p')$ of $K_{i+1}^b(x)$ holds because the strategy f is assumed to be full and winning for $\exists$, so in particular it is full and surviving for $\exists$ in $E_n \cap T.s = \|p'\|$. Concerning the subformula $State_{b,B}(x)$, by assumption

(b,s) is a winning position for $\exists$. This means that b is the unique set-variable marking $s = \|x\|$ according to (4.1), so that $State_{b,B}(x)$ holds. It remains to show the statement

$$\forall y\ (y \in Front(p') \rightarrow (\bigvee_{b \in F} State_{b,B}(y) \wedge K_i^b(y))). \tag{4.2}$$

For this purpose, let $\|y\|$ be some node on the frontier of $E_n = \|p'\|$. By (4.1) and the fact that f is full and functional, there is a unique set-variable $\|b\|$ marking $\|y\|$, such that $(b,\|y\|)$ is a node of $\mathbb{T}_f$. Therefore $(b,\|y\|)$ is a winning position for $\exists$ in $\mathcal{G}$, and $K_i^b(y)$ holds by inductive hypothesis. The fact that b is in F follows from properties of the frontier of E_n as in definition 4.6. ∎

By applying claim 11 to the winning position (b_I, s_I) we have that $\mathbb{T} \models K_n^{b_I}(x)$ for each $n < \omega$, with x witnessed by s_I. Then in particular $\mathbb{T} \models \exists x\ (Root(x) \wedge K_m^{b_I}(x))$. This completes the proof of direction $(\Rightarrow)$.

$(\Leftarrow)$ By assumption $\varphi_{\mathbb{B},\overline{\mathbb{B}}}$ is true in $\mathbb{T}$. We need to show that $\mathbb{T}$ is accepted by $\mathbb{B}$.

The idea of the proof is as follows. Suppose by way of contradiction that $\mathbb{B}$ does not accept $\mathbb{T}$. Then the tree $\mathbb{T}$ is accepted by $\overline{\mathbb{B}}$. Let $\overline{f}$ be the winning strategy of $\exists$ in $\mathcal{A}(\overline{\mathbb{B}},\mathbb{T})$. By proposition 2.36 and remark 4.3, we can assume $\overline{f}$ to be full and functional. Suppose that we can prove from the previous assumptions the existence of an m-trap for $\mathbb{B}$ and $\overline{\mathbb{B}}$. Then by proposition 4.9 we have that $L(\mathbb{B}) \cap L(\overline{\mathbb{B}}) \neq \emptyset$, contradicting the fact that $L(\overline{\mathbb{B}}) = \overline{L(\mathbb{B})}$.

In order to complete the proof of direction $(\Leftarrow)$, it remains to verify the following claim.

Claim 12. There exists an m-trap for $\mathbb{B}$ and $\overline{\mathbb{B}}$.

Proof of Claim By definition 4.8, we have to to provide the following components:

1. a strictly increasing sequence $(E_i)_{i \le m}$ of prefixes of $\mathbb{T}$, with $E_0 = \{s_I\}$;

2. a strategy $f^{\mathbb{B}}$ for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{B},\mathbb{T})@(b_I,s_I)$ which is surviving for $\exists$ in E_m;

3. a strategy $f^{\overline{\mathbb{B}}}$ for $\exists$ in $\overline{\mathcal{G}} = \mathcal{A}(\overline{\mathbb{B}},\mathbb{T})@(\overline{b}_I,s_I)$ which is surviving for $\exists$ in E_m;

4. an m-accepting sequence $(G_i^B)_{i \le m}$ for $f^{\mathbb{B}}$ over $\mathbb{B}$ and $\mathbb{T}$;

5. an m-accepting sequence $(G_i^{\overline{B}})_{i \le m}$ for $f^{\overline{\mathbb{B}}}$ over $\overline{\mathbb{B}}$ and $\mathbb{T}$.

Moreover, $(E_i)_{i \le m}$, $(G_i^B)_{i \le m}$ and $(G_i^{\overline{B}})_{i \le m}$ have to present the interleaving behavior described in definition 4.8.

We put the strategy $\overline{f}$ as witness for $f^{\overline{\mathbb{B}}}$. By assumption $\overline{f}$ is a winning strategy for $\exists$ in $\overline{\mathcal{G}}$. Then, by proposition 4.7, we are also given with an ω-accepting sequence $(E_i^{\overline{f}})_{i<\omega}$ for $\overline{f}$ over $\overline{\mathbb{B}}$ and $\mathbb{T}$.

It remains to define the other components of the m-trap, which is what we do next. The idea is to define the surviving strategy $f^{\mathbb{B}}$, the sequences $(E_i)_{i \le m}$ and $(G_i^B)_{i \le m}$ by using the assumption that $\mathbb{T} \models \varphi_{\mathbb{B},\overline{\mathbb{B}}}$. The last component, namely the sequence $(G_i^{\overline{B}})_{i \le m}$, will be defined from $(E_i^{\overline{f}})_{i<\omega}$.

The construction of the strategy $f^{\mathbb{B}}$ and the sequences $(E_i)_{i \le m}$, $(G_i^B)_{i \le m}$ and $(G_i^{\overline{B}})_{i \le m}$ proceeds in stages, by induction on $i \le m$. In particular, $f^{\mathbb{B}}$ will be defined as the last element $f_m^{\mathbb{B}}$ in a sequence of strategies $(f_i^{\mathbb{B}})_{i \le m}$.

Given $i \le m$, the inductive hypothesis that we want to maintain along the construction can be expressed as follows.

1. If $i < m$ then $Ft(E_i) \leq Ft(G_i^B) < Ft(E_{i+1})$.
   Otherwise $i = m$ and $Ft(E_i) \leq Ft(G_i^B)$.
2. If $i < m$ then $Ft(E_i) \leq Ft(G_i^{\overline{B}}) < Ft(E_{i+1})$.
   Otherwise $i = m$ and $Ft(E_i) \leq Ft(G_i^{\overline{B}})$.
3. The sets E_i, G_i^B $G_i^{\overline{B}}$ are prefixes of $\mathbb{T}$.
4. The function $f_i^{\mathbb{B}}$ is a strategy $\exists$ in $\mathcal{G}$ which is full, functional and surviving in G_i^B. If $i \geq 1$, then $f_i^{\mathbb{B}}$ extends $f_{i-1}^{\mathbb{B}}$.
5. For each node $s \in Ft(G_i^{\overline{B}})$, there is a unique $\overline{b}_s \in \overline{B}$ such that the position $(\overline{b}_s, s)$ is a node of the tree representation $\mathbb{T}_{\overline{f}}$ of $\overline{f}$; in addition, $\overline{b}_s$ is in $\overline{F}$.
6. For each node $s \in Ft(G_i^B)$, there is a unique $b_s \in B$ such that the formula $K_{m-i}^{b_s}(x)$ holds for $s = \|x\|$. The position (b_s, s) is a node of the tree representation $\mathbb{T}_{f_i^{\mathbb{B}}}$ of $f_i^{\mathbb{B}}$; in addition, b_s is in F.

(‡)

Let us first show why the different components form an m-trap if condition (‡) can be maintained. By (‡.4) the strategy $f^{\mathbb{B}} = f_m^{\mathbb{B}}$ for $\exists$ in $\mathcal{G}$ would be full, functional and surviving in G_m^B. By (‡.1) we have that $Ft(E_m) \leq Ft(G_m^B)$, meaning that $f^{\mathbb{B}}$ is also surviving in E_m, as requested by point 1 of the definition of m-trap (definition 4.8). For $f^{\overline{\mathbb{B}}} = \overline{f}$, we know by assumption that $\overline{f}$ is full, functional and winning for $\exists$ in $\overline{\mathcal{G}}$. Since E_m is a subset of T, then $f^{\overline{B}}$ is also surviving in E_m, as requested by point 2 of definition 4.8.

For points 3 and 4 of definition 4.8, we have to check that $(G_i^B)_{i \leq m}$ and $(G_i^{\overline{B}})_{i \leq m}$ are m-accepting sequences respectively for $f^{\mathbb{B}}$ and $f^{\overline{\mathbb{B}}}$. For this purpose, there are three conditions to check according to the definition of accepting sequence (definition 4.6). The first condition is that $(G_i^B)_{i \leq m}$ and $(G_i^{\overline{B}})_{i \leq m}$ are sequences of prefixes, which is given by (‡.3). The second condition, on the relation between frontiers of each G_i^B, $G_i^{\overline{B}}$ and E_i, is given by (‡.1) and (‡.2). Concerning the third condition of definition 4.6, for each $i \leq m$, the requirements on $Ft(G_i^B)$ and $Ft(G_i^{\overline{B}})$ are fulfilled by (‡.5) and (‡.6).

The last two points of definition 4.8, concerning the interleaving of the frontiers of $(E_i)_{i \leq m}$, $(G_i^B)_{i \leq m}$ and $(G_i^{\overline{B}})_{i \leq m}$, just correspond to (‡.1) and (‡.2). Therefore what we obtain is indeed an m-trap, provided that we are able to maintain condition (‡).

Now we proceed with the inductive construction. For the base case, let $E_0 := \{s_I\}$. We define the first element $G_0^{\overline{B}}$ in the sequence $(G_i^{\overline{B}})_{i \leq m}$ as the smallest prefix in the sequence $(E_i^{\overline{f}})_{i < \omega}$ such that $E_0 \subseteq G_0^{\overline{B}}$, that is simply $E_0^{\overline{f}}$ because $(E_i^{\overline{f}})_{i < \omega}$ is monotone.

In order to define G_0^B, we observe that the unique witness for x in $\exists x\, (Root(x) \wedge K_m^{b_I}(x))$ must be s_I. Then, by putting E_0 as the witness of the variable p in $K_m^{b_I}(x)$, we are provided with a prefix G_0^B witnessing the variable p' in $K_m^{b_I}(x)$. We let such G_0^B be the first element in the sequence $(G_i^B)_{i \leq m}$.

In order to define the first surviving strategy $f_0^{\mathbb{B}}$ in the sequence $(f_i^{\mathbb{B}})_{i \leq m}$, we fix valuations $\|p\| = E_0$ and $\|p'\| = G_0^B$ in the formula $K_m^{b_I}(x)$ and we consider the witnesses for set-variables $b_1, \ldots, b_k$ in $K_m^{b_I}(x)$. By definition of $K_m^{b_I}(x)$, for each node s in G_0^B there is a unique $b_s \in \{b_1, \ldots, b_k\}$ such that $s \in \|b_s\|$. This yields a strategy $f_0^{\mathbb{B}}$ for $\exists$ in $\mathcal{G}$, which we define as follows:

1. $f_0^{\mathbb{B}}$ is defined at the basic position (b_I, s_I);
2. given a basic position $(b_s, s) \in B \times G_0^B$ with $s \notin Ft(G_0^B)$, we let $f_0^{\mathbb{B}}$ suggest to $\exists$ a marking assigning b_t to t, for each $t \in \sigma_R(s)$;
3. we leave $f_0^{\mathbb{B}}$ undefined on all other basic positions from $B \times T$.

To make the definition of $f_0^{\mathbb{B}}$ more clear, observe that by construction the tree representation of $f_0^{\mathbb{B}}$ is in bijective correspondence with G_0^B. In other words, $f_0^{\mathbb{B}}$ is a strategy for partial match of $\mathcal{G}$ which are played along nodes of G_0^B.

Given prefixes $G_0^{\overline{B}}$ and G_0^B as above, we define E_1 to be the smallest prefix of $\mathbb{T}$ such that $Ft(G_0^B) < Ft(E_1)$ and $Ft(G_0^{\overline{B}}) < Ft(E_1)$.

It remains to check that conditions $1-5$ in (‡) hold for the base case. Condition (‡.1), (‡.2) and (‡.3) are clear by construction of E_0, $G_0^{\overline{B}}$, G_0^B and E_1. For condition (‡.4), by assumption we have that $State_{b_I,B}(y)$ and $Surv_{B,P}(p')$ hold, being subformulae of $K_m^{b_I}(x)$, with $\|p'\| = G_0^B$ and $\|y\| = s_I$. By construction of the strategy $f_0^{\mathbb{B}}$, this means that $f_0^{\mathbb{B}}$ is full, functional and surviving for $\exists$ in G_0^B. Analogously, (‡.5) holds because the subformula of $K_m^{b_I}(x)$ given as in (4.2) is true, meaning that every node on the frontier of G_0^B is associated with a unique accepting state of $\mathbb{B}$ according to $f_0^{\mathbb{B}}$. In order to fulfill condition (‡.6), we observe that, by definition of $K_m^{b_I}(x)$, every node $s \in Ft(G_0^B)$ is associated with a basic position $(b_s, s) \in B \times T$, such that $b_s \in F$ and $K_{m-1}^{b_s}(x)$ holds for $s = \|x\|$.

