Complete Axiomatizations of Fragments of Monadic Second-Order Logic on Finite Trees

Amélie Gheerbrant
School of Informatics
University of Edinburgh
Email: agheerbr@inf.ed.ac.uk

Balder ten Cate
Department of Computer Science
University of California Santa Cruz
Email: btencate@ucsc.edu

***Abstract*—We present a general model-theoretic technique that we developed and used in [3], [4] to obtain complete axiomatizations of fragments of MSO on finite trees. There is much interest in studying logics on finite trees, and many logics of interest are fragments of MSO. Previously FO axiomatizations were known. To produce axiomatizations beyond FO, we had to develop a new technique that combines classical tools from infinite model theory (Henkin semantics for higher-order logics) with those more typical in finite model theory (Ehrenfeucht-Fraïssé games, and their composition). The key idea behind the technique is to analyze infinite Henkin models of our axioms, and use games to show that they are elementarily equivalent to finite trees.**

Given the general interest in the LICS community in logics on finite trees, and a new set of tools developed by us (that combine classical and finite model theory), we believe that a brief account of this work will be of interest to the LICS audience.

I. Introduction

Recently there has been much interest in studying logics over ordered unranked trees, mainly due to connections with XML research, since labeled unranked trees serve as a standard abstraction of XML documents. Logics are used to describe the structure of XML documents, and to query data they contain, and absolute majority of those used in this context happen to be fragments of MSO (see [8]).

The goal of this work is to obtain *complete axiomatizations* of MSO and its fragments on finite node-labeled sibling-ordered trees. Such axiomatizations have previously been presented for FO-theories [5] but extending the work to MSO presents a number of challenges. We address those by developing a new *model-theoretic technique* by which we obtain complete axiomatizations not only of MSO but also of some of its fragments, such as monadic transitive closure logic and monadic least fixed-point logic.

The new tools we developed combine traditional model-theoretic techniques used to show completeness, with techniques more common in finite model theory, namely Ehrenfeucht-Fraïssé games as well as techniques for composing games. These results have not been presented in the forums traditionally attended by LICS attendees: the conference version appeared in [3] (the journal version is to appear in [4]). As this work addresses traditional LICS topics via using a new set of techniques, we believe that the LICS community could be interested in a short presentation of this work.

To give a flavor of our results and techniques, below we describe the approach for MSO. We give a brief account of the key ingredients: Henkin completeness and Feferman-Vaught theorems, that we need to obtain our results. References [3], [4] can be consulted for details, as well as for extensions to various MSO fragments.

II. Henkin Completeness

It is well known that MSO is highly undecidable on arbitrary standard structures and hence not recursively enumerable. However, Henkin [7] formulated a non-standard semantics for logics of even higher order, and showed that under this interpretation, they can be completely axiomatized. In the case of MSO, the procedure amounts to allowing "non-standard" or "Henkin" interpretations of MSO-formulas in addition to their standard interpretations. In such non standard structures, the set quantifier is interpreted as ranging not over the whole powerset of the domain, but over one of its explicitly given subsets, required to satisfy some good closure conditions. This means that each Henkin structure is given as a pair, containing a usual relational structure together with a subset of the powerset of its domain. A point of particular interest to us is that on finite structures, the mandatory closure conditions are only satisfied by the whole powerset of the domain. It follows that finite Henkin structures are always equivalent to standard structures. This point matters here for the following reason. As a first step of our proof, we show that our axioms are complete on the class of their Henkin models, but the problematic thing at this stage is that some of these models might not be finite trees. However, it is straightforward to infer from our axioms that a finite structure satisfies them if and only if it is a tree. Hence, if there are Henkin models of our axioms which are not finite trees, they have to be infinite. In what remains, we need to show that such infinite models can be "dismissed".

III. Feferman-Vaught Theorems

Even though our main completeness result concerns finite trees, inside the proof we need to consider infinite Henkin structures. In this context even such basic notions as substructures, as well as methods for forming new structures out of existing ones have to be redefined carefully. There is a whole range of model-theoretic methods to form new structures out of existing ones [6], [9]. Familiar constructions like disjoint unions are redefined as particular cases of a notion of generalized product of FO-structures and abstract properties of such products are studied. Results telling how theories of complex structures can be obtained from theories of the components they are built from are known as Feferman-Vaught theorems (who proved the first such result in [6]).

Here we are particularly interested in a type of Feferman-Vaught theorem which establishes that generalized products of relational structures preserve elementary equivalence. We show such a result for a particular case of generalized product of Henkin-structures called fusion. These preservation results are shown with the crucial help of Ehrenfeucht-Fraïssé games that are suitable to use on Henkin structures. More precisely, by combining winning strategies in these games, we show that for every $n \in \mathbb{N}$, the MSO n-theory of the fusion structure reduces to the MSO n-theories of the components structures (by MSO n-theory, we mean the restriction of the theory to MSO-formulas of quantifier depth n). We believe that such general combination techniques for Henkin structures have independent interest. We refer to [9] for an extensive discussion of the question in the more restricted context of standard structures.

