London Mathematical Society
Lecture Note Series 379

Finite and Algorithmic Model Theory

Edited by
Javier Esparza, Christian Michaux
and Charles Steinhorn

The London
Mathematical
Society

CAMBRIDGE

London Mathematical Society Lecture Notes series: 379

Finite and Algorithmic Model Theory

Edited by

JAVIER ESPARZA
Technische Universität München, Germany

CHRISTIAN MICHAUX
Université de Mons, Belgium

CHARLES STEINHORN
Vassar College, New York, USA

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521718202

First published 2011

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Finite and algorithmic model theory / edited by Javier Esparza, Christian Michaux,
Charles Steinhorn.
p. cm. – (London Mathematical Society lecture note series ; 379)
Includes index.
ISBN 978-0-521-71820-2 (pbk.)
1. Finite model theory. 2. Computer science – Mathematics. I. Esparza, Javier.
II. Michaux, Christian. III. Steinhorn, Charles. IV. Title. V. Series.
QA9.7.F565 2011
511.3′4 – dc22 2010048062

ISBN 978-0-521-71820-2 Paperback

LONDON MATHEMATICAL SOCIETY LECTURE NOTE SERIES

Managing Editor: Professor M. Reid, Mathematics Institute,
University of Warwick, Coventry CV4 7AL, United Kingdom

The titles below are available from booksellers, or from Cambridge University Press at
www.cambridge.org/mathematics

280 Characters and automorphism groups of compact Riemann surfaces, T. BREUER
281 Explicit birational geometry of 3-folds, A. CORTI & M. REID (eds)
282 Auslander–Buchweitz approximations of equivariant modules, M. HASHIMOTO
283 Nonlinear elasticity, Y.B. FU & R.W. OGDEN (eds)
284 Foundations of computational mathematics, R. DEVORE, A. ISERLES & E. SÜLI (eds)
285 Rational points on curves over finite fields, H. NIEDERREITER & C. XING
286 Clifford algebras and spinors (2nd Edition), P. LOUNESTO
287 Topics on Riemann surfaces and Fuchsian groups, E. BUJALANCE, A.F. COSTA & E. MARTÍNEZ (eds)
288 Surveys in combinatorics, 2001, J.W.P. HIRSCHFELD (ed)
289 Aspects of Sobolev-type inequalities, L. SALOFF-COSTE
290 Quantum groups and Lie theory, A. PRESSLEY (ed)
291 Tits buildings and the model theory of groups, K. TENT (ed)
292 A quantum groups primer, S. MAJID
293 Second order partial differential equations in Hilbert spaces, G. DA PRATO & J. ZABCZYK
294 Introduction to operator space theory, G. PISIER
295 Geometry and integrability, L. MASON & Y. NUTKU (eds)
296 Lectures on invariant theory, I. DOLGACHEV
297 The homotopy category of simply connected 4-manifolds, H.-J. BAUES
298 Higher operads, higher categories, T. LEINSTER (ed)
299 Kleinian groups and hyperbolic 3-manifolds, Y. KOMORI, V. MARKOVIC & C. SERIES (eds)
300 Introduction to Möbius differential geometry, U. HERTRICH-JEROMIN
301 Stable modules and the D(2)-problem, F.E.A. JOHNSON
302 Discrete and continuous nonlinear Schrödinger systems, M.J. ABLOWITZ, B. PRINARI & A.D. TRUBATCH
303 Number theory and algebraic geometry, M. REID & A. SKOROBOGATOV (eds)
304 Groups St Andrews 2001 in Oxford I, C.M. CAMPBELL, E.F. ROBERTSON & G.C. SMITH (eds)
305 Groups St Andrews 2001 in Oxford II, C.M. CAMPBELL, E.F. ROBERTSON & G.C. SMITH (eds)
306 Geometric mechanics and symmetry, J. MONTALDI & T. RATIU (eds)
307 Surveys in combinatorics 2003, C.D. WENSLEY (ed.)
308 Topology, geometry and quantum field theory, U.L. TILLMANN (ed)
309 Corings and comodules, T. BRZEZINSKI & R. WISBAUER
310 Topics in dynamics and ergodic theory, S. BEZUGLYI & S. KOLYADA (eds)
311 Groups: topological, combinatorial and arithmetic aspects, T.W. MÜLLER (ed)
312 Foundations of computational mathematics, Minneapolis 2002, F. CUCKER *et al* (eds)
313 Transcendental aspects of algebraic cycles, S. MÜLLER-STACH & C. PETERS (eds)
314 Spectral generalizations of line graphs, D. CVETKOVIĆ, P. ROWLINSON & S. SIMIĆ
315 Structured ring spectra, A. BAKER & B. RICHTER (eds)
316 Linear logic in computer science, T. EHRHARD, P. RUET, J.-Y. GIRARD & P. SCOTT (eds)
317 Advances in elliptic curve cryptography, I.F. BLAKE, G. SEROUSSI & N.P. SMART (eds)
318 Perturbation of the boundary in boundary-value problems of partial differential equations, D. HENRY
319 Double affine Hecke algebras, I. CHEREDNIK
320 L-functions and Galois representations, D. BURNS, K. BUZZARD & J. NEKOVÁŘ (eds)
321 Surveys in modern mathematics, V. PRASOLOV & Y. ILYASHENKO (eds)
322 Recent perspectives in random matrix theory and number theory, F. MEZZADRI & N.C. SNAITH (eds)
323 Poisson geometry, deformation quantisation and group representations, S. GUTT *et al* (eds)
324 Singularities and computer algebra, C. LOSSEN & G. PFISTER (eds)
325 Lectures on the Ricci flow, P. TOPPING
326 Modular representations of finite groups of Lie type, J.E. HUMPHREYS
327 Surveys in combinatorics 2005, B.S. WEBB (ed)
328 Fundamentals of hyperbolic manifolds, R. CANARY, D. EPSTEIN & A. MARDEN (eds)
329 Spaces of Kleinian groups, Y. MINSKY, M. SAKUMA & C. SERIES (eds)
330 Noncommutative localization in algebra and topology, A. RANICKI (ed)
331 Foundations of computational mathematics, Santander 2005, L.M PARDO, A. PINKUS, E. SÜLI & M.J. TODD (eds)
332 Handbook of tilting theory, L. ANGELERI HÜGEL, D. HAPPEL & H. KRAUSE (eds)