Inductively, we consider the stage $j < m$ of the construction. By inductive hypothesis, we are given with sequences $(E_i)_{i \le j+1}$, $(G_i^B)_{i \le j}$, $(G_i^{\overline{B}})_{i \le j}$ and a surviving strategy $f_i^{\mathbb{B}}$ for $\exists$ in $\mathcal{G}_{E_i}$ as in (‡).

For the definition of $G_{j+1}^{\overline{B}}$, as in the base case we let it be the smallest prefix in the sequence $(E_i^{\overline{f}})_{i<\omega}$ which contains E_{j+1}. For the definition of G_{j+1}^B and $f_{j+1}^{\mathbb{B}}$, the key observation is that, by inductive hypothesis, for each node $s \in Ft(G_j^B)$ we can make the following assumptions:

1. the formula $K_{m-j}^{b_s}(x)$ holds, with $s = \|x\|$;

2. the position (b_s, s) is a node of the tree representation of $f_j^{\mathbb{B}}$.

We let $T.s \cap E_{j+1}$ be the witness for the set-variable p occurring in $K_{m-j}^{b_s}(x)$. Then by definition of $K_{m-j}^{b_s}(x)$ we are provided with a prefix $G_{j+1}^{B,s}$ of $\mathbb{T}.s$ witnessing the variable p', such that $T.s \cap E_{j+1} \subseteq G_{j+1}^{B,s}$. Also we are provided with well-closed sets of nodes witnessing variables $b_1, \dots, b_k$. Analogously to the base case, this yields a strategy $f_{j+1}^{B,s}$ for $\exists$ in $\mathcal{G}$, which is defined as follows:

1. $f_{j+1}^{\mathbb{B}}$ is defined at the basic position (b_s, s);

2. for each basic position $(b_t, t) \in B \times G_{j+1}^{B,s}$ with $t \notin Ft(G_{j+1}^{B,s})$, we let $f_{j+1}^{\mathbb{B}}$ suggest to $\exists$ a marking assigning b_r to r, for each $r \in \sigma_R(t)$.

3. we leave $f_{j+1}^{\mathbb{B}}$ undefined on all other basic positions from $B \times T$.

To make the definition of $f_{j+1}^{B,s}$ more clear, observe that $f_{j+1}^{B,s}$ is a strategy for $\exists$ in partial matches of $\mathcal{A}(\mathbb{A}, \mathbb{T})@(b_s, s)$, which is full, functional, surviving in $G_{j+1}^{B,s}$ and marks each node $t \in Ft(G_{j+1}^{B,s})$ with a unique state from F.

We define G_{j+1}^B by putting

$$G_{j+1}^B := G_j^B \cup \bigcup_{s \in Ft(G_j^B)} G_{j+1}^{B,s}.$$

Since G_j^B is a prefix of $\mathbb{T}$ and for each $s \in Ft(G_j^B)$ the set $G_{j+1}^{B,s}$ is a prefix of $\mathbb{T}.s$, we have that G_{j+1}^B is a prefix of $\mathbb{T}$. Next, we define f_{j+1}^B by putting

$$f_{j+1}^B := f_j^B \cup \bigcup_{s \in Ft(G_j^B)} f_{j+1}^{B,s},$$

where the union of strategies just means the union of their graphs. In order to check that f_{j+1}^B is indeed a function, observe that by inductive hypothesis the domain of f_j^B is $B \times (G_j^B \setminus Ft(G_j^B))$. By construction, for each $s \in Ft(G_j^B)$, the domain of the strategy $f_{j+1}^{B,s}$ is the union of (b_s, s) and $B \times (G_{j+1}^{B,s} \setminus (G_j^B \cup Ft(G_{j+1}^{B,s})))$. Since $(b_s, s) \in Ft(G_j^B)$, then the domains of f_j^B and each $f_{j+1}^{B,s}$ are all disjoints. Therefore f_{j+1}^B is uniquely defined on each basic position in its domain.

Given $G_{j+1}^{\overline{B}}$ and G_{j+1}^B as above, if $j+1 < m$ then we define E_{j+2} to be the smallest prefix of $\mathbb{T}$ such that $Ft(G_{j+1}^B) < Ft(E_{j+2})$ and $Ft(G_{j+1}^{\overline{B}}) < Ft(E_{j+2})$. The check that all conditions in (‡) hold for G_{j+1}^B, f_{j+1}^B and E_{j+2}^B is completely analogous to the base case.

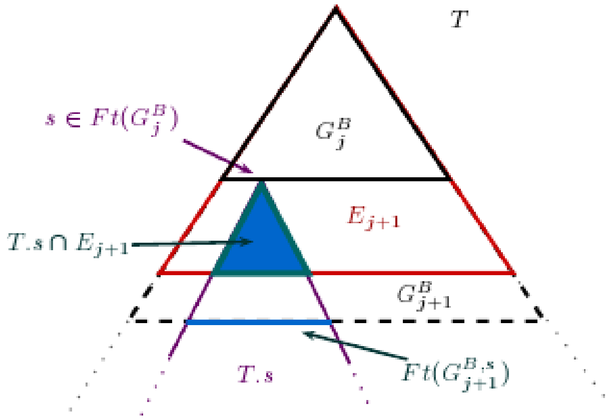

Figure 4.3: construction of G^B_{j+1}.

We have just defined a strategy $f^{\mathbb{B}}$, sequences $(E_i)_{i\leq m}$, $(G^B_i)_{i\leq m}$ and $(G^{\overline{B}}_i)_{i\leq m}$, such that for each $i \leq m$ condition $(\ddagger)$ is respected. It follows that $\overline{f}$ and $f^{\mathbb{B}}$ witness a trap for $\mathbb{B}$ and $\overline{B}$ according to definition 4.8. This concludes the proof of the claim. ■

The proof of claim 12 completes the proof of direction $(\Leftarrow)$.

□

We are now ready to supply a proof for the main statement.

Proof of theorem 4.14 Given a weak *MSO*-automaton $\mathbb{A}$, let $\overline{\mathbb{A}}$ be the weak *MSO*-automaton recognizing the complement of $\mathbb{A}$ according to proposition 2.31. Let $\mathbb{B}$ and $\overline{\mathbb{B}}$ be *NDB*-automata equivalent respectively to $\mathbb{A}$ and $\overline{\mathbb{A}}$, as in definition 4.4. Consider the formula $\varphi_{\mathbb{B},\overline{\mathbb{B}}} \in WFMSO$ defined for $\mathbb{B}$ and $\overline{\mathbb{B}}$ as in definition 4.15. By proposition 4.16 the formula $\varphi_{\mathbb{B},\overline{\mathbb{B}}}$ is equivalent to $\mathbb{B}$, meaning that it is also equivalent to $\mathbb{A}$. □

As a corollary we obtain the following characterization of *WFMSO*.

Corollary 4.17. *A tree language $\mathcal{L}$ is WFMSO-definable if and only if there are non-deterministic Büchi automata $\mathbb{B}$ and $\overline{\mathbb{B}}$ such that $\mathcal{L} = L(\mathbb{B})$ and $\overline{\mathcal{L}} = L(\overline{\mathbb{B}})$.*

Proof Let $\varphi \in WFMSO$ be a formula. By proposition 3.13 we can construct weak *MSO*-automata $\mathbb{A}$ and $\overline{\mathbb{A}}$ equivalent respectively to φ and to $\neg\varphi$. By proposition 4.17 we can also construct non-deterministic Büchi automata $\mathbb{B}$ and $\overline{\mathbb{B}}$ that are equivalent respectively to φ and to $\neg\varphi$.

Conversely, suppose that we have a tree language $\mathcal{L}$ and non-deterministic Büchi automata $\mathbb{B}$ and $\overline{\mathbb{B}}$ recognizing respectively $\mathcal{L}$ and its complement. By theorem 4.14 we can construct a formula $\varphi \in WFMSO$ that is equivalent to $\mathbb{B}$. □

Remark 4.18. Contrary to the case of weak *MSO*-automata, corollary 4.17 does not provide a full logical characterization of *NDB* automata in terms of *WFMSO*. The reason is that the tree languages recognized by *NDB* automata are not closed under complementation [28]. From this result it also follows that weak *MSO*-automata are strictly weaker than *NDB* automata. ◂

Historical Notes

Non-deterministic Büchi automata are a generalization on arbitrarily branching trees of the 'special automata', working on binary trees, which have been introduced by Rabin [28]. The whole construction leading to a logical characterization of *NDB*-automata, including the notion of k-trap (definition 4.8) and the inductive definition of $K^a_i(x)$ (proposition 4.14), is essentially a (game-theoretic) generalization of Rabin's argument showing how special automata relate to *WMSO*-definable languages on binary trees [28].

Chapter 5

Expressivity Results

The previous part of this thesis was devoted to the characterization of *MSO* and *WFMSO* by means of automata. Since the work of Rabin [27], the expressive power of these automata has been studied and compared on different kinds of structures. In this chapter we use some of these results to investigate expressivity questions on the side of logic. In particular, we are interested in comparing the expressive power of *MSO*, *WFMSO* and *WMSO* on different classes of trees. The automata-theoretic perspective plays an essential role in revealing the precise nature of their expressiveness, and how the landscape of logical definability changes by considering trees with different branching degree.

A key result, underlying most of the theory developed in this chapter, is Rabin's observation that there is a tree language which is not accepted by any non-deterministic Büchi automaton [28].

Proposition 5.1 ([28]). *Let $\mathcal{T}_{Bin}$ be the class of binary trees. Let $\mathcal{L}_{Fp}$ be the tree language defined by putting*

$$\mathbb{T} \in \mathcal{L}_{Fp} \quad \textit{iff} \quad \textit{every path of } \mathbb{T} \textit{ contains only finitely many nodes labeled with } p. \tag{5.1}$$

For any NDB automaton $\mathbb{B}$, we have that

$$\mathcal{L}_{Fp} \cap \mathcal{T}_{Bin} \neq L(\mathbb{B}) \cap \mathcal{T}_{Bin}.$$

Proof reference We refer to [28], section 3 and [11], theorem 8.6 for a proof of this result. □

In order to show the significance of this result, recall from the previous chapter that an essential feature of *NDB* automata is the so-called *bounded information property*. Intuitively, if $\mathcal{L}_{Fp}$ was accepted by some *NDB* automaton, that would mean that checking whether $\mathbb{T} \in \mathcal{L}_{Fp}$ amounts to check some property on well-founded subtrees of $\mathbb{T}$. Suppose by way of contradiction that this is the case for some *NDB* automaton $\mathbb{B}$. It is possible to construct a tree $\mathbb{T}$, where the letter p occurs finitely many times on each path, but we can 'reassemble' well-founded subtrees of $\mathbb{T}$ to obtain a tree $\mathbb{T}'$ with a path where p occurs infinitely often. Since $\mathbb{T}$ is in $\mathcal{L}_{Fp}$, then $\exists$ has a winning strategy f in $\mathcal{A}(\mathbb{B},\mathbb{T})@(b_I,s_I)$. However, by the fact that $\mathbb{B}$ is a *NDB* automaton, what f does is essentially to check the presence of accepting states for infinitely many well-founded subtrees of $\mathbb{T}$ visited along the play. This means that we can decompose f into a bundle of strategies for matches on well-founded subtrees of $\mathbb{T}$. Since $\mathbb{T}'$ has been obtained by patching together well-founded subtrees of $\mathbb{T}$, we can construct a strategy f' for $\exists$ in $\mathcal{A}(\mathbb{B},\mathbb{T}')@(b_I,s_I)$ out of the bundle, such that f' is winning because each strategy in the bundle allows $\exists$ to encounter an accepting state of $\mathbb{B}$. Then we have that $\mathbb{T}'$ is in $L(\mathbb{B})$. However, by construction it also holds that $\mathbb{T}' \notin \mathcal{L}_{Fp}$, contradicting the assumption that $L(\mathbb{B}) = \mathcal{L}_{Fp}$.

The informal argument that we just sketched shows that the tree property associated with $\mathcal{L}_{Fp}$ cannot be tested on finite segments of paths (as an *NDB* automaton would do), but we need to store information on each path of $\mathbb{T}$ taken as a whole. This is a key example showing why *NDB* automata are not suitable to capture properties on the *vertical dimension* of trees, such as the one expressed by $\mathcal{L}_{Fp}$.

Since every weak *MSO*-automaton can be turned into an equivalent *NDB* automaton (proposition 4.5), $\mathcal{L}_{Fp}$ is also not accepted by any weak *MSO*-automaton. However, it is quite easy to define an *MSO*-automaton for $\mathcal{L}_{Fp}$.