IV. "Real" Completeness

With the key ingredients – Henkin models and Feferman-Vaught theorems - we can obtain complete axiomatizations for finite trees. We cannot present the entire axiomatization in this very short abstract. The axioms are roughly subdivided into three group: "generic" axioms true in well-behaved logics (e.g., propositional tautologies, properties of substitution), axioms stating properties of binary predicates defining the trees (e.g., transitivity of the descendant relation), and, crucially, the induction scheme.

To use the previous ingredients to obtain completeness, we define quasi-trees as Henkin models of our axioms. We then use our Feferman-Vaught results to show that MSO cannot distinguish quasi-trees from finite trees. Since finiteness is definable on trees in MSO, it then follows that every model of our axioms is a finite tree. Note that this is in sharp contrast with the FO-theory of finite trees, which does have infinite models.

Let us now briefly sketch the details of our final completeness argument. In order to proceed inductively, it is more convenient to consider a stronger version of the result concerning Henkin substructures of quasi-trees that we call quasi-forests. To grab some intuition consider a finite tree and remove the root node; then it is no longer a finite tree. Instead it is a finite sequence of trees, whose roots stand in a linear sibling order. It does not have a unique root, but it does have a unique left-most root: it is a finite forest. Now given a node a in a quasi-tree T, we let T_a be the Henkin substructure of T generated by the set of its siblings to the right and of their descendants. We call T_a a quasi-forest. Using our game composition results, we finally complete our proof by showing that for each n and for each node a in a quasi-tree, the quasi-forest T_a is n-equivalent to a finite forest. The argument essentially relies on an inductive axiom scheme, which for simplicity we only give here for the restricted case of MSO on finite words ($\varphi(x)$ standing here for any definable MSO property of finite words):

$$\forall x(\forall y((x < y \to \varphi(y)) \to \varphi(x)) \to \forall z\varphi(z)$$

Extending our approach to other classes of finite structures would involve finding comparable induction schemes. This suggests that other natural candidates would be fragments of MSO on classes of finite structures for which MSO satisfiability is decidable (e.g., structures of bounded treewidth).

V. Conclusion

In [3], [4] we obtained complete axiomatizations not only for MSO but also for monadic transitive closure and least fixed point logics on finite trees. The method we developed is quite uniform and can be used for other logics as well. While it follows the route used in modal logic, where "canonical models" are often transformed in order to obtain intended models [1], its key new element is the use of Henkin semantics: the model we first create is a Henkin model, and then we modify it to obtain a model that is among our intended ones. There are related complete axiomatizations on infinite models [2], [10], [11], where completeness proofs are based on automata-theoretic techniques, which are probably harder to adapt to obtain axiomatizations in the finite case. This leads to an intriguing question whether some of our model-theoretic techniques could also be used as an alternative to automata, in order to show other sorts of results, not necessarily related to complete axiomatizations.

References

[1] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2000.

[2] J. R. Büchi and D. Siefkes. *Decidable Theories: Vol. 2: The Monadic Second Order Theory of All Countable Ordinals*. Lectures Notes in Mathematics. Springer, Berlin, Heidelberg, 1973.

[3] B. ten Cate and A. Gheerbrant. Complete Axiomatizations of MSO, FO(TC^1) and FO(LFP^1) on finite trees. In *LFCS*, pages 180–196, 2009.

[4] B. ten Cate and A. Gheerbrant. Complete Axiomatizations of fragments of MSO on finite trees. To appear in *LMCS*, preprint here: http://homepages.inf.ed.ac.uk/agheerbr/lmcs.pdf.

[5] K. Doets. *Completeness and Definability : Applications of the Ehrenfeucht Game in Second-Order and Intensional Logic*. PhD thesis, Universiteit van Amsterdam, 1987.

[6] S. Feferman and R. Vaught. The first-order properties of algebraic systems. *Fundamenta Mathematicae*, 47:57–103, 1959.

[7] L. Henkin. Completeness in the Theory of Types. *The Journal of Symbolic Logic*, 15(2):81–91, 1950.

[8] L. Libkin. Logics for Unranked Trees: An Overview. *Logical Methods in Computer Science*, 2(3), 2006.

[9] J. A. Makowsky. Algorithmic uses of the Feferman Vaught Theorem. *Annals of Pure and Applied Logic*, 126(1–3):159–213, 2004.

[10] D. Siefkes. An Axiom System for the Weak Monadic Second Order Theory of Two Successors. *Isr. J. of Mathematics*, 30(3):264–284, 1978.

[11] C. Zaiontz. Axiomatization of the Monadic Theory of Ordinals $< \omega^2$. *Mathematical Logic Quarterly*, 29(6):337–356, 1983.