333 Synthetic differential geometry (2nd Edition), A. KOCK
334 The Navier–Stokes equations, N. RILEY & P. DRAZIN
335 Lectures on the combinatorics of free probability, A. NICA & R. SPEICHER
336 Integral closure of ideals, rings, and modules, I. SWANSON & C. HUNEKE
337 Methods in Banach space theory, J.M.F. CASTILLO & W.B. JOHNSON (eds)
338 Surveys in geometry and number theory, N. YOUNG (ed)
339 Groups St Andrews 2005 I, C.M. CAMPBELL, M.R. QUICK, E.F. ROBERTSON & G.C. SMITH (eds)
340 Groups St Andrews 2005 II, C.M. CAMPBELL, M.R. QUICK, E.F. ROBERTSON & G.C. SMITH (eds)
341 Ranks of elliptic curves and random matrix theory, J.B. CONREY, D.W. FARMER, F. MEZZADRI & N.C. SNAITH (eds)
342 Elliptic cohomology, H.R. MILLER & D.C. RAVENEL (eds)
343 Algebraic cycles and motives I, J. NAGEL & C. PETERS (eds)
344 Algebraic cycles and motives II, J. NAGEL & C. PETERS (eds)
345 Algebraic and analytic geometry, A. NEEMAN
346 Surveys in combinatorics 2007, A. HILTON & J. TALBOT (eds)
347 Surveys in contemporary mathematics, N. YOUNG & Y. CHOI (eds)
348 Transcendental dynamics and complex analysis, P.J. RIPPON & G.M. STALLARD (eds)
349 Model theory with applications to algebra and analysis I, Z. CHATZIDAKIS, D. MACPHERSON, A. PILLAY & A. WILKIE (eds)
350 Model theory with applications to algebra and analysis II, Z. CHATZIDAKIS, D. MACPHERSON, A. PILLAY & A. WILKIE (eds)
351 Finite von Neumann algebras and masas, A.M. SINCLAIR & R.R. SMITH
352 Number theory and polynomials, J. MCKEE & C. SMYTH (eds)
353 Trends in stochastic analysis, J. BLATH, P. MÖRTERS & M. SCHEUTZOW (eds)
354 Groups and analysis, K. TENT (ed)
355 Non-equilibrium statistical mechanics and turbulence, J. CARDY, G. FALKOVICH & K. GAWEDZKI
356 Elliptic curves and big Galois representations, D. DELBOURGO
357 Algebraic theory of differential equations, M.A.H. MACCALLUM & A.V. MIKHAILOV (eds)
358 Geometric and cohomological methods in group theory, M.R. BRIDSON, P.H. KROPHOLLER & I.J. LEARY (eds)
359 Moduli spaces and vector bundles, L. BRAMBILA-PAZ, S.B. BRADLOW, O. GARCÍA-PRADA & S. RAMANAN (eds)
360 Zariski geometries, B. ZILBER
361 Words: Notes on verbal width in groups, D. SEGAL
362 Differential tensor algebras and their module categories, R. BAUTISTA, L. SALMERÓN & R. ZUAZUA
363 Foundations of computational mathematics, Hong Kong 2008, F. CUCKER, A. PINKUS & M.J. TODD (eds)
364 Partial differential equations and fluid mechanics, J.C. ROBINSON & J.L. RODRIGO (eds)
365 Surveys in combinatorics 2009, S. HUCZYNSKA, J.D. MITCHELL & C.M. RONEY-DOUGAL (eds)
366 Highly oscillatory problems, B. ENGQUIST, A. FOKAS, E. HAIRER & A. ISERLES (eds)
367 Random matrices: High dimensional phenomena, G. BLOWER
368 Geometry of Riemann surfaces, F.P. GARDINER, G. GONZÁLEZ-DIEZ & C. KOUROUNIOTIS (eds)
369 Epidemics and rumours in complex networks, M. DRAIEF & L. MASSOULIÉ
370 Theory of p-adic distributions, S. ALBEVERIO, A.YU. KHRENNIKOV & V.M. SHELKOVICH
371 Conformal fractals, F. PRZYTYCKI & M. URBAŃSKI
372 Moonshine: The first quarter century and beyond, J. LEPOWSKY, J. MCKAY & M.P. TUITE (eds)
373 Smoothness, regularity and complete intersection, J. MAJADAS & A. G. RODICIO
374 Geometric analysis of hyperbolic differential equations: An introduction, S. ALINHAC
375 Triangulated categories, T. HOLM, P. JØRGENSEN & R. ROUQUIER (eds)
376 Permutation patterns, S. LINTON, N. RUŠKUC & V. VATTER (eds)
377 An introduction to Galois cohomology and its applications, G. BERHUY
378 Probability and mathematical genetics, N.H. BINGHAM & C.M. GOLDIE (eds)
379 Finite and algorithmic model theory, J. ESPARZA, C. MICHAUX & C. STEINHORN (eds)
380 Real and complex singularities, M. MANOEL, M.C. ROMERO FUSTER & C.T.C WALL (eds)
381 Symmetries and integrability of difference equations, D. LEVI, P. OLVER, Z. THOMOVA & P. WINTERNITZ (eds)
382 Forcing with random variables and proof complexity, J. KRAJÍČEK
383 Motivic integration and its interactions with model theory and non-Archimedean geometry I, R. CLUCKERS, J. NICAISE & J. SEBAG (eds)
384 Motivic integration and its interactions with model theory and non-Archimedean geometry II, R. CLUCKERS, J. NICAISE & J. SEBAG (eds)
385 Entropy of hidden Markov processes and connections to dynamical systems, B. MARCUS, K. PETERSEN & T. WEISSMAN (eds)