Remark 5.2. Let $\mathbb{A}_{Fp} = \langle A, a_I, \Delta, \Omega \rangle$ be the *MSO*-automaton defined by putting

$$\begin{array}{rcl} A & := & \{a_0, a_1\} \\ a_I & := & a_0 \\ \Delta(a_0, c) & := & \forall x\, (a_0(x) \vee a_1(x)) \\ \Delta(a_1, c) & := & \begin{cases} \forall x\, (a_0(x) \vee a_1(x)) & \text{If } p \notin c \\ \bot & \text{Otherwise} \end{cases} \\ \Omega(a_0) & := & 1 \\ \Omega(a_1) & := & 2. \end{array}$$

It is not hard to see that $\mathcal{L}_{Fp} = L(\mathbb{A}_{Fp})$. In order to check this, let $\mathbb{T}$ be a tree. We look at which strategies the two players should follow in $\mathcal{G} = \mathcal{A}(\mathbb{A}_{Fp}, \mathbb{T})@(a_I, s_I)$. Given any position $(a,s) \in A \times T$, player $\exists$ should mark with a_0 each node $t \in \sigma_R(s)$ which is labeled with p, and with a_1 all the others. If there is a path S containing infinitely many nodes labeled with p, player $\forall$ is allowed to keep any match of $\mathcal{G}$ on S. Then the minimum parity occurring infinitely often along the play is 1 and $\exists$ looses. If $\mathbb{T}$ does not contain any such path S, then the only parity occurring infinitely often is 2 and $\exists$ wins. By this argument it is clear that indeed $\mathcal{L}_{Fp} = L(\mathbb{A}_{Fp})$.

As a side observation, note that the automaton $\mathbb{A}_{Fp}$ has a parity map ranging over $\{1,2\}$. This can be seen as a 'co-Büchi' acceptance condition, because $\forall$ wins an infinite match if and only if some state with parity 1 occurs infinitely often along the play. The essential difference with a Büchi condition is that, in case both an accepting and a rejecting state occur infinitely often, the rejecting one prevails. ◂

On the base of these observations on the language $\mathcal{L}_{Fp}$, we can relate the expressive power of weak *MSO*-automata and *MSO*-automata.

Proposition 5.3 ([28])**.** *The class of tree languages recognized by weak MSO-automata is strictly included in the class of tree languages recognized by MSO-automata. The same result holds if we restrict to finitely branching trees.*

Proof The inclusion is immediate by the fact that every weak *MSO*-automaton is also an *MSO*-automaton, as observed in remark 3.2. In order to check that the inclusion is strict, let $\mathcal{L}_{Fp}$ be a tree language defined as in (5.1). By proposition 5.1 it follows that $\mathcal{L}_{Fp}$ is not definable by any weak *MSO*-automaton. However, by remark 5.2, the language $\mathcal{L}_{Fp}$ is recognized by some *MSO*-automaton. The same result holds on finitely branching trees because proposition 5.1 holds already on binary trees. □

5.1 The Finitely Branching Case

The result of proposition 5.3 suggests that the class of finitely branching trees is already a reliable benchmark to study the expressive power of *MSO* and *WFMSO*. In fact, the branching degree is not relevant as far as properties of the *vertical dimension* of trees are concerned.

The essential difference between the general setting and the setting of finitely branching trees concerns the *horizontal dimension* of trees, which turns out to be essentially simpler to describe in the finitely branching case. From the point of view of expressiveness, this amounts to a coarser distinction between tree properties. The key example is given by notion of *well-founded* tree, which collapses by Kőnig's Lemma to the one of *finite* tree in the finitely branching case. By this observation, *WMSO* and *WFMSO* become indistinguishable logics on finitely branching trees.

Remark 5.4. Recall the semantics of *WMSO* as in definition 1.10. We observe that *WMSO* and *WFMSO* are the same logic on finitely branching trees. In order to see that, let φ be a formula written in the monadic second-order language as in definition 1.9. We can easily verify by induction that φ defines the same tree language, either if we interpret φ according to the semantics of *WMSO* or *WFMSO*. For the non-trivial case, let $\varphi = \exists p.\psi$. If we interpret φ according to the semantics of *WMSO*, then we have that

$$\mathbb{T} \vDash \exists p.\psi \quad \textit{iff} \quad \mathbb{T}^p \vDash \psi \text{ for some finite } p\text{-variant } \mathbb{T}^p \text{ of } \mathbb{T}. \tag{5.2}$$

Otherwise, if we interpret φ according to the semantics of *WFMSO*, then we have that

$$\mathbb{T} \vDash \exists p.\psi \quad \textit{iff} \quad \mathbb{T}^p \vDash \psi \text{ for some well-closed } p\text{-variant } \mathbb{T}^p \text{ of } \mathbb{T}. \tag{5.3}$$

Now observe that, if $\mathbb{T}$ is a finitely branching tree, then by Kőnig's Lemma a well-closed subset of $\mathbb{T}$ is just a finite subset of $\mathbb{T}$. This means that (5.2) and (5.3) express equivalent conditions on finitely branching trees. ◂

Since *WMSO* and *WFMSO* are the same logic from the finitely branching point of view, we can exploit the automata-theoretic perspective we introduced for *WFMSO* to derive expressivity results also for *WMSO*.

Proposition 5.5. *On finitely branching trees, the class of WMSO-definable tree languages is strictly included in the class of MSO-definable tree languages.*

Proof By proposition 5.3 and remark 2.32 it suffices to show that, on finitely branching trees, the class of *WMSO*-definable tree languages is included in the class of tree languages recognized by weak *MSO*-automata.

For this purpose, let $\varphi \in WMSO$ be a formula. By remark 5.4 there is a formula $\varphi' \in WFMSO$ which is equivalent to φ on finitely branching trees. By theorem 4.14, there is a weak *MSO*-automaton $\mathbb{A}$ which is equivalent to φ'. Then clearly $\mathbb{A}$ is also equivalent to φ' on finitely branching trees. □

5.2 The Arbitrarily Branching Case

As we observed, the restriction to finitely branching trees essentially affects the horizontal dimension of trees, making it sufficiently coarse to identify *WMSO*-expressivity and *WFMSO*-expressivity. Both these logics turn out to be essentially weaker than *MSO* in expressing properties of the *vertical dimension* on trees, such as the language $\mathcal{L}_{Fp}$ presented in proposition 5.1. Since this is the only 'relevant' dimension of the finitely branching case, *WFMSO* and *WMSO* are strictly weaker than *MSO* on finitely branching trees, as shown in proposition 5.5.

In this section we observe how this landscape of connections between *WMSO*, *WFMSO* and *MSO* is affected by considering the more general setting of trees with arbitrary branching degree.

The first observation is that the relation between *WFMSO* and *MSO* remains unaltered.

Proposition 5.6. *The class of WFMSO-definable tree languages is strictly included in the class of MSO-definable tree languages.*

Proof This is an immediate consequence of proposition 5.3, theorem 4.14 and remark 2.32. □

The weakness of *WFMSO* on the vertical dimension is somehow reflected by the semantics of *WFMSO* quantifiers: a well-closed set of nodes is limited on the vertical dimension, being included in a well-founded tree. However, it has no limitations of cardinality on the horizontal dimension, meaning that *MSO*-quantification and *WFMSO*-quantification do not differ much in this respect.

The deeper reason underlying this observation is that *MSO* itself is a rather coarse logic, once it comes to specify properties on the horizontal dimension of trees. In fact, *MSO* cannot even distinguish between trees with finite or infinite branching degree. We indicate this phenomenon as the *Finite Branching Property* of *MSO*. Its significance is best explained through the automata-theoretic perspective, as shown by the next proposition.

Proposition 5.7 (**Finite Branching Property**). *Let $\mathbb{A}$ be an MSO-automaton. If $L(\mathbb{A})$ is non-empty then there is a finitely branching tree $\mathbb{T}$ in $L(\mathbb{A})$.*

Proof Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be an *MSO*-automaton and $\mathbb{T}$ a tree that is accepted by $\mathbb{A}$. By proposition 2.14 we can assume $\mathbb{A}$ to be non-deterministic. Let f be a winning strategy for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$, which we can assume to be full and functional by proposition 2.36 and 2.36. Let $\mathbb{T}_f$ and $\pi_2^f : T_f \to T$ be respectively the tree representation of f and its projection function. As observed in remark 2.35, since f is full and functional then π_2^f is 1-1 and onto.

The main idea of the proof is that we can prune $\mathbb{T}_f$ until we get a finitely branching subtree $\mathbb{T}'_f$ of $\mathbb{T}_f$. We can do it in such a way that $\mathbb{T}'_f$ is the tree representation of a winning strategy for $\exists$ in $\mathcal{G}' = \mathcal{A}(\mathbb{A},\mathbb{T}')@(a_I,s_I)$, with $\mathbb{T}'$ a finitely branching subtree of $\mathbb{T}$ induced by $\mathbb{T}'_f$ itself. This implies that $\mathbb{T}'$ is in $L(\mathbb{A})$.

We proceed with the formal part of the proof. We define a tree $\mathbb{T}'_f$ by induction as follows:

1. The root of $\mathbb{T}'_f$ is the root of $\mathbb{T}_f$, i.e. the position (a_I,s_I).

2. Suppose that (a,s) is a node of $\mathbb{T}'_f$ of height i. By inductive hypothesis the position (a,s) is also a node of $\mathbb{T}_f$ of height i, and the label of (a,s) in $\mathbb{T}'_f$ is the label of (a,s) in $\mathbb{T}_f$.

   Let $m_f : A \to \wp\sigma_R(s)$ be the marking provided by f from position (a,s) and $\varphi \in SBF^+(A)$ be a disjunct of $\Delta(a,\sigma_C(s)) \in SLatt(SBF^+(A))$, such that $(\sigma_R(s),m) \models \varphi$. Let $S_\varphi = \{t_1,\dots,t_k\}$ be the set of nodes in $\sigma_R(s)$ witnessing the existential part of φ. If S_φ is empty, that is, $k = 0$, then at least we know that there is a node $t_\forall \in \sigma_R(s)$ witnessing the universal part of φ. In this case, we let $\{(b_\forall,t_\forall)\}$ be the set of successors of (a,s)

in $\mathbb{T}'_f$, where $b_\forall$ is such that t is in $m(b_\forall)$. Otherwise, S_φ is non-empty and we define the successors of (a,s) in $\mathbb{T}'_f$ to be the elements of the following set.

$$\{(b,t) \mid t \in m(b) \cap S_\varphi\} \tag{5.4}$$

Observe that this is a finite subset of $\{(b,t) \mid t \in m(b)\}$. It follows that each successor of (a,s) in $\mathbb{T}'_f$ is also a successor of (a,s) in $\mathbb{T}_f$ and we can maintain our inductive hypothesis at height $i+1$.

It should be clear by construction that $\mathbb{T}'_f$ is a finitely branching subtree of $\mathbb{T}_f$. Let $\mathbb{T}'$ be the subtree of $\mathbb{T}$ obtained by projecting $\mathbb{T}'_f$ on $\mathbb{T}$, that is, its carrier T' is given as $\pi_2^f[T'_f]$. Since $\mathbb{T}'_f$ is finitely branching then also $\mathbb{T}'$ is finitely branching. The proof of the main statement is completed by showing the following claim.

CLAIM 13. The tree $\mathbb{T}'$ is accepted by $\mathbb{A}$.

PROOF OF CLAIM The key observation of the proof is that $\mathbb{T}'_f$ is the tree representation of a winning strategy f' for $\exists$ in $\mathcal{G}'$. The strategy f' is defined from each basic position $(a,s) \in A \times T'$ occurring in $\mathbb{T}'_f$ as a node. The marking suggested by $\exists$ from position (a,s) is uniquely defined by (5.4). The proof of claim is completed by showing that:

1. f' is a full and surviving strategy for $\exists$ in $\mathcal{G}'$;
2. f' is a winning strategy for $\exists$ in $\mathcal{G}'$.

In order to prove the first fact, recall that by assumption f is full and surviving strategy for $\exists$ in $\mathcal{G}$. By construction f' suggests the same markings of f on all basic positions in $\mathcal{G}'$, meaning that it is full in $\mathcal{G}'$. Given a position $(a,s) \in T_{f'}$, let $\sigma'_R(s)$ be the set of successors of s in $\mathbb{T}'$ and m' the marking suggested by f' from position (a,s). By definition of f' in terms of $\mathbb{T}'_f$, there is a disjunct $\varphi \in SBF^+(A)$ of $\Delta(a,\sigma_C(s))$ such that:

a) if φ only consists of an universal part, the A-structure $(\sigma'_R(s), m')$ has a unique element, witnessing the universal part of φ;

b) otherwise, φ has a non-empty existential part; in this case, by definition $(\sigma'_R(s), m')$ contains exactly the witnesses for the existential part of φ, meaning that it also verifies (vacuously) the universal part of φ.

In any of the two cases, m' makes φ true in $\sigma'_R(s)$, implying that it is a legitimate move for $\exists$. It follows that f' is a surviving strategy for $\exists$ in $\mathcal{G}'$.

