INDIAN INSTITUTE OF TECHNOLOGY – BOMBAY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SYNOPSIS

of the Ph.D. thesis entitled

A Generalization of the Łoś-Tarski Preservation Theorem

proposed to be submitted in partial fulfillment
of the requirements of the degree of

DOCTOR OF PHILOSOPHY

of the

INDIAN INSTITUTE OF TECHNOLOGY – BOMBAY

by

Abhisekh Sankaran
(Roll No. 07405001)

under the guidance of

Professor Supratik Chakraborty
and
Professor Bharat Adsul

Introduction

Classical model theory is a branch of mathematical logic which studies the relationship between a formal language and its interpretations, also called models. The most well-studied formal language in classical model theory is first order logic (henceforth called FO), that is a language built up from predicates, functions and constant symbols using boolean connectives, and existential and universal quantification. Amongst the earliest areas of study in classical model theory, is a class of results called *FO preservation theorems*. An FO preservation theorem for a model-theoretic operation syntactically characterizes classes of structures that are defined using FO theories, and that are preserved under that operation. One of the earliest preservation theorems is the Łoś-Tarski theorem (1954-55) [20], which states that a class of structures that is defined by an FO theory is preserved under substructures if, and only if, it is definable by a theory of universal sentences, the latter being FO sentences in which only universal quantifications appear (see Theorem 3.2.2 in [7]). In "dual" form, this theorem states that a class of structures that is defined by an FO theory is preserved under extensions if, and only if, it is definable by a theory of existential sentences which are FO sentences that use only existential quantifications. Historically speaking, the study of preservation theorems began with Marczewski asking in 1951, which FO definable classes of structures are preserved under surjective homomorphisms. This question triggered off an extensive study of preservation theorems in which a variety of model-theoretic operations like substructures, extensions, homomorphisms, unions of chains, direct products, reduced products, etc. were taken up and FO preservation theorems for these operations were proven.

While a preservation theorem can be seen as providing a syntactic characterization of a preservation property, the same theorem, flipped around, can also be seen as providing a semantic characterization (and furthermore, via a preservation property) of a syntactic class of FO theories. Thus, the Łoś-Tarski theorem provides semantic characterizations of existential and universal theories, in terms of preservation under extensions and preservation under substructures respectively. Existential and universal theories are equivalent respectively to, what are known in the literature as Σ^0_1 and Π^0_1 theories. For $n \geq 1$, a Σ^0_n theory is a set of Σ^0_n sentences, where a Σ^0_n sentence is an FO sentence in which from left to right, there are n blocks of quantifiers (equivalently, $n-1$ alternations of quantifiers) beginning with a block of existential quantifiers, followed by a quantifier-free formula. Likewise a Π^0_n theory is a set of Π^0_n sentences, where a

Π^0_n sentence is an FO sentence in which from left to right, there are n blocks of quantifiers beginning with a block of universal quantifiers, followed by a quantifier-free formula. As already mentioned, the Łoś-Tarski theorem provides semantic characterizations of Σ^0_1 and Π^0_1 theories. For Σ^0_n sentences and Π^0_n theories for $n \geq 2$, semantic characterizations were proven using preservation properties defined in terms of *ascending chains* and *descending chains* (see Theorem 5.2.8 in [7]). Finally in 1960, Keisler proved the *n-sandwich theorem* [24] that provides a characterization of Σ^0_n and Π^0_n theories, for each $n \geq 1$, using preservation properties defined uniformly in terms of the notion of *n-sandwiches*.

While the aforementioned properties characterize Σ^0_n and Π^0_n theories, and hence Σ^0_n and Π^0_n sentences, "as a whole", none of these characterize Σ^0_n and Π^0_n sentences/theories in which for some given block, the number of quantifiers in that block is *fixed* to a given natural number k. Further, all of these properties are in terms of notions that are "infinitary", i.e. notions that are non-trivial only when arbitrary (i.e. finite and infinite) structures are considered, and that become trivial when restricted to only finite structures. Given the active interest in preservation theorems in the finite model theory context (which we soon describe), none of the properties mentioned above can be used to characterize Σ^0_n and Π^0_n sentences in the finite, for $n \geq 2$.