Preface

This volume is based on the satellite workshop on *Finite and Algorithmic Model Theory* that took place at the University of Durham, January 9–13, 2006, to inaugurate the scientific program *Logic and Algorithms* held at the Isaac Newton Institute for Mathematical Sciences during the first six months of 2006. The goal of the workshop was to explore the emerging and potential connections between finite and infinite model theory, and their applications to theoretical computer science. The primarily tutorial format introduced researchers and graduate students to a number of fundamental topics. The excellent quality of the tutorials suggested to the program organizers, Anuj Dawar and Moshe Vardi, that a volume based on the workshop presentations could serve as a valuable and lasting reference. They proposed this to the workshop scientific committee; this volume is the outcome.

The *Logic and Algorithms* program focused on the connection between two chief concerns of theoretical computer science: (i) how to ensure and verify the correctness of computing systems; and (ii) how to measure the resources required for computations and ensure their efficiency. The two areas historically have interacted little with each other, partly because of the divergent mathematical techniques they have employed. More recently, areas of research in which model-theoretic methods play a central role have reached across both sides of this divide. Results and techniques that have been developed have found applications to fields such as database theory, complexity theory, and verification.

Some brief historical remarks help situate the context for this volume. The study of the model-theoretic properties of finite structures emerged initially as a branch of classical model theory, with its focus primarily on first-order logic. Beginning in the late 1980s, however, research concerning logics on finite structures diverged sharply from work in classical model theory. Classical model theory, with its emphasis on infinite structures, had made dramatic advances both theoretically and in applications to other areas of mathematics. Work on finite structures focused on connections with discrete complexity theory and verification. Indeed, the connections between finite model theory, descriptive complexity theory, parameterized complexity, and state machine verification are now so strong that boundaries between them are hard to distinguish.