In order to prove the second fact, let π be an infinite f'-conform match of $\mathcal{G}'$, with basic positions

$$B_\pi := (a_I, s_I), (a_1, s_1), \ldots, (a_n, s_n), \ldots.$$

By definition of f', the sequence B_π is just a branch of $\mathbb{T}'_f$. By construction, $\mathbb{T}'_f$ is a finitely branching subtree of $\mathbb{T}_f$ with the same root. Therefore B_π is also a branch of $\mathbb{T}_f$, meaning that there is an f-conform match of $\mathcal{G}$ with the same basic positions. By assumption the strategy f is winning for $\exists$ in $\mathcal{G}$. It follows that the minimum parity occurring along π is even. Therefore $\exists$ wins the match π. ■

The proof of the claim 5.2 completes the proof of the main statement.

□

The Finite Branching Property qualifies *MSO* as a logic which is not very expressive on the horizontal dimension of trees. This aspect is not revealed if we restrict to finitely branching trees. In the same way, the precise nature of *WMSO* expressiveness is disclosed only if we take trees with arbitrary branching degree into account. As we will show in the sequel, it turns out that *WMSO* does not have the Finite Branching Property, being able to distinguish between trees with finite and infinite branching degree. It follows that *WMSO* is essentially stronger than *MSO* in expressing properties of the horizontal dimension of trees.

Proposition 5.8. *The class of WMSO-definable tree languages and the class of MSO-definable tree languages are incomparable.*

Proof By proposition 5.5 there is a tree language which is *MSO*-definable but not *WMSO*-definable. For the converse direction, consider the language $\mathcal{L}_{Nfb}$ defined by putting

$$\mathbb{T} \in \mathcal{L}_{Nfb} \quad \textit{iff} \quad \text{the tree } \mathbb{T} \text{ is not finitely branching.} \tag{5.5}$$

Let χ_{Nfb} be a *WMSO*-formula defined by putting

$$\chi_{Nfb} \quad := \quad \exists x \neg \exists p \, \forall y \, (xRy \to y \in p). \tag{5.6}$$

Intuitively, χ_{Nfb} says that there is a node $\|x\|$ such that no finite set $\|p\|$ can contain all the successors of $\|x\|$. It is easy to see that the tree language $\mathcal{L}_{Nfb}$ is defined by χ_{Nfb}.

Suppose by way of contradiction that $\mathcal{L}_{Nfb}$ is *MSO*-definable. By proposition 2.26 there is an *MSO*-automaton $\mathbb{A}$ such that $L(\mathbb{A}) = \mathcal{L}_{Nfb}$. Since we are considering trees with arbitrary branching degree, the language $L(\mathbb{A})$ is not empty. Then, by proposition 5.7, there is a finitely branching tree $\mathbb{T}'$ in $L(\mathbb{A})$. This also means that $\mathbb{T}'$ is in $\mathcal{L}_{Nfb}$, contradicting the definition of $\mathcal{L}_{Nfb}$ as in (5.5). Therefore the language $\mathcal{L}_{Nfb}$ is not *MSO*-definable. □

Proposition 5.8 does not only state the incomparability of *MSO* and *WMSO*, but also reveals the nature of their relation. The two logics are in some sense *orthogonal*: *MSO* is weaker than *WMSO* on the horizontal dimension, while *WMSO* is weaker than *MSO* on the vertical dimension on trees.

In the sequel we bring further our investigation by comparing the expressive power of *WMSO* and *WFMSO*. As a preliminary observation, it is immediate that *WFMSO* is weaker than *WMSO* on the horizontal dimension, being a fragment of *MSO*. The orthogonal question, namely how *WFMSO* and *WMSO* relate on the vertical dimension, requires a finer analysis. What we have seen so far is that both *WMSO* and *WFMSO* are weaker than *MSO* on this respect. What we are going to show is that *WFMSO* is still stronger than *WMSO* on the vertical dimension, implying that the two logics have incomparable expressive power.

5.3 A Janin-Walukiewicz Theorem for *WFMSO*

Janin and Walukiewicz [16] have shown that the bisimulation-invariant fragment of *MSO* is as expressive as the modal μ-calculus (μMC) on trees. In this section we consider the same question for the bisimulation-invariant fragment of *WFMSO*. It turns out that *WFMSO* is still weaker than *MSO* on this respect, being as expressive as the alternation-free fragment of the modal μ-calculus (*AFMC*). This outcome is coherent with the perspective on *WFMSO* and weak *MSO*-automata that we have suggested throughout this thesis. Indeed, there is a tight connection between fixpoint operators of the μ-calculus and parities occurring infinitely often in parity games [34]. The absence of alternation in formulae from *AFMC* intuitively corresponds to at most one parity occurring infinitely often along infinite matches of a parity game, which is exactly the property of weak *MSO*-automata described in remark 3.3.

Theorem 5.9. *Let $\mathcal{L}$ a tree language that is closed under bisimulation. The following are equivalent.*

1. *The language $\mathcal{L}$ is AFMC-definable.*
2. *The language $\mathcal{L}$ is WFMSO-definable.*

For proving this result, once again we use an automata-theoretic argument. Roughly, the idea is that automata for *AFMC* are the weak counterpart of automata for μMC, just as automata for *WFMSO* are the weak version of *MSO*-automata. Then, the same argument, used by Janin and Walukiewicz [16] to show that automata for μMC and *MSO* have the same expressive power modulo bisimulation, can be restricted to show an analogous result for the weak counterparts.

For this purpose, the main observation is that automata for μMC and for *MSO* are the same class of automata, modulo a certain transformation on the first-order sentences associated with the transition function. Since every formula of μMC is bisimulation invariant, the first-order language associated with the transition of the corresponding automata should reflect this invariance at the level of the successors of a given node. This language cannot be $FO^+(A)$, because the equality symbol = allows for sentences which count the number of elements, such as

$$\varphi_{Two} \quad := \quad \exists x_1 \, \exists x_2 \, (x_1 \neq x_2 \wedge \forall z \, ((z \neq x_1 \wedge z \neq x_2) \to \bot)).$$

We can define an *MSO*-automaton corresponding to the property that each node in a tree has exactly two successors, by using φ_{Two} in the transition function. This is an example of a tree language which is *MSO*-definable but not bisimulation-invariant, meaning that it is not μMC-definable.

In the sequel we introduce a translation, which we call ∇-*translation*, transforming sentences in special basic form into sentences of $FO^+(A)$ without equality. The set of sentences which are ∇-translations of some sentence in special basic form will provide the first-order language associated with the automata for μMC.

Definition 5.10 (∇-translation)**.** Given a set A of unary predicates, let $\varphi \in SBF^+(A)$ be a sentence in special basic form of shape

$$\varphi \quad = \quad \exists x_1 \ldots \exists x_k \; (\mathit{diff}(\bar{x}) \wedge \bigwedge_{1 \le i \le k} a_i(x_i) \wedge \forall z \; (\mathit{diff}(\bar{x}, z) \to \bigvee_{1 \le l \le j} b_l(z))).$$

We define φ^∇ by putting

$$\varphi^\nabla \quad := \quad \exists x_1 \ldots \exists x_k \; (\bigwedge_{1 \le i \le k} a_i(x_i) \wedge \forall z \bigvee_{1 \le l \le j} b_l(z)).$$

We denote with $SBF^\nabla(A)$ the set $\{\varphi^\nabla \mid \varphi \in SBF^+(A)\}$. ◁

The modal μ-calculus is characterized by a class of automata which are defined as non-deterministic *MSO*-automata but for the transition function, which ranges over sentences from $SLatt(SBF^\nabla(A))$ instead of $SLatt(SBF^+(A))$ [15]. If we restrict to the alternation-free fragment, then a weaker version of these automata suffices [20]. We use the name *modal non-deterministic Büchi automata* to emphasize their connection with non-deterministic Büchi automata as we introduced with definition 4.1.

Definition 5.11 ([20])**.** A *modal non-deterministic Büchi automaton* on alphabet C is an *MSO*-automaton $\mathbb{B} = \langle B, b_I, \Delta, \Omega \rangle$ with Δ of type $B \times C \to SLatt(SBF^\nabla(B))$ and Ω of type $B \to \{0,1\}$. ◁

In [20] an automata characterization for *AFMC* in terms of modal non-deterministic Büchi automata is provided.

Proposition 5.12 ([20])**.** *Let $\mathcal{L}$ be a tree language. The following are equivalent.*

- *There exists $\varphi \in AFMC$ such that $\mathcal{L} = \|\varphi\|$.*
- *There are modal non-deterministic Büchi automata $\mathbb{B}^\nabla$ and $\mathbb{B}^\nabla_C$ such that $\mathcal{L} = L(\mathbb{B}^\nabla)$ and $\overline{\mathcal{L}} = \mathcal{L}(\mathbb{B}^\nabla_C)$.*

Proof Reference We refer to [20] for a proof of the statement. Direction $(1 \Rightarrow 2)$ follows from theorem 6, 4 and 5 in [20]. Direction $(2 \Rightarrow 1)$ follows from theorem 2 and 3 in [20]. □

We introduce a translation between *NDB* automata and modal *NDB* automata. Let $\mathbb{B}$ be a *NDB* automaton. In analogy with Janin and Walukiewicz's argument, we are going to show that, if $L(\mathbb{B})$ is closed under bisimulation, then the modal *NDB* automaton $\mathbb{B}^\nabla$ that we obtain from $\mathbb{B}$ through the translation is such that $\mathbb{B} \equiv \mathbb{B}^\nabla$. This is the content of proposition 5.14.

Definition 5.13. Let $\mathbb{B} = \langle B, b_I, \Delta, \Omega \rangle$ be an *NDB* automaton. We define an automaton $\mathbb{B}^\nabla = \langle B, b_I, \Delta^\nabla, \Omega \rangle$ by putting

$$\Delta^\nabla(a,c) \quad := \quad \bigvee \{\varphi^\nabla \mid \varphi \text{ is a disjunct of } \Delta(a,c)\}.$$

By definition 5.10 the transition function Δ^∇ has type $B \times C \to Slatt(SBF^\nabla(B))$, meaning that $\mathbb{B}^\nabla$ is a modal *NDB* automaton. ◁

Proposition 5.14. *Let $\mathbb{B}$ be an NDB automaton and $\mathbb{B}^\nabla$ the modal NDB automaton constructed from $\mathbb{B}$ as in definition 5.11. If $L(\mathbb{B})$ is closed under bisimulation, then $\mathbb{B} \equiv \mathbb{B}^\nabla$.*

Proof The argument is the same used in [15], proof of Lemma 12. Let $\mathbb{B} = \langle B, b_I, \Delta, \Omega \rangle$ and $\mathbb{B}^\nabla = \langle B, b_I, \Delta^\nabla, \Omega \rangle$ be given as in the statement and suppose that $L(\mathbb{B})$ is closed under bisimulation. Given a tree $\mathbb{T}$, we want to show that

$$\mathbb{B} \text{ accepts } \mathbb{T} \quad \textit{iff} \quad \mathbb{B}^\nabla \text{ accepts } \mathbb{T}. \tag{5.7}$$

Let $\mathbb{T}_\omega$ be the ω-expansion of $\mathbb{T}$, given as in definition 1.6. By remark 1.7, we know that $\mathbb{T}$ and $\mathbb{T}_\omega$ are bisimilar. This means that, in order to show (5.7), it suffices to show that

$$\mathbb{B} \text{ accepts } \mathbb{T}_\omega \quad \textit{iff} \quad \mathbb{B}^\nabla \text{ accepts } \mathbb{T}. \tag{5.8}$$

$(\Rightarrow)$ Suppose that $\mathbb{B}$ accepts $\mathbb{T}_\omega$ and let f be a winning strategy for $\exists$ in $\mathcal{G} = \mathcal{A}(\mathbb{B}, \mathbb{T}_\omega)@(a_I, (s_I, 0))$. We want to define a strategy f^∇ that is winning for $\exists$ in $\mathcal{G}^\nabla = \mathcal{A}(\mathbb{B}, \mathbb{T})@(a_I, s_I)$. As usual, f^∇ is provided for each stage of the construction of a match π^∇ of $\mathcal{G}^\nabla$, while playing in parallel an f-conform shadow match π of $\mathcal{G}$. For each round z that is played in π and π^∇, we want to maintain the following relation between the two matches.