These observations raise the following two natural questions:

(Q1) Are there properties that semantically characterize (over arbitrary structures) Σ^0_n and Π^0_n sentences/theories in which the number of quantifiers appearing in a given block(s) is fixed to a given natural number(s)?

(Q2) Are there characterizations of Σ^0_n and Π^0_n sentences/theories in terms of notions that are finitary and combinatorial? Whereby these notions can possibly serve to characterize Σ^0_n and Π^0_n sentences over finite structures as well.

In the classical model theory part of this thesis, we consider the case when $n = 2$, and present our partial results towards addressing the above questions. Specifically, for the case of Σ^0_2 and Π^0_2 sentences, in which the number of quantifiers in the *leading block* is fixed to a given natural number, we identify preservation properties that positively answer both (Q1) and (Q2). In other words, we present quantitative dual parameterized preservation properties that are finitary and combinatorial, and that characterize (over arbitrary structures) Σ^0_2 and Π^0_2 sentences whose quantifier prefixes are of the form $\exists^k\forall^*$ or $\forall^k\exists^*$. Our properties for the case of $k = 0$ are exactly the classical properties of preservation under substructures and preservation under extensions, whereby our characterizations of $\exists^k\forall^*$ and $\forall^k\exists^*$ sentences yield the Łoś-Tarski theorem for sentences for the case of $k = 0$. We hence call our characterizations collectively as the *generalized Łoś-Tarski theorem for sentences at level* k, and denote it as $\mathsf{GLT}(k)$. We describe this result, its extension to theories, and other related results in more detail in Section .

The Σ^0_2 and Π^0_2 classes have always been of interest and importance in the literature. After Hilbert posed the *Entscheidungsproblem* in the 1900s, namely the problem of deciding if a given FO sentence is satisfiable, one of the first classes of FO sentences for which satisfiability was shown to be decidable, was the Σ^0_2 class. This was shown by Bernays and Schönfinkel in 1928 for Σ^0_2 sentences without equality, and subsequently extended to full Σ^0_2 by Ramsey [30] (on a historical note: it was in showing this result that Ramsey proved the famous *Ramsey's theorem*). In a subsequent extensive research of about 70 years on the satisfiability problem for prefix classes, it was shown [5] that Σ^0_2 is indeed one of the *maximal* prefix classes for which the satisfiability problem is decidable. With the growth of parameterized complexity theory [9], it became interesting to study the computational complexity of the satisfiability problem for the Σ^0_2 class, in terms of *counts of quantifiers* as parameters. As shown in [5], satisfiability for the Σ^0_2 class is in $\mathsf{NTIME}((n \cdot k^m)^c)$, where n is the length of the input sentence, k and m are the number of existential and universal quantifiers respectively in the sentence, and c is a suitable constant. (On the other hand, for Π^0_2 class, it turns out that if the number of universal quantifiers is at least two, then satisfiability checking for this class is undecidable.) In recent years, there has been significant interest in the Σ^0_2 class from the program verification and program synthesis communities as well [12, 18, 29, 32]. Here, the Σ^0_2 class is also referred to as *effectively propositional logic*. For the Π^0_2 class on the other hand, the database community has shown a lot of active interest in this class in the context of data exchange, data integration and data interoperability [6, 14, 25, 27], and much more recently, in the context of query answering over RDF and OWL knowledge [22, 23].