The methods employed in these two facets of model theory also grew apart during this period. Probabilistic techniques and machine simulations have played a prominent role in the study of finite structures, and stand in contrast to the geometric, algebraic, and analytic methods that pervade classical (infinite) model theory. Although both classical and finite model theory deal with restricted classes of structures, the conditions by which such classes are delimited also have been quite different. Finite model theory and verification typically concentrate on classes linked to particular computing formalisms, or to which decomposition methods from finite graph theory can be applied. In contrast, infinitary model theory usually places restrictions on combinatorial or geometric properties of the definable sets of a structure.

Yet, there are recent indications of a re-convergence of classical model theory and logical aspects of computer science. This has resulted both from the interest of computer scientists in new computing and specification models that make use of infinitary structures, and from the development of powerful model-theoretic techniques that provide insight into finite structures. If there is an overarching theme, it is how various "tameness" hypotheses used to delimit classes of structures *and* logics have deeply impacted the study of those aspects of theoretical computer science in which model-theory naturally comes into play. The chapters that comprise this volume survey many of the common themes that have emerged and gained attention, and point to the significant potential for wider interaction.

The chapter of Bárány, Grädel, and Rubin, *Automata-based presentations of infinite structures* develops what the authors call *algorithmic model theory*. The authors direct their attention to the "tame" class of *automatic structures*, that is structures that have a presentation in a precise sense by automata operating on finite or infinite words or trees. The goal of this work, to extend algorithmic and logical methods from finite structures to finitely presented infinite structures, has been a focal point for research in computer science, combinatorics, and mathematical logic. This point of view allows structures to be viewed alternately from both a finite and infinite model theoretic perspective. The theory that has emerged makes use of techniques both from classical model theory and theoretical computer science, and has found appealing applications to several areas, including database theory, complexity theory and verification.

Classical model theory by and large concentrates on the analysis of the first-order definable sets over a structure, that is, those sets of n-tuples of the universe whose definition is given by a first-order formula. This analysis has predominantly taken two forms. The first is based on the "structural complexity" of the formula, e.g., the number of alternations of blocks of existential and universal quantifiers appearing in its prenex normal form. This theme is best illustrated by *quantifier elimination*, in which definable sets over a structure are shown to have quantifier-free definitions. The second involves assigning a dimension (with a corresponding notion of independence) to the definable sets that is combinatorially, algebraically, or geometrically motivated. Stability theory, with its combinatorial/algebraic account of dimension and independence, is perhaps the most widely known and longest-studied exemplar, its development traceable to Morley's seminal work in the 1960's and to Shelah's deep and extensive work in the 1970's. More recently, o-minimality, and in particular its focus on o-minimal expansions of the ordered field of real numbers, provides another important class of examples. The imposition of "tameness" assumptions in classical model theory such as stability and o-minimality – often verified in examples by quantifier elimination – make the analysis of the structures satisfying these hypotheses not only tractable but also amenable to applications in mathematics outside of logic.

Tarski's quantifier-elimination for real-closed fields which thereby (effectively) equates the first-order definable sets over the field of real numbers with the semialgebraic sets, has long proved a fertile ground for framing and addressing computational issues. Kuijpers and Van den Bussche, in their chapter, *Logical aspects of spatial databases*, model spatial data via semialgebraic subsets of n-dimensional Euclidean space, and investigate the expressive power of several logic-based languages to query these databases. They first characterize the topological properties of planar spatial databases that are first-order expressible over the usual language for the ordered field of real numbers – of interest from the point of view of geographical information systems, for example – in terms of the query language "cone logic". The second half of their chapter deals with query languages that extend first-order logic over the real field by some form of recursion, including spatial Datalog, and first-order logic extended with a while loop or with a transitive closure operator.

Koponen, in her chapter, *Some connections between finite and infinite model theory*, discusses how stability theoretic considerations, as well as other properties and techniques from classical model theory such as smooth approximation, can be imported successfully into the study of finite structures by restricting to bounded variable logic, that is, first-order logic under the restriction that there is a fixed value k such that only formulas in which no more than k variables occur. In particular, Koponen investigates when a theory in bounded variable logic with an infinite model has arbitrarily large finite models and isolates conditions for effectively determining least upper bounds for the size of the smallest such finite model.

The chapter of Macpherson and Steinhorn, *Definability in classes of finite structures*, contains two distinct threads that draw their motivation from classical model theory. The first, inspired by the model theory of finite and pseudofinite fields, concerns asymptotic classes of finite structures. These are non-elementary classes of finite structures whose first-order definable sets asymptotically satisfy cardinality constraints that permit the assignation of a dimension and measure, and have an intimate connection in classical model theory to so-called simple theories. The second theme concerns so-called *robust classes* of finite structures, whose origin lies in attempting to "finitize" classical model-theoretic tameness conditions, such as o-minimality, that are provably excluded in asymptotic classes. Robust classes consist of directed systems of finite structures in which the truth value of a formula requires "looking ahead" into a larger structure in the system.