For some $a \in A$, $s \in T$ and $i < \omega$, the current basic positions in π^∇ and π are respectively (a,s) and $(a,(s,i))$.	(‡)

Condition (‡) holds at the initial round, where we initialize π and π^∇ respectively at position $(a_I,(s_I,0))$ and (a_I,s_I). Inductively, suppose that we are given at round z_i with basic positions respectively $(a,(s,i))$ and (a,s) in π and π^∇, for some $i < \omega$. By assumption we are provided with a marking $m : A \to \wp(\sigma_R^\omega(s,i))$ that makes $\Delta(a,\sigma_C^\omega(s,i))$ true in $\sigma_R^\omega(s,i)$. We define a marking $m^\nabla : A \to \wp(\sigma_R(s))$ by putting

$$t \in m^\nabla(a) \quad \textit{iff} \quad (t,i) \in m(a)$$

for each $t \in T$, $i < \omega$ and $a \in A$. We let m^∇ be the choice of $\exists$ from position (a,s) in π^∇. If m^∇ is a legitimate move for $\exists$, then it is clear that condition (‡) can be maintained at round z_{i+1}: any next basic position $(b,t) \in A \times T$ picked by $\forall$ in π^∇ corresponds to an admissible move $(b,(t,i)) \in A \times (T \times \omega)$ for $\forall$ in π, for any $i < \omega$.

Thus it remains to show that m^∇ is a legitimate choice for $\exists$, that is,

$$(\sigma_R(s), m^\nabla) \quad \vDash \quad \Delta^\nabla(a, \sigma_C(s)). \tag{5.9}$$

For this purpose, suppose that $\varphi \in SBF^+(B)$ is a disjunct of $\Delta(a,\sigma_C^\omega(s,i))$, depending on sequences of states $a_1 \dots a_k, b_1 \dots b_j$ of B, such that $(\sigma_R^\omega(s,i),m) \vDash \psi$. By definition of Δ^∇, the sentence φ^∇ is a disjunct of $\Delta^\nabla(a,\sigma_C^\omega(s,i))$. By definition $\sigma_C(s) = \sigma_C^\omega(s,i)$, that is, the label of (s,i) in $\mathbb{T}_\omega$ is just the label of s in $\mathbb{T}$. It follows that φ^∇ is a disjunct of $\Delta^\nabla(a,\sigma_C(s))$. In order to show (5.9), it suffices to show that

$$(\sigma_R(s), m^\nabla) \quad \vDash \quad \varphi^\nabla.$$

By the syntactic shape of φ, we have k nodes $(t_1,n_1),\dots,(t_k,n_k)$ in $\sigma_R(s,i)$ witnessing the variables $x_1,\dots,x_k$ in the existential part of φ. This means that m assigns a_i to (t_i,n_i), for each i with $1 \le i \le k$. By definition m^∇ assigns a_i to t_i, for each i with $1 \le i \le k$. This means that $t_1,\dots,t_k$ are witnesses for variables $x_1,\dots,x_k$ occurring in the existential part of φ^∇. It remains to consider the condition given by the universal part of φ^∇, that is, to each node $t \in \sigma_R(s)$ is assigned some state $b \in \{b_1,\dots,b_j\}$ according to m^∇. For this purpose, take $t \in \sigma_R(s)$ and any natural number n different from $n_1,\dots,n_k$. By the syntactic shape of φ, the node (t,n) is a witness for the variable z in the universal part of φ. Therefore m assigns to (t,n) some state $b \in \{b_1,\dots,b_j\}$, implying that t is in $m^\nabla(b)$. Since t was an arbitrary node of $\sigma_R(s)$, this shows that indeed $(\sigma_R(s),m) \vDash \forall z \bigvee_{l \le j} b_l(z)$ and completes the proof of (5.9).

We have shown that condition (‡) can be maintained for each round that is played in the two matches. In order to show that $\exists$ wins π^∇, first suppose that π^∇ is a finite match. Since the suggestion of f^∇ is always legitimate, this means that $\forall$ gets stuck and $\exists$ wins. Otherwise, π and π^∇ are infinite and by condition (‡) the same states from B are visited along the play. Since $\mathbb{B}$ and $\mathbb{B}^\nabla$ have the same parity map, the minimum parity occurring infinitely often in π and π^∇ is the same. By the fact that f is a winning strategy for $\exists$ in $\mathcal{G}$ and π is f-conform, this suffices to show that $\exists$ wins π^∇.

($\Leftarrow$) Let f^∇ be a winning strategy for $\exists$ in $\mathcal{G}^\nabla = \mathcal{A}(\mathbb{B},\mathbb{T})@(a_I,s_I)$. We define a strategy f for each stage of the construction of some match π of $\mathcal{G}$, while playing in parallel an f-conform shadow match π^∇ of $\mathcal{G}^\nabla$. For each round z that is played in π and π^∇, we want to maintain the same condition (‡) that we used in showing the converse direction.

At the initial round, position $(a_I,(s_I,0))$ occurs in π and we initialize π^∇ at position (a_I,s_I). Inductively, suppose that we are given at round z_i with basic positions respectively $(a,(s,i))$ and (a,s) in π and π^∇, for some $i < \omega$. By assumption we are provided with a marking $m^\nabla : A \to \wp(\sigma_R^\omega(s))$ that makes $\Delta^\nabla(a,\sigma_C(s))$ true in $\sigma_R(s)$. We define a marking $m : A \to \wp(\sigma_R(s,i))$ by putting

$$(t,i) \in m(a) \quad \textit{iff} \quad t \in m^\nabla(a).$$

We let m be the choice of $\exists$ in π. In order to see that m is a legitimate move, let ψ be a disjunct of $\Delta^\nabla(a,\sigma_C(s))$ that is true in $(\sigma_R(s),m^\nabla)$. By definition of Δ^∇ we know that ψ is equal to φ^∇, for some disjunct φ of $\Delta(a,\sigma_C(s)) = \Delta(a,\sigma_C^\omega(s,i))$. In order to show that m is legitimate, it suffices to show that φ is true in $(\sigma_R^\omega(s,i),m)$. For this purpose, let $t_1,\dots,t_k$ be (not necessarily distinct) elements of $\sigma_R(s)$ witnessing variables $x_1,\dots,x_k$ of φ^∇, that is, for each i with $1 \le i \le k$, we have that $t_i \in m(a_i)$. By definition of m^∇, nodes $(t_1,1),\dots,(t_k,k)$ are k distinct elements of $\sigma_R^\omega(s,i)$ such that, for each i with $1 \le 1 \le k$, we have that $(t_i,i) \in m(a_i)$. It follows that they witness the existential part of φ. For the universal part of φ, consider any node $(t,n) \in \sigma_R^\omega(s,i)$ which is different from $(t_1,1),\dots,(t_k,k)$. By definition of φ^∇ the node $t \in \sigma_R(s)$ witnesses the universal part of φ^∇, that is, for some l with $1 \le l \le j$, we have that $t \in m^\nabla(b_l)$. Then, by definition of m, the node (t,n) is in $m(b_l)$ and witnesses the universal part of φ. Therefore φ is true in $(\sigma_R^\omega(s,i),m)$ and m is a legitimate move for $\exists$ in π.

Using the same argument of the converse direction, we can show that condition (‡) is maintained at round z_{i+1} and $\exists$ wins π.

□

Proposition 5.14 provides the key result to prove that *AFMC* is at least as expressive as the bisimulation-invariant fragment of *WFMSO*. This is one direction of theorem 5.9. For the converse direction, we need to show that the bisimulation-invariant fragment of *WFMSO* is not weaker than *AFMC*. This is in fact the easy direction of theorem 5.9, being a corollary of the automata-theoretic characterization of *AFMC* provided in [20].

Proposition 5.15 ([20]). *Let $\varphi \in AFMC$ be a sentence. There is a weak MSO-automaton $\mathbb{A}_\varphi$ such that φ is equivalent to $\mathbb{A}_\varphi$.*

Proof reference In [20], theorem 3, it is shown that a class of automata called *symmetric weak alternating automata* characterizes *AFMC*. It is easy to check that symmetric weak alternating automata are just a particular case of weak *MSO*-automata, where the transition function Δ ranges over sentences from $FO^+(A)$ without the equality symbol =. □

We are now ready to prove the main result of this section.

Proof of theorem 5.9 Let $\mathcal{L}$ be a tree language that is closed under bisimulation. The direction $(1 \Rightarrow 2)$ follows by proposition 5.15 and theorem 4.14. The proof of direction $(2 \Rightarrow 1)$ is given by the following derivation.

there is $\varphi_1 \in$ *WFMSO* such that $\|\varphi_1\| = \mathcal{L}$	$\Leftrightarrow$	(closure of *WFMSO* under negation)
there are $\varphi_1, \varphi_2 \in$ *WFMSO* such that $\|\varphi_1\| = \mathcal{L}$ and $\|\varphi_2\| = \overline{\mathcal{L}}$	$\Leftrightarrow$	(theorem 4.14)
there are *NDB* automata $\mathbb{B}_1, \mathbb{B}_2$ such that $L(\mathbb{B}_1) = \mathcal{L}$ and $L(\mathbb{B}_2) = \overline{\mathcal{L}}$	$\Rightarrow$	(proposition 5.14)
there are modal *NDB* automata $\mathbb{B}_1^\nabla, \mathbb{B}_2^\nabla$ such that $L(\mathbb{B}_1^\nabla) = \mathcal{L}$ and $L(\mathbb{B}_2^\nabla) = \overline{\mathcal{L}}$	$\Leftrightarrow$	(proposition 5.12)
there is $\varphi_1 \in$ *AFMC* such that $\|\varphi_1\| = \mathcal{L}$		

□

As a corollary of theorem 5.9, we obtain an incomparability result for *WFMSO* and *WMSO*. We can see this as a strengthening of proposition 5.8, *WFMSO* being weaker than *MSO*.

Corollary 5.16. *The class of WMSO-definable tree languages and the class of WFMSO-definable tree languages are incomparable.*

Proof By the same argument used for proposition 5.8, the language $\mathcal{L}_{Nfb}$ defined as in (5.5) is *WMSO*-definable but not *WFMSO*-definable. For the converse direction, consider the sentence $\chi_{Wf} \in AFMC$ defined by putting

$$\chi_{Wf} \quad = \quad \mu q.(\Box q \vee p).$$

Intuitively, χ_{Wf} holds in a tree $\mathbb{T}$ if and only if on each branch of $\mathbb{T}$ there is a node labeled with p. By proposition 5.9 the tree language $\|\chi_{Wf}\|$ is *WFMSO*-definable. In order to see that $\|\chi_{Wf}\|$ is not *WMSO*-definable, recall the prefix topology on trees, defined as in section 1.6. By proposition 1.20 the tree language $\|\chi_{Wf}\|$ is not a Borel set of this topology. However, by proposition 1.19, only the Borel tree languages are *WMSO*-definable. This means that $\|\chi_{Wf}\|$ is not *WMSO*-definable. □

Historical Notes

To the best of our knowledge, the result that weak *MSO*-automata characterize *WMSO* on finitely branching trees is folklore, just as the incomparability result of proposition 5.8. The former statement is mentioned in [14], section 4.3, whereas the latter statement is mentioned in [9], remark 1.8.

Modal correspondence theory is the comparative study of expressiveness of modal languages and classical languages (such as first-order logic) on transition systems. Perhaps the most important contribution in this area is van Benthem's Bisimulation Theorem stating that modal logic is the bisimulation-invariant fragment of first-order logic [30]. Another landmark result is Janin and Walukiewicz's theorem stating that the modal μ-calculus is the bisimulation-invariant fragment of *MSO* [16].

We refer to Kupferman and Vardi [18], [20] for the automata-theoretic characterization of *AFMC* on trees of arbitrary branching degree. For the case of binary trees, an algebraic approach has been proposed [1] [2] to relate the expressive power of *AFMC*, *WMSO* and weak automata. For an overview on automata for modal fixpoint logics we refer to [34] and [31].

Conclusions

We recapitulate what we consider the main contributions of this thesis. In the second and third chapter we rephrase the existing results on *MSO*-automata and weak *MSO*-automata in a unified framework. We emphasize the role of the simulation theorem and the two-sorted construction in providing a guidance to compare the expressive power of *MSO*-automata, weak *MSO*-automata and the corresponding logics. For this purpose, we introduce and reformulate many concepts that were implicit or differently expressed in the theory of tree automata, such as the tree representation of a strategy, the notions of minimal, full and functional strategy, the completion construction for *MSO*-automata, the two-sorted construction and the Büchi powerset construction for weak *MSO*-automata. An analogous contribution is also provided in the fourth chapter, where we reformulate Rabin's constructions for non-deterministic Büchi automata in a game-theoretic fashion, generalizing the associated results from the binary case to the arbitrarily branching case. In Appendix *A* we also present a game-theoretic perspective on the complement construction for *MSO*-automata. Our proof of the complementation lemma is based on the observation that acceptance games for *MSO*-automata are essentially asymmetric, assigning a prominent role to player $\exists$. We show that acceptance games can be equivalently defined in a completely symmetric way, and also asymmetric on the side of player $\forall$. We believe that this perspective on acceptance games brings useful insights on the concept of *alternation*, which is central in the theory of tree automata [24] [11].

On the connection between automata and logic, our main contribution is the logical characterization of weak *MSO*-automata on trees of arbitrary branching degree. For this purpose we introduce a new variant of *MSO* which we call well-founded monadic second-order logic (*WFMSO*). In the third chapter we prove that, for each formula $\varphi \in WFMSO$, there is a weak *MSO*-automaton which is equivalent on trees of arbitrary branching degree. In the fourth chapter we show the converse direction, namely that each tree language which is recognized by some weak *MSO*-automaton is also defined by some *WFMSO*-formula. The proof passes through non-deterministic Büchi automata, that generalize Rabin's 'special automata' [28] working on binary trees. We give a second characterization for *WFMSO* in connection with this class of automata: a tree language $\mathcal{L}$ is *WFMSO*-definable if and only if both $\mathcal{L}$ and its complement are recognized by non-deterministic Büchi automata. This can be seen as a generalization of an analogous result of Rabin for *WMSO* on binary trees [28].

We believe that the fifth chapter provides a quite broad overview of the expressivity results connecting *MSO*, *WMSO* and *WFMSO* which are disseminated in the literature. Some (apparently) folklore results, such as the Finite Branching Property and the incomparability between *MSO* and *WMSO*, are disclosed and their consequences investigated. In particular, we emphasize the difference between properties on the *horizontal* and *vertical* dimension of trees. We believe that this perspective provides a guidance in understanding and comparing the expressing power of different monadic second-order logics.

Another main contribution is the modal characterization of the bisimulation-invariant fragment of *WFMSO*, which is proven to be as expressive as the alternation-free fragment of the modal μ-calculus. This result somehow completes the net of correspondences between *WFMSO* and *MSO*, the bisimulation-invariant fragment of *MSO* being as expressive as the modal μ-calculus [16]. As a corollary, we obtain the result that *WFMSO* and *WMSO* have incomparable expressive power on trees. This can be seen as a strengthening of the incomparability result for *MSO* and *WMSO*, *WFMSO* being a fragment of *MSO*.

Logic	Corresponding automata	Bisimulation-invariant fragment
MSO	*MSO*-automata	μMC
WFMSO	weak *MSO*-automata	*AFMC*
WMSO	?	?

Table 5.1: overview of the characterization results on arbitrarily branching trees.

Future Work

The main theme of our work has been the understanding of the expressive power of weak *MSO*-automata on trees of arbitrary branching degree. This was motivated by the observation that weak *MSO*-automata do not characterize *WMSO* on this class of structures. Thus a natural continuation would be to provide a different class of automata, which characterizes *WMSO* on trees of arbitrary branching degree. The crux of the matter is to understand how to define these automata, in such a way that their expressive power is incomparable with respect to *MSO*-automata. They should be able to express those specifications concerning the *horizontal dimension* of trees, which are *WMSO*-definable but not *MSO*-definable. This also means to avoid the finite branching property: then a problem arises, for all the projection constructions that we considered so far are tightly connected to such property. Therefore, in order to give a projection construction corresponding to *WMSO*-quantification, essentially different methods should be proposed.

A second line of research concerns the bisimulation-invariant fragment of *WMSO*. This investigation is motivated by the fact that all *WMSO*-definable tree languages are Borel [9]. If the bisimulation-invariant fragment of *WMSO* is strictly weaker than the modal μ-calculus, then it would correspond to a sort of 'Borelian' fragment, providing a better understanding of the topological complexity of modal fixpoint logics. In fact there are reasons to believe that this is the case. To the best of our knowledge, all examples of tree languages that are *WMSO*-definable but not *MSO*-definable are not bisimulation closed. As an example, consider the language of finitely branching trees: each finitely branching tree is bisimilar to its ω-expansion, which is not finitely branching. This motivates the conjecture that the bisimulation-fragment of *WMSO* 'collapses inside' the bisimulation-invariant fragment of *MSO*, that is the modal μ-calculus. In fact, we hypothesize that it is even included in the alternation-free fragment of the modal μ-calculus, for the intuitive reason that *WMSO* is not stronger than *WFMSO* in expressing properties on the vertical dimension of trees. Analogously to other characterization results, we believe that this question should be tackled from an automata-theoretic perspective. The aforementioned automata characterization of *WMSO* would probably be a decisive step towards the proof of the conjecture.

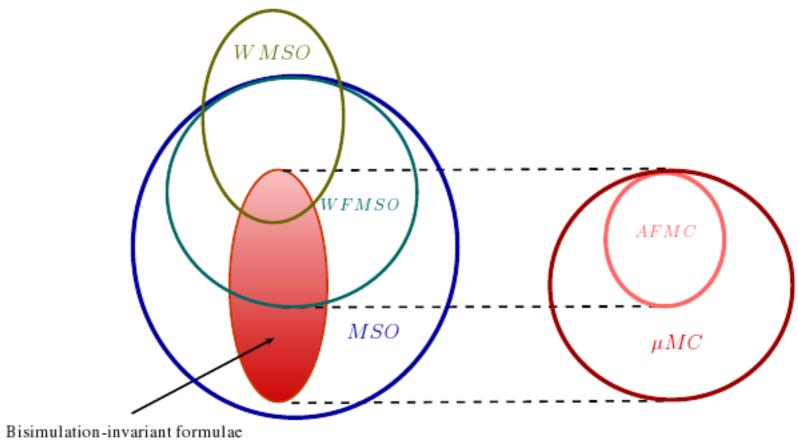

Figure 5.1: overview of the relations between classes of definable tree languages (the position of *WMSO*$\cap$*MSO*-definable tree languages is conjectural).

As a conclusive remark, we observe that all results that we obtained concern leafless trees. We believe that it is not too harmful to extend the same results taking also trees with leaves into account. The idea is that a C-labeled tree $\mathbb{T}$ with leaves can be represented as a leafless $\wp(P\cup\{r\})$-labeled tree $\mathbb{T}'$, where all the nodes of the original tree are labeled in $\mathbb{T}'$ with a propositional letter $r \notin P$. Then each formula φ of the monadic second-order language on P can be translated into a formula φ' on $P\cup\{r\}$, such that $\|\varphi\|_P$ and $\|\varphi'\|_{P\cup\{r\}}$ are the same tree language modulo the translation of trees with leaves into leafless trees.

Acknowledgements

My sincere thanks go to my supervisors Alessandro Facchini and Yde Venema, both providing constant support and helpful suggestions throughout this thesis project. I want to thank Alessandro for introducing me to the intriguing problem of the automata characterization of *WMSO*, and his guidance (even on Skype) and great patience in examining my several attempts to tackle it. I am also grateful to Yde: his perspective on *MSO*-automata was essential for my understanding of the topic, and his sharp comments stimulated me to improve the clarity of my ideas and the way to present them.

Appendix A

Symmetric and ∀-Asymmetric Acceptance Games

In this appendix we want to show a complementation lemma for *MSO*-automata.

Proposition A.1. *Given an MSO-automaton $\mathbb{A}$, there is an effectively constructible MSO-automaton $\overline{\mathbb{A}}$ such that*

$$L(\overline{\mathbb{A}}) \quad = \quad \overline{L(\mathbb{A})}.$$

The idea is to use a game-theoretical argument. Given an *MSO*-automaton $\mathbb{A}$, we want to define an automaton $\overline{\mathbb{A}}$, with the same carrier A of $\mathbb{A}$, such that for each tree $\mathbb{T}$, basic position $(a,s) \in A \times T$, the following holds:

$$(a,s) \text{ is a winning position for } \exists \text{ in } \mathcal{A}(\mathbb{A},\mathbb{T}) \quad \textit{iff} \quad (a,s) \text{ is a winning position for } \forall \text{ in } \mathcal{A}(\overline{\mathbb{A}},\mathbb{T}). \quad \text{(A.1)}$$

The main problem with this argument is to prove that a basic position is winning for ∀. We should define a strategy for ∀ which can face any legitimate marking that ∃ could pick at each round. This is trickier than dealing with strategies for ∃, because ∃'s choice of a marking always comes first in a round, and her move is a more 'complicated' object to handle than ∀'s move. In fact the acceptance game has an *asymmetric* definition, assigning a prominent role to player ∃.

The key idea is to make our task easier by creating another asymmetric version of the acceptance game, which privileges ∀. At each round we want ∀ to pick first a marking, and ∃ answer by picking a basic position. If we can prove that such version of the acceptance game is in fact equivalent to the standard one, then it becomes relatively easy to define an automaton $\overline{\mathbb{A}}$ accepting the complement language of $\mathbb{A}$. Indeed a winning strategy f for ∃ in $\mathcal{G}$ becomes a legal strategy for ∀ in the ∀-asymmetric version of the acceptance game. If we can enforce condition (A.1), then f is also winning for ∀ and makes $\overline{\mathbb{A}}$ reject $\mathbb{T}$.

A.1 Equivalence between Acceptance Games

In this section we introduce ∀-asymmetric acceptance games and prove that they are equivalent to the standard acceptance games (table 2.1). As an intermediate step, we introduce a third version of the acceptance game, which we call *symmetric* because similar roles are assigned to the two players.

Definition A.2. Let $\mathbb{A}$ be an *MSO*-automaton and $\mathbb{T}$ a tree. We let n_A denote the natural number $Max(\Omega[A])$.

- The ∀*-asymmetric acceptance game* of $\mathbb{A}$ on $\mathbb{T}$ - notation $\mathcal{A}_\forall(\mathbb{A},\mathbb{T})$ - is defined according to the rules of table A.1.

Position	Player	Admissible moves	Parity
$(a,s) \in A \times S$	∀	$\{m : A \to \wp(\sigma_R(s)) \mid (\sigma_R(s),m) \nvDash \Delta(a,\sigma_v(s))\}$	$\Omega(a)$
$m : A \to \wp(\sigma_R(s))$	∃	$\{(b,t) \mid t \in (\sigma_R(s) \setminus m(b))\}$	n_A

Table A.1: ∀-asymmetric acceptance game for *MSO* automata

- For each node t of $\mathbb{T}$, let c_t be a constant which is interpreted on t in $\mathbb{T}$. We denote with $FO^+_T(A)$ the language of sentences from $FO^+(A)$ where constants from $\{c_t \mid t \in T\}$ can occur in place of individual variables. Given a formula φ, we denote with $\varphi[c_t/x]$ the formula obtained by substituting each free occurrence of x in φ with c_t. The *symmetric acceptance game* of $\mathbb{A}$ on $\mathbb{T}$ - notation $\mathcal{A}_{sym}(\mathbb{A},\mathbb{T})$ - is defined according to the rules of table A.2.

Position	Player	Admissible moves	Parity
$(a,s) \in A \times S$	-	$\{(\Delta(a,\sigma_C(s)),s)\}$	$\Omega(a)$
$(\psi_1 \vee \psi_2,s)$	$\exists$	$\{(\psi_1,s),(\psi_2,s)\}$	n_A
$(\psi_1 \wedge \psi_2,s)$	$\forall$	$\{(\psi_1,s),(\psi_2,s)\}$	n_A
$(\exists x.\varphi,s)$	$\exists$	$\{(\varphi[c_t/x],s) \mid t \in \sigma_R(s)\}$	n_A
$(\forall x.\varphi,s)$	$\forall$	$\{(\varphi[c_t/x],s) \mid t \in \sigma_R(s)\}$	n_A
$(c_{t_1} \approx c_{t_2},s)$ and $t_1 \neq t_2$	$\exists$	$\varnothing$	n_A
$(c_{t_1} \approx c_{t_2},s)$ and $t_1 = t_2$	$\forall$	$\varnothing$	n_A
$(c_{t_1} \not\approx c_{t_2},s)$ and $t_1 = t_2$	$\exists$	$\varnothing$	n_A
$(c_{t_1} \not\approx c_{t_2},s)$ and $t_1 \neq t_2$	$\forall$	$\varnothing$	n_A
$(\bot,s)$	$\exists$	$\varnothing$	n_A
$(\top,s)$	$\forall$	$\varnothing$	n_A
$(a(c_t),s)$	-	$\{(a,t)\}$	n_A

Table A.2: Symmetric acceptance game for *MSO* automata

The *basic positions* of $\mathcal{A}_\forall(\mathbb{A},\mathbb{T})$ and $\mathcal{A}_{sym}(\mathbb{A},\mathbb{T})$ are the same of $\mathcal{A}(\mathbb{A},\mathbb{T})$. We call a position $(\varphi,s) \in (FO^+_T(A)) \times T$ *atomic* if φ is of the form $c_{t_1} \approx c_{t_2}$, $c_{t_1} \not\approx c_{t_2}$ or $a(c_t)$ for some $a \in A$ and $t \in \sigma_R(s)$. Winning conditions for $\forall$ and $\exists$ in $\mathcal{A}_\forall(\mathbb{A},\mathbb{T})$ and $\mathcal{A}_{sym}(\mathbb{A},\mathbb{T})$ are standardly defined in terms of the parities occurring along the play. $\lhd$

The idea of $\forall$-asymmetric acceptance games is that at each round $\forall$ assigns states $b \in A$ to nodes $t \in \sigma_R(s)$ in such a way that continuing the match from a position (b,t) leads to a rejecting run of $\mathbb{A}$ on $\mathbb{T}$. For this reason, $\exists$ will pick the next basic position among the ones that *are not* induced by $\forall$'s marking. Dually with respect to the case of standard acceptance games, $\forall$ will try to assign *as many states as possible* to each node, so that less basic positions are available choices for $\exists$.