For the model theory of finite structures that has been developed with great success within theoretical computer science, "tameness" assumptions do not apply only to isolate classes of structures that are well-behaved with respect to a preferred logic, such as first-order logic. Research has prospered by striking a balance between appropriate logics or fragments thereof and classes of finite structures: that is, tame logics matched with tame classes. This theme appears already in Koponen's chapter, with its emphasis on bounded variable logic combined with classical tameness assumptions, and strongly emerges in the chapters of Otto and Kreutzer. As these chapters furthermore show, this point of view can furnish significant computational insights.

Kreutzer's chapter, *Algorithmic meta-theorems*, discusses how constraining both classes of (finite) structures and logics yields a wealth of algorithmic results. An algorithmic meta-theorem has the form that every computational problem that can be expressed in some logic can be solved efficiently on every class of structures that satisfy certain constraints. This is usually accomplished by showing that the model-checking problem for formulas in some logic – typically first-order or monadic second-order – is what is called *fixed-parameter tractable* for a class of structures, typically based on graphs with well-behaved tree decompositions. This point of view goes back to well-known work of Courcelle and his collaborators.

Otto takes as the focus of his chapter the application of game-oriented methods and explicit model constructions in the analysis of fragments of first-order logic restricted to well-behaved (non-elementary) classes of structures, particularly finite structures. Whereas the model-theoretic compactness theorem plays an essential role in the classical setting, paradigmatically in proving *expressive completeness* results such as the Łos-Tarski theorem characterizing those formulas preserved under extensions as the existential formulas, its failure for restricted classes of structures, e.g., classes of finite structures, motivates the introduction of the methods and techniques that Otto places at the center of the chapter. The chapter also surveys how by restricting to classes of finite structures defined by tree-width and locality considerations, expressive completeness results that fail for the class of finite structures can be regained.

The workshop organizer was Professor Iain Stewart (Durham). The members of the Scientific Committee for the workshop included : Michael Benedikt (Oxford), Javier Esparza (Munich), Bradd Hart (McMaster), Christian Michaux (Mons-Hainaut), Charles Steinhorn (Vassar), and Katrin Tent (Münster). Financial support from the Newton Institute and EPSRC is gratefully acknowledged. We also wish to express our appreciation to the staff at Cambridge University Press, in particular Clare Dennison, our maths/computer science editor, and Sabine Koch, our production editor, for their remarkable thoughtfulness, patience, and efficiency throughout the process of bringing this volume into print.

Javier Esparza
Christian Michaux
Charles Steinhorn

Contents

1

Automata-based presentations of infinite structures

VINCE BÁRÁNY[1], ERICH GRÄDEL[2] AND SASHA RUBIN[3]

1.1 Finite presentations of infinite structures

The model theory of finite structures is intimately connected to various fields in computer science, including complexity theory, databases, and verification. In particular, there is a close relationship between complexity classes and the expressive power of logical languages, as witnessed by the fundamental theorems of descriptive complexity theory, such as Fagin's Theorem and the Immerman-Vardi Theorem (see [78, Chapter 3] for a survey).

However, for many applications, the strict limitation to finite structures has turned out to be too restrictive, and there have been considerable efforts to extend the relevant logical and algorithmic methodologies from finite structures to suitable classes of infinite ones. In particular this is the case for databases and verification where infinite structures are of crucial importance [130]. *Algorithmic model theory* aims to extend in a systematic fashion the approach and methods of finite model theory, and its interactions with computer science, from finite structures to finitely-presentable infinite ones.

There are many possibilities to present infinite structures in a finite manner. A classical approach in model theory concerns the class of *computable structures*; these are countable structures, on the domain of natural numbers, say, with a finite collection of computable functions and relations. Such structures can be finitely presented by a collection of algorithms, and they have been intensively

[1] Oxford University Computing Laboratory
Wolfson Building, Parks Road, Oxford OX1 3QD, United Kingdom
vbarany@logic.rwth-aachen.de
[2] Mathematical Foundations of Computer Science
RWTH Aachen, D-52056 Aachen, Germany
graedel@logic.rwth-aachen.de
[3] Department of Mathematics and Applied Mathematics
University of Cape Town, Private Bag, Rondebosch 7701, South Africa
srubin@math.cornell.edu

studied in model theory since the 1960s. However, from the point of view of algorithmic model theory the class of computable structures is problematic. Indeed, one of the central issues in algorithmic model theory is the effective evaluation of logical formulae, from a suitable logic such as first-order logic (FO), monadic second-order logic (MSO), or a fixed point logic like LFP or the modal μ-calculus. But on computable structures, only the quantifier-free formulae generally admit effective evaluation, and already the existential fragment of first-order logic is undecidable, for instance on the computable structure $(\mathbb{N}, +, \cdot)$.