The idea of symmetric acceptance games is that at each round $\exists$ and $\forall$ play a little sub-game, following the syntactic shape of the sentence $\Delta(a,\sigma_C(s)) \in FO^+(A)$ associated with the transition function. In fact this sub-game closely resembles the *evaluation game* providing the standard game semantics to first-order logic [12]. Observe that the cases for atomic positions in table A.2 are exhaustive, because $\Delta(a,\sigma_C(s))$ is a sentence. This means that all first-order variables appearing in $\Delta(a,\sigma_C(s))$ are bound and the two players replace all of them with individual constants before the match arrives to an atomic position.

Proposition A.3. *Let $\mathbb{A}$ be an MSO-automaton and $\mathbb{T}$ a tree. The following are equivalent.*

1. *Player $\exists$ has a winning strategy in $\mathcal{G} = \mathcal{A}(\mathbb{A},\mathbb{T})@(a_I,s_I)$.*

2. *Player $\exists$ has a winning strategy in $\mathcal{G}_{sym} = \mathcal{A}_{sym}(\mathbb{A},\mathbb{T})@(a_I,s_I)$.*

3. *Player $\exists$ has a winning strategy in $\mathcal{G}_\forall = \mathcal{A}_\forall(\mathbb{A},\mathbb{T})@(a_I,s_I)$.*

Proof $(1 \Rightarrow 2)$ Suppose that $\exists$ has a winning strategy f in $\mathcal{G}$. We define a strategy f' for $\exists$ in $\mathcal{G}_{sym}$, by induction on the construction of a match π' of $\mathcal{G}_{sym}$, while maintaining an f-conform shadow match π of $\mathcal{G}$. For each round that is played in π' and π, we want that either $\forall$ gets stuck in π' or the same basic position in π' and π can be maintained for the next round. This is the case for the initial round, where we initialize both matches at position (a_I,s_I). Inductively, suppose that we are at round z_i with the same basic position $(a,s) \in A \times T$ occurring both in π' and π. Since π is f-conform, the strategy f suggests a marking $m : A \to \wp(\sigma_R(s))$ that makes $\Delta(a,\sigma_C(s))$ true. We want to define how $\exists$ should play in π' at round z_i, in such a way that she can maintain the following condition for each position $(\varphi,s) \in FO^+_T(A)$ encountered in round z_i.

The marking m makes φ true in $\sigma_R(s)$.	(†)

Condition (†) is true for position $(\Delta(a,\sigma_C(s)),s)$ because m is suggested by the winning strategy f. Intuitively, this was the initial position of the sub-game of round z_i. For each non-atomic position $(\varphi,s) \in FO^+_T(A) \times T$ that is encountered while playing the sub-game, we distinguish the following cases.

- If $\varphi = \psi_1 \wedge \psi_2$, then position $(\psi_1 \wedge \psi_2,s)$ belongs to $\forall$. Let $i \in \{1,2\}$ be such that (ψ_i,s) is the next position picked by $\forall$. By inductive hypothesis $\psi_1 \wedge \psi_2$ is a sentence in $FO^+_T(A)$ such that m makes $\psi_1 \wedge \psi_2$ true in $\sigma_R(s)$. It follows that m makes ψ_i true in $\sigma_R(s)$.

- If $\varphi = \psi_1 \vee \psi_2$, then position $(\psi_1 \vee \psi_2, s)$ belongs to $\exists$. By inductive hypothesis $\psi_1 \vee \psi_2$ is a sentence in $FO^+_T(A)$ such that m makes $\psi_1 \vee \psi_2$ true in $\sigma_R(s)$. It follows that, for some $i \in \{1,2\}$, the marking m makes ψ_i true in $\sigma_R(s)$. We let $\exists$ pick (ψ_i, s) as next position in round z.

- If $\varphi = \exists x.\psi$, then position $(\exists x.\psi, s)$ belongs to $\exists$. By inductive hypothesis $\exists x.\psi$ is a sentence in $FO^+_T(A)$ such that m makes $\exists x.\psi$ true in $\sigma_R(s)$. It follows that, for some node $t \in \sigma_R(s)$, the marking m makes $\psi[c_t/x]$ true in $\sigma_R(s)$. We let $\exists$ pick $(\psi[c_t/x], s)$ as next position in round z.

- If $\varphi = \forall x.\psi$, then position $(\forall x.\psi, s)$ belongs to $\forall$. Let $t \in \sigma_R(s)$ be a node such that $(\psi[c_t/x], s)$ is the next position picked by $\forall$. By inductive hypothesis $\forall x.\psi$ is a sentence in $FO^+_T(A)$ such that m makes $\forall x.\psi$ true in $\sigma_R(s)$. It follows that the marking m makes $\psi[c_t/x]$ true in $\sigma_R(s)$.

It is immediate to check that the above strategy allows $\exists$ to maintain condition (†) through round z_i in π'. This means that we arrive at some atomic position of the sub-game of round z_i, which is associated with an atomic subformula of $(\Delta(a, \sigma_C(s)), s)$. If it is of the form $(c_{t_1} \approx c_{t_2}, s)$ or $(c_{t_1} \not\approx c_{t_2}, s)$, then by condition (†) player $\forall$ gets stuck. Otherwise, the sub-game arrives to a position of the form $(b(c_t), s)$. By definition of $\mathcal{G}_{sym}$, $(b,t) \in A \times T$ is the next basic position in π', associated with round z_{i+1}. By condition (†), the marking m makes $b(c_t)$ true in $\sigma_R(s)$, meaning that t is in $m(b)$. Therefore (b,t) is an admissible choice for $\forall$ in π and we let it be the next basic position associated with round z_{i+1} in π.

In order to see that $\exists$ wins π', observe that the move suggested by f' is always legitime. Then either $\forall$ gets stuck at some round, or π' and π are both infinite matches, with the same parities occurring along the play. Since π is f-conform and f is winning for $\exists$ in $\mathcal{G}$, this suffices to show that the minimum parity occurring infinitely often in π' is even and then $\exists$ wins π'.

$(2 \Rightarrow 1)$ Suppose that $\exists$ has a winning strategy f' in $\mathcal{G}_{sym}$. We define a strategy f for $\exists$ in $\mathcal{G}$, by induction on the construction of a match π of $\mathcal{G}$, while maintaining an f'-conform shadow match π' of $\mathcal{G}_{sym}$. We want f to be such that, for each round that is played in π and π', either player $\forall$ gets stuck in π, or it is always possible to maintain the same basic position in π and π' for the next round.

For the base case, we initialize both π and π' at position (a_I, s_I). Inductively, suppose that we are at round z_i with the same basic position $(a,s) \in A \times T$ occurring in π' and π. We want to suggest a marking $m : A \to \wp(\sigma_R(s))$ to $\exists$ in π, on the base of how she plays in the sub-game associated with round z_{i+1} in π'. The key observation is that this sub-game resembles an *evaluation game* on the sentence $\Delta(a, \sigma_C(s))$, for which we consider the possible outcomes. By playing according to f', player $\exists$ is guaranteed to not get stuck in the sub-game. Depending on how $\forall$ plays, an atomic position of the form $(c_{t_1} \approx c_{t_2}, s)$, $(c_{t_1} \not\approx c_{t_2}, s)$ or $(b(c_t), s)$ is reached. In any of the former two cases, by definition of $\mathcal{G}_{sym}$ one of the two players gets stuck. Since this player cannot be $\exists$, then it is $\forall$, meaning that the relation between t_1 and t_2 is the one depicted respectively by $c_{t_1} \approx c_{t_2}$ and $c_{t_1} \not\approx c_{t_2}$. Intuitively, the truth of $c_{t_1} \approx c_{t_2}$ and $c_{t_1} \not\approx c_{t_2}$ does not depend from which unary predicates mark t_1 and t_2. This means that, for the purpose of giving our marking m, we just need to focus on the remaining case, namely the (f'-conform) plays of the sub-game in which an atomic position of the form $(b(c_t), s)$ is reached. This motivates the following definition of m.

- ⋆ For each $t \in \sigma_R(s)$, let $A_{t,f'} \subseteq A$ be defined by putting

$$b \in A_{t,f'} \quad \textit{iff} \quad (b,t) \text{ occurs in some } f'\text{-conform match } \pi'_t \text{ extending } \pi'. \tag{A.2}$$

- ⋆ For each $b \in A$, we define $m : A \to \wp(\sigma_R(s))$ by putting

$$m(b) \quad := \quad \{t \in \sigma_R(s) \mid b \in A_{t,f'}\}.$$

Following the intuition given above, it can be easily verified that m makes $\Delta(a, \sigma_C(s))$ true in $\sigma_R(s)$, meaning that it is a a legitimate move for $\exists$ in π. If $m(b) = \varnothing$ for all $b \in A$, then $\forall$ gets stuck and $\exists$ wins the match π. Otherwise, suppose that $\forall$ picks $(b,t) \in A \times T$ as next basic position in π. Since t is in $m(b)$, then by definition of m there is an f-conform match π'_t extending π' where the position (b,t) occurs. We let $\exists$ and $\forall$ play in the sub-game of round z_i according to the prescription of π'_t. In this way, we make the shadow match π' coincide with π'_t for the first $i+1$ rounds. By assumption, π'_t is f'-conform and position (b,t) occurs at round z_{i+1}.

It follows that, for each round that is played in π, either $\forall$ gets stuck or we can maintain the same basic positions in π and π' in the next round. This suffices to show that $\exists$ wins π.

$(2 \Rightarrow 3)$ By contraposition, we want to show that if there is no winning strategy for $\exists$ in $\mathcal{G}_\forall$, then there is no winning strategy for $\exists$ also in $\mathcal{G}_{sym}$. Since parity games are determined (theorem 1.29), it suffices to show that if $\forall$ has a winning strategy f in $\mathcal{G}_\forall$, then he has a winning strategy f' in $\mathcal{G}_{sym}$. The argument showing this statement works dually with respect to the one given for direction $(1 \Rightarrow 2)$ and we confine ourselves to a sketch. We define the strategy f' for $\forall$ by induction on the construction of a match π' of $\mathcal{G}_{sym}$, while maintaining an f-conform shadow match π of $\mathcal{G}_\forall$. For each round that is played in π and π', we want that either $\exists$ gets stuck in π' or the same basic position can be maintained in π and π' for the next round. This suffices to show that $\forall$ wins π'.

For this purpose, we standardly initialize π' and π at the same initial position (a_I, s_I). Inductively, let z be a round where the same basic position $(a,s) \in A \times T$ occurs both in π and π'. Since π is f-conform, a marking m is suggested to player $\forall$ which makes $\Delta(a, \sigma_C(s))$ false in $\sigma_R(s)$. Analogously to direction $(1 \Rightarrow 2)$, we can tell $\forall$ how to play in π' the sub-game associated with $\Delta(a, \sigma_C(s))$, in such a way that the following condition is maintained for each position $(\varphi, s) \in FO^+_T(A)$ occurring at round z.

The marking m makes φ false in $\sigma_R(s)$.	(†)

If the sub-game of round z_i in π' reaches an atomic position $(c_{t_1} \approx c_{t_2}, s)$ or $(c_{t_1} \not\approx c_{t_2}, s)$, by condition (†) we have respectively that $t_1 \neq t_2$ and $t_1 = t_2$, implying that player $\exists$ gets stuck in π'. In the remaining case, the sub-game of round z in π' reaches a position of the form $(b(c_t), s)$. By condition (†), the sentence $b(c_t)$ is false in $(\sigma_R(s), m)$, meaning that t is not in $m(b)$. It follows that (b,t) is an admissible choice for player $\exists$ in π and we can maintain the same basic position in π and π' for the next round.

$(3 \Rightarrow 2)$ As we observed in the proof of direction $(2 \Rightarrow 3)$, by contraposition and determinacy of parity games, it suffices to show that, if $\forall$ has a winning strategy f' in $\mathcal{G}_{sym}$, then he has a winning strategy f in $\mathcal{G}_\forall$. The argument works dually with respect to the one provided for direction $(2 \Rightarrow 1)$ and we confine ourselves to a sketch. The strategy f is defined by induction on the construction of a match π of $\mathcal{G}_\forall$, while maintaining an f'-conform shadow match π' of $\mathcal{G}_{sym}$. We want the strategy f be such that either the same basic positions can be maintained in the two matches or $\exists$ gets stuck at some round in π.

For this purpose, we standardly initialize π' and π at the same initial position (a_I, s_I). Inductively, let z be a round where the same basic position $(a,s) \in A \times T$ occurs both in π and π'. We suggest a marking $m : A \to \wp(\sigma_R(s))$ to $\forall$ in π as follows.