This leads us to the central requirement that for a suitable logic L (depending on the intended application) the model-checking problem for the class $\mathcal{C}$ of finitely presented structures should be algorithmically solvable. At the very least, this means that the L-theory of individual structures in $\mathcal{C}$ should be decidable. But for most applications somewhat more is required:

Effective semantics: There should be an algorithm that, given a finite presentation of a structure $\mathfrak{A} \in \mathcal{C}$ and a formula $\psi(\bar{x}) \in L$, expands the given presentation to include the relation $\psi^{\mathfrak{A}}$ defined by ψ on $\mathfrak{A}$.

This also implies that the class $\mathcal{C}$ should be closed under some basic operations (such as logical interpretations). Thus we should be careful to restrict the model of computation. Typically, this means using some model of *finite automata* or a very restricted form of rewriting.

In general, the finite means for presenting infinite structures may involve different approaches: logical interpretations; finite axiomatisations; rewriting of terms, trees, or graphs; equational specifications; the use of synchronous or asynchronous automata, etc. The various possibilities can be classified along the following lines:

Internal: a set of finite or infinite words or trees/terms is used to represent the domain of (an isomorphic copy of) the structure. Finite automata/ rewriting-rules compute the domain and atomic relations (eg. prefix-recognisable graphs, automatic structures).

Algebraic: a structure is represented as the least solution of a finite set of recursive equations in an appropriately chosen algebra of finite and countable structures (eg. VR-equational structures).

Logical: structures are described by interpreting them, using a finite collection of formulae, in a fixed structure (eg. tree-interpretable structures). A different approach consists in (recursively) axiomatising the isomorphism class of the structure to be represented.

Transformational: structures are defined by sequences of prescribed transformations, such as graph-unraveling, or Muchnik's iterations applied to certain fixed initial structures (which are already known to have a decidable theory). Transformations can also be transductions, logical interpretations, etc. [23]

The last two approaches overlap somewhat. Also, the algebraic approach can be viewed *generatively*: convert the equational system into an appropriate *deterministic grammar* generating the solution of the original equations [44]. The grammar is thus the finite presentation of the graph. One may also say that internal presentations and generating grammars provide descriptions of the *local structure* from which the whole arises, as opposed to descriptions based on *global symmetries* typical of algebraic specifications.

Prerequisites and notation

We assume rudimentary knowledge of finite automata on finite and infinite words and trees, their languages and their correspondence to monadic second-order logic (MSO) [133, 79]. Undefined notions from logic and algebra (congruence on structures, definability, isomorphism) can be found in any standard textbook. We mainly consider the following logics $\mathcal{L}$: first-order (FO), monadic second order (MSO), and weak monadic second-order (wMSO) which has the same syntax as MSO, but the intended interpretation of the set variables is that they range over *finite* subsets of the domain of the structure under consideration.

We mention the following to fix notation: infinite words are called ω-words and infinite trees are called ω-trees (to distinguish them from finite ones); relations computable by automata will be called *regular*; the domain of a *structure* $\mathfrak{B}$ is usually written B and its relations are written $R^{\mathfrak{B}}$. An MSO-formula $\phi(X_1, \ldots, X_j, x_1, \ldots, x_k)$ interpreted in $\mathfrak{B}$ *defines* the set $\phi^{\mathfrak{B}} := \{(B_1, \ldots, B_j, b_1, \ldots, b_k) \mid B_i \subset B, b_i \in B, \mathfrak{B} \models \phi(B_1, \ldots, B_j, b_1, \ldots, b_k)\}$. A wMSO-formula is similar except that the B_i range over finite subsets of B. The *full binary tree* $\mathfrak{T}_2$ is defined as the structure

$$\left(\{0, 1\}^*, \mathrm{suc}_0, \mathrm{suc}_1\right)$$

where the successor relation suc_i consists of all pairs (x, xi). Tree automata operate on Σ-*labelled trees* $T : \{0, 1\}^* \to \Sigma$. Such a tree is identified with the structure

$$\left(\{0, 1\}^*, \mathrm{suc}_0, \mathrm{suc}_1, \{T^{-1}(\sigma)\}_{\sigma \in \Sigma}\right).$$

Rabin proved the decidability of the MSO-theory of $\mathfrak{T}_2$ and the following fundamental correspondence between MSO and tree automata (see [132] for an overview):

For every monadic second-order formula $\varphi(\overline{X})$ in the signature of $\mathfrak{T}_2$ there is a tree automaton $\mathcal{A}$ (and vice versa) such that

$$L(\mathcal{A}) = \{T_{\overline{X}} \mid \mathfrak{T}_2 \models \varphi(\overline{X})\} \tag{1.1}$$

where $T_{\overline{X}}$ denotes the tree with labels for each X_i.