- ⋆ For each $t \in \sigma_R(s)$, let $A_{t,f'} \subseteq A$ be defined as in (A.2). For each $b \in A$, we define $m : A \to \wp(\sigma_R(s))$ by putting

$$m(b) \quad := \quad \{t \in \sigma_R(s) \mid b \notin A_{t,f'}\}.$$

Dually with respect to direction $(2 \Rightarrow 1)$, in order to show that m makes $\Delta(a, \sigma_C(s))$ false in $\sigma_R(s)$, the key observation is that the sub-game at round z in π' resembles an *evaluation game* on $\Delta(a, \sigma_C(s))$, where we know by assumption that player $\forall$ does not get stuck.

If $m(b) = \sigma_R(s)$ for all $b \in A$, then $\exists$ gets stuck in π at round z_i and $\forall$ immediately wins. Otherwise, suppose that player $\exists$ chooses a next basic position (b,t) in π. Then t is not in $m(b)$ by definition of $\mathcal{G}_\forall$ and b is in $A_{t,f'}$ by definition of m. As for direction $(2 \Rightarrow 1)$, this suffices to show that we can make $\exists$ and $\forall$ play the sub-game of round z_i in π', in such a way that $\exists$'s movements follow the suggestions of f' and the basic position (b,t) occurs in π' at round z_{i+1}. □

A.2 A Complementation Lemma for *MSO*-Automata

In this section we use the equivalence between standard and $\forall$-symmetric acceptance games to show that the tree languages recognized by *MSO*-automata are closed under complementation.

Definition A.4. Let $\varphi \in For^+(A)$ be a formula. The *dual* $\varphi^\delta \in FO^+(A)$ of φ is defined by induction as follows.

$$\begin{array}{lcl}
(a(x))^\delta & := & a(x) \\
(\top)^\delta & := & \bot \\
(\bot)^\delta & := & \top \\
(x \approx y)^\delta & := & x \not\approx y \\
(x \not\approx y)^\delta & := & x \approx y \\
(\varphi \wedge \psi)^\delta & := & (\varphi)^\delta \vee (\psi)^\delta \\
(\varphi \vee \psi)^\delta & := & (\varphi)^\delta \wedge (\psi)^\delta \\
\forall x.\psi & := & \exists x.(\psi)^\delta \\
\exists x.\psi & := & \forall x.(\psi)^\delta
\end{array}$$

Let X be a set and $m : A \to \wp(X)$ a marking. The *dual* $m^\delta : A \to \wp(X)$ of m is defined by putting

$$m^\delta(a) \quad := \quad X \setminus m(a)$$

for each $a \in A$. ◁

The following property of the dual transformation is immediate by definition A.4.

Proposition A.5. *Let $\varphi \in FO^+(A)$ be a sentence, X a set and $m : A \to \wp(X)$ a marking. The following are equivalent.*

1. $(X, m) \models \varphi$.
2. $(X, m^\delta) \not\models \varphi^\delta$.

We have now all the ingredients to prove the complementation lemma for *MSO*-automata.

Proof of proposition A.1 Let $\mathbb{A} = \langle A, a_I, \Delta, \Omega \rangle$ be an *MSO*-automaton. Define an *MSO*-automaton $\overline{\mathbb{A}} = \langle A, a_I, \Delta^\delta, \Omega^\delta \rangle$ by putting $\Omega^\delta(a) := \Omega(a) + 1$ and $\Delta^\delta(a, c) := (\Delta(a, c))^\delta$ for each $a \in A$ and $c \in C$. Let $\mathbb{T}$ be a tree. We want to show that

$$\mathbb{A} \text{ accepts } \mathbb{T} \quad \textit{iff} \quad \overline{\mathbb{A}} \text{ does not accept } \mathbb{T}.$$

By proposition A.3 and determinacy of parity games, it suffices to prove that $\exists$ has a winning strategy in $\mathcal{G}_\exists = \mathcal{A}(\mathbb{A}, \mathbb{T})$ if and only if $\forall$ has a winning strategy in $\mathcal{G}_\forall = \mathcal{A}_\forall(\overline{\mathbb{A}}, \mathbb{T})$.

($\Rightarrow$) Suppose that $\exists$ has a winning strategy $f_\exists$ in $\mathcal{G}_\exists$. We define a strategy $f_\forall$ for $\forall$ in $\mathcal{G}_\forall$, by induction on the construction of a match $\pi_\forall$ of $\mathcal{G}_\forall$, while maintaining an $f_\exists$-conform shadow match $\pi_\exists$ of $\mathcal{G}_\exists$. For each round z that is played in $\pi_\forall$ and $\pi_\exists$, we want maintain the same basic position in the two matches.

At the initial round we initialize the two matches from position (a_I, s_I). Inductively, suppose that we are at round z_i and the same basic position $(a, s) \in A \times T$ occurs in $\pi_\forall$ and $\pi_\exists$. Let $m_\exists : A \to \wp(\sigma_R(s))$ be the suggestion of $f^\exists$ from position (a, s) in $\pi_\exists$. We let $m_\forall := (m_\exists)^\delta$ be the suggestion of $f^\forall$ from position (a, s) in $\pi_\forall$. By proposition A.5, the marking $m_\forall$ makes $(\Delta(a, \sigma_C(s)))^\delta = \Delta^\delta(a, \sigma_C(s))$ false in $\sigma_R(s)$, meaning that it is a legitimate move for $\forall$ in $\pi_\forall$. At this point we distinguish two cases.

1. If $m_\exists$ made $\forall$ get stuck in $\pi_\exists$, it means that $m_\exists(b) = \varnothing$ for all $b \in A$. Then, by definition, $m_\forall(b) = \sigma_R(s)$ for all $b \in A$. By definition of $\mathcal{G}_\forall$, no move is available for $\exists$ and she gets stuck in the match $\pi_\forall$, which is immediately won by $\forall$.
2. Otherwise, let (b, t) be the next position picked by $\exists$ in $\pi_\forall$. By definition of $\mathcal{G}_\forall$ the node t is not in $m_\forall(b)$. By definition of $m_\forall$, this means that t is in $m_\exists(b)$. Therefore (b, t) is a legitimate move for $\forall$ in $\pi_\exists$ and we let (b, t) be the next basic position in $\pi_\exists$. In this way the same basic positions are maintained in the two matches at round z_{i+1}.

By construction, either the two matches $\pi_\forall$ and $\pi_\exists$ end in the same round, respectively with player $\exists$ and player $\forall$ getting stuck, or they are both infinite. In the latter case, by the fact that $\pi_\exists$ is $f_\exists$-conform and $f_\exists$ is winning for $\exists$, the minimum parity n occurring infinitely often in $\pi_\exists$ in even. By definition of Ω^δ, the minimum parity occurring infinitely often in $\pi_\forall$ is $n + 1$, which is odd. Therefore $\forall$ wins $\pi_\forall$.

($\Leftarrow$) The same argument provided for direction ($\Rightarrow$) works dually for the converse direction. □

Bibliography

[1] ARNOLD, A. Logical definability of fixed points. *Theor. Comput. Sci. 61* (1988), 289–297.

[2] ARNOLD, A., AND NIWINSKI, D. Fixed point characterization of weak monadic logic definable sets of trees. In *Tree Automata and Languages*. 1992, pp. 159–188.

[3] BLACKBURN, P., RIJKE, M. D., AND VENEMA, Y. *Modal logic*. Cambridge University Press, Cambridge England New York, 2001.

[4] BRADFIELD, J. The modal μ-calculus alternation hierarchy is strict. In *CONCUR '96: Concurrency Theory*, U. Montanari and V. Sassone, Eds., vol. 1119 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 1996, pp. 233–246.

[5] DAVEY, B. A., AND PRIESTLEY, H. A. *Introduction to Lattices and Order (2. ed.)*. Cambridge University Press, 2002.

[6] EMERSON, E. A. Temporal and modal logic. In *Handbook of Theoretical Computer Science* (1995), Elsevier, pp. 995–1072.

[7] EMERSON, E. A., AND JUTLA, C. S. Tree automata, μ-calculus and determinacy (extended abstract). In *FOCS* (1991), pp. 368–377.

[8] EMERSON, E. A., AND LEI, C.-L. Efficient model checking in fragments of the propositional μ-calculus (extended abstract). In *LICS* (1986), pp. 267–278.

[9] FACCHINI, A. *A Study on the Expressive Power of Some Fragments of the Modal μ-Calculus*. PhD thesis, University of Lausanne and University of Bordeaux 1.

[10] FONTAINE, G. *Modal fixpoint logic: some model-theoretic questions*. PhD thesis, ILLC (University of Amsterdam), 2010.

[11] GRÄDEL, E., THOMAS, W., AND WILKE, T., Eds. *Automata, Logics, and Infinite Games: A Guide to Current Research* (2002), vol. 2500 of *Lecture Notes in Computer Science*, Springer.

[12] HINTIKKA, J. Game-theoretical semantics: Insights and prospects. In *The Game of Language*, J. Hintikka and J. Kulas, Eds. D. Reidel Publishing Company, 1983, pp. 1–31.

[13] HODGES, W. *Model theory*. Cambridge University Press, Cambridge England New York, 1993.

[14] JANIN, D., AND LENZI, G. On the relationship between monadic and weak monadic second order logic on arbitrary trees, with applications to the μ-calculus. *Fundam. Inform. 61*, 3-4 (2004), 247–265.

[15] JANIN, D., AND WALUKIEWICZ, I. Automata for the modal μ-calculus and related results. In *MFCS* (1995), pp. 552–562.

[16] JANIN, D., AND WALUKIEWICZ, I. On the expressive completeness of the propositional μ-calculus with respect to monadic second order logic. In *Proceedings of the 7th International Conference on Concurrency Theory* (London, UK, 1996), CONCUR '96, Springer-Verlag, pp. 263–277.

[17] JECH, T. *Set theory, Second Edition*. Perspectives in Mathematical Logic. Springer, 1997.

[18] KUPFERMAN, O., AND VARDI, M. Y. On bounded specifications. In *LPAR* (2001), pp. 24–38.

[19] KUPFERMAN, O., AND VARDI, M. Y. Weak alternating automata are not that weak. *ACM Trans. Comput. Logic 2*, 3 (July 2001), 408–429.

[20] KUPFERMAN, O., AND VARDI, M. Y. $\Pi_2 \cap \Sigma_2 \equiv AFMC$. In *ICALP* (2003), pp. 697–713.

[21] KUPFERMAN, O., VARDI, M. Y., AND WOLPER, P. An automata-theoretic approach to branching-time model checking. *J. ACM 47*, 2 (Mar. 2000), 312–360.

[22] MOSTOWSKI, A. Games with forbidden positions. Tech. Rep. 78, University of Gdansk, 1991.

[23] MOSTOWSKI, A. W. Hierarchies of weak automata and weak monadic formulas. *Theor. Comput. Sci. 83*, 2 (1991), 323–335.

[24] MULLER, D. E., SAOUDI, A., AND SCHUPP, P. E. Alternating automata, the weak monadic theory of trees and its complexity. *Theor. Comput. Sci. 97*, 2 (1992), 233–244.

[25] NEUMANN, J., SZEPIETOWSKI, A., AND WALUKIEWICZ, I. Complexity of weak acceptance conditions in tree automata. *Inf. Process. Lett. 84*, 4 (2002), 181–187.

[26] PNUELI, A. The temporal logic of programs. In *Proceedings of the 18th Annual Symposium on Foundations of Computer Science* (Washington, DC, USA, 1977), SFCS '77, IEEE Computer Society, pp. 46–57.

[27] RABIN, M. O. Decidability of second-order theories and automata on infinite trees. *BULLETIN of the American Mathematical Society 74* (July 1968), 1025–1029.

[28] RABIN, M. O. Weakly definable relations and special automata. In *Proceedings of the Symposium on Mathematical Logic and Foundations of Set Theory (SMLFST'70)* (1970), Y. Bar-Hillel, Ed., North-Holland, pp. 1–23.

[29] SAFRA, S. On the complexity of ω-automata. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science* (Washington, DC, USA, 1988), SFCS '88, IEEE Computer Society, pp. 319–327.

[30] VAN BENTHEM, J. *Modal Correspondence Theory*. PhD thesis, Universiteit van Amsterdam, 1977.

[31] VENEMA, Y. Lectures on the modal μ-calculus. Lecture Notes, Institute for Logic, Language and Computation, University of Amsterdam, 2008.

[32] WALUKIEWICZ, I. An extension of Rabin's tree theorem to tree-like structures.

[33] WALUKIEWICZ, I. Monadic second order logic on tree-like structures. In *STACS* (1996), pp. 401–413.

[34] WILKE, T. Alternating tree automata, parity games, and modal μ-calculus. *Bull. Soc. Math. Belg 8* (2001).