Similar definitions and results hold for r-ary trees, in which case the domain is $[r]^*$ where $[r] := \{0, \ldots, r-1\}$, and finite trees.

In section 1.2.2 and elsewhere we do not distinguish between a term and its natural representation as a tree. Thus we may speak of infinite terms. We consider countable, vertex- and edge-labelled graphs possibly having distinguished vertices (called sources), and no parallel edges of the same label. A graph is *deterministic* if each of its vertices is the source of at most one edge of each edge label.

Interpretations

Interpretations allow one to define an isomorphic copy of one structure in another. Fix a logic $\mathcal{L}$. A d-dimensional $\mathcal{L}$-*interpretation* $\mathcal{I}$ of structure $\mathfrak{B} = (B; (R_i^{\mathfrak{B}})_i)$ in structure $\mathfrak{A}$, denoted $\mathfrak{B} \leq_{\mathcal{L}}^{\mathcal{I}} \mathfrak{A}$, consists of the following $\mathcal{L}$-formulas in the signature of $\mathfrak{A}$,

- a domain formula $\Delta(\overline{x})$,
- a relation formula $\Phi_{R_i}(\overline{x}_1, \ldots, \overline{x}_{r_i})$ for each relation symbol R_i, and
- an equality formula $\epsilon(\overline{x}_1, \overline{x}_2)$,

where each $\Phi_{R_i}^{\mathfrak{A}}$ is a relation on $\Delta^{\mathfrak{A}}$, each of the tuples $\overline{x}_i, \overline{x}$ contain the same number of variables, d, and $\epsilon^{\mathfrak{A}}$ is a congruence on the structure $(\Delta^{\mathfrak{A}}, (\Phi_{R_i}^{\mathfrak{A}})_i)$, so that $\mathfrak{B}$ is isomorphic to

$$(\Delta^{\mathfrak{A}}, (\Phi_{R_i}^{\mathfrak{A}})_i) \,/\, \epsilon^{\mathfrak{A}} \,.$$

If $\mathcal{L}$ is FO then the free $\overline{x}$ are FO and we speak of a *FO interpretation*. If $\mathcal{L}$ is MSO (wMSO) but the free variables are FO, then we speak of a *(weak) monadic second-order interpretation.*

We associate with $\mathcal{I}$ a transformation of formulas $\psi \mapsto \psi^{\mathcal{I}}$. For illustration we define it in the first-order case: the variable x_i is replaced by the d-tuple $\overline{y}_i$, $(\psi \vee \phi)^{\mathcal{I}}$ by $\psi^{\mathcal{I}} \vee \phi^{\mathcal{I}}$, $(\neg\psi)^{\mathcal{I}}$ by $\neg\psi^{\mathcal{I}}$, $(\exists x_i \psi)^{\mathcal{I}}$ by $\exists \overline{y}_i \Delta(\overline{y}_i) \wedge \psi^{\mathcal{I}}$, and $(x_i = x_j)^{\mathcal{I}}$ is replaced by $\epsilon(\overline{y}_i, \overline{y}_j)$. Thus one can translate $\mathcal{L}$ formulas from the signature of $\mathfrak{B}$ into the signature of $\mathfrak{A}$.

Proposition 1.1.1 *If $\mathfrak{B} \leq_{\mathcal{L}}^{\mathcal{I}} \mathfrak{A}$, say the isomorphism is f, then for every formula $\psi(x_1, \ldots, x_k)$ in the signature of $\mathfrak{B}$ and all k-tuples $\overline{b}$ of elements of*

$\mathfrak{B}$ it holds that

$$\mathfrak{B} \models \psi(b_1, \ldots, b_k) \iff \mathfrak{A} \models \psi^{\mathcal{I}}(f(b_1), \ldots, f(b_k))$$

In particular, if $\mathfrak{A}$ has decidable $\mathcal{L}$-theory, then so does $\mathfrak{B}$.

Set interpretations

When $\mathcal{L}$ is MSO (wMSO) and the free variables are MSO (wMSO) the interpretation is called a *(finite) set interpretation*. In this last case, we use the notation $\mathfrak{B} \leq^{\mathcal{I}}_{\text{set}} \mathfrak{A}$ or $\mathfrak{B} \leq^{\mathcal{I}}_{\text{fset}} \mathfrak{A}$. We will only consider (finite) set interpretations of dimension 1.

If finiteness of sets is MSO-definable in some structure $\mathfrak{A}$ (as for linear orders or for finitely branching trees) then every structure $\mathfrak{B}$ having a finite-set interpretation in $\mathfrak{A}$ can also be set interpreted in $\mathfrak{A}$.

Example 1.1.2 An interpretation $(\mathbb{N}, +) \leq^{\mathcal{I}}_{\text{fset}} (\mathbb{N}, 0, \texttt{suc})$ based on the binary representation is given by $\mathcal{I} = (\varphi(X), \varphi_+(X, Y, Z), \varphi_=(X, Y))$ with $\varphi(X)$ always true, $\varphi_=$ the identity, and $\varphi_+(X, Y, Z)$ is

$$\exists C\, \forall n\ [(Zn \leftrightarrow Xn \oplus Yn \oplus Cn) \land (C(\texttt{suc}n) \leftrightarrow \mu(Xn, Yn, Cn)) \land \neg C0]$$

where C stands for carry, $\oplus$ is exclusive or, and $\mu(x_0, x_1, x_2)$ is the majority function, in this case definable as $\bigvee_{i \neq j} x_i \land x_j$.

To every (finite) subset interpretation $\mathcal{I}$ we associate, as usual, a transformation of formulas $\psi \mapsto \psi^{\mathcal{I}}$, in this case mapping first-order formulas to (weak) monadic second-order formulas.

Proposition 1.1.3 *Let $\mathfrak{B} \leq^{\mathcal{I}}_{\text{(f)set}} \mathfrak{A}$ be a (finite) subset interpretation with isomorphism f. Then to every first-order formula $\psi(x_1, \ldots, x_k)$ in the signature of $\mathfrak{B}$ one can effectively associate a (weak) monadic second-order formula $\psi^{\mathcal{I}}(X_1, \ldots, X_k)$ in the signature of $\mathfrak{A}$ such that for all k-tuples $\overline{b}$ of elements of $\mathfrak{B}$ it holds that*

$$\mathfrak{B} \models \psi(b_1, \ldots, b_k) \iff \mathfrak{A} \models \psi^{\mathcal{I}}(f(b_1), \ldots, f(b_k)).$$

Consequently, if the (weak) monadic-second order theory of $\mathfrak{A}$ is decidable then so is the first-order theory of $\mathfrak{B}$.

For more on subset interpretations we refer to [23].

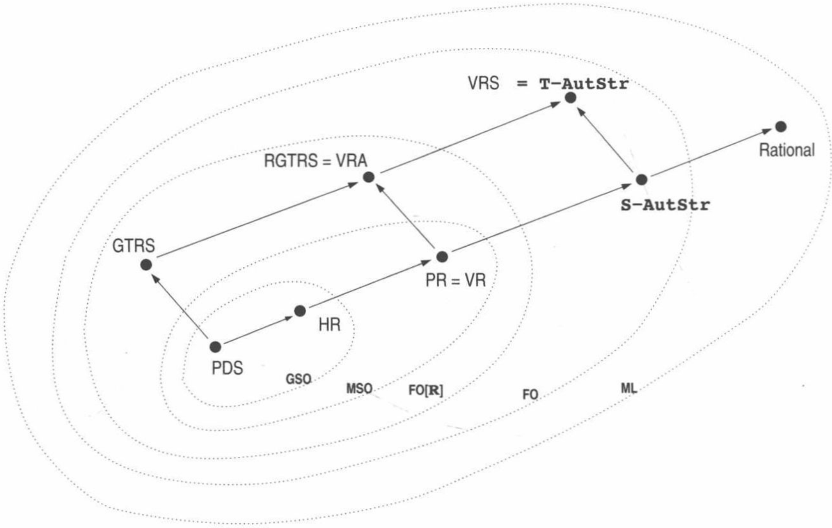

Figure 1.1 Relationship of graph classes and logical decidability boundaries.

1.2 A hierarchy of finitely presentable structures

This section provides an overview of some of the prominent classes of graphs and their various finite presentations.

These developments are the product of over two decades of research in diverse fields. We begin our exposition with the seminal work of Muller and Schupp on context-free graphs, we mention prefix-recognisable structures, survey hyperedge-replacement and vertex-replacement grammars and their corresponding algebraic frameworks leading up to equational graphs in algebras with asynchronous or synchronous product operation. These latter structures are better known in the literature by their automatic presentations, and constitute the topic of the rest of this survey.

As a unifying approach we discuss how graphs belonging to individual classes can be characterised as least fixed-point solutions of finite systems of equations in a corresponding algebra of graphs. We illustrate on examples how to go from graph grammars through equational presentations and interpretations to internal presentations and vice versa.

We briefly summarise key results on Caucal's pushdown hierarchy and more recent developments on simply-typed recursion schemes and collapsible pushdown automata.

Figure 1.1 provides a summary of some of the graph classes discussed in this section together with the boundaries of decidability for relevant logics.