Returning to preservation theorems, with the advent of finite model theory with the work of Fagin who proved the first descriptive complexity theory result characterizing the complexity class NP in terms of existential second order logic [13], it became interesting to study preservation theorems in the context of finite structures. It turns out that many important results and techniques of classical model theory fail in the context of finite structures. The most stark failure is that of the compactness theorem, which is the most central tool in classical model theory. Consequently, all proofs based on the compactness theorem – indeed this includes the proofs of almost every preservation theorem – fail when restricted to only finite structures. But worse still, the statements of most preservation theorems fail too. The Łoś-Tarski theorem fails in the finite; Tait [33] showed there is an FO sentence that is preserved under extensions over the class of all finite structures, but that is not equivalent over this class, to any existential sentence (this result was rediscovered later by Gurevich and Shelah in 1984). The other preservation theorems from the classical model theory literature mentioned earlier, namely those characterizing Σ_n^0 and Π_n^0 sentences/theories for $n \geq 2$, fail in the finite too; this is simply because the characterizing notions become trivial over finite structures. The homomorphism preservation theorem is a rare theorem however that survives passage into the finite, and this is a landmark result due to Rossman [31]. But then this is an exception.

To "recover" classical preservation theorems in the finite model theory setting, recent research [3, 4, 8, 10, 19] has focussed attention on studying these theorems over "well-behaved" classes of finite structures. In particular, Atserias, Dawar and Grohe showed in [4] that under suitable closure assumptions, classes of structures that are acyclic or of bounded degree admit the Łoś-Tarski theorem for sentences. Likewise, the class of all structures of tree-width at most k also admits the Łoś-Tarski theorem, for each natural number k. Some of the aforesaid classes, like those of bounded tree-width, have proved especially useful in modern graph structure theory and also from an algorithmic point of view. For instance, many computational problems that are otherwise intractable, become tractable when restricted to structures of bounded treewidth [9]. Atserias, Dawar and Kolatis showed that for the aforesaid classes of structures, the homomorphism preservation theorem also holds [3] (Note that this theorem being true over all structures does not imply that it would be true over subclasses of finite structures; restricting attention to a subclass weakens both the hypothesis and the consequent of the statement of the theorem). Subsequently, Harwath, Heimberg and Schweikardt [19] studied the bounds for an effective version of the Łoś-Tarski theorem and the homomorphism preservation theorem over bounded degree structures. In [10], Duris showed that the Łoś-Tarski theorem holds for structures that are acyclic in a more general sense.

The above results give characterizations of the Σ_1^0 and Π_1^0 classes of sentences over the special classes of structures mentioned. What happens to our characterizations of the Σ_2^0 and Π_2^0 classes of sentences, given by $\mathsf{GLT}(k)$, over the aforesaid special classes? It unfortunately turns out that none of the above classes, in general, admits $\mathsf{GLT}(k)$ for any $k \geq 2$. We show that the existence of induced paths of unbounded length in a class is, under reasonable assumptions on the class, the reason for the failure of our characterizations over the class. Since these assumptions are satisfied by the aforesaid special classes and the latter allow unbounded induced path lengths in general, $\mathsf{GLT}(k)$ fails over these classes in general. This therefore motivates us to ask the following:

(Q3) Can we identify structural properties (possibly abstract) of classes of finite structures, that are satisfied by interesting classes of finite structures, and that admit $\mathsf{GLT}(k)$? And furthermore, admit $\mathsf{GLT}(k)$ in effective form?

We answer this question affirmatively. In the finite model theory part of this thesis, we define a parameterized logic-based combinatorial property of classes of finite structures that entails $\mathsf{GLT}(k)$, and even in effective form, under suitable additional assumptions on our property. We call our property as the *$\mathcal{L}$-equivalent bounded substructure property*, denoted $\mathcal{L}\text{-}\mathsf{EBSP}(\mathcal{S}, k)$, where $\mathcal{S}$ is a class of finite structures, k is a natural number, and $\mathcal{L}$ is either FO or an extension of FO called monadic second order logic (MSO). We show that a variety of classes of structures that are of interest to computer science and finite model theory satisfy $\mathcal{L}\text{-}\mathsf{EBSP}(\mathcal{S}, k)$; examples of these include words, trees (unorderd, ordered or ranked), nested words, graphs classes of bounded tree-depth, graph classes of bounded shrub-depth and n-partite cographs. All of these classes of structures have enjoyed (and continue to enjoy) extensive interest from the computer science community. Further these classes, except for words and trees, are very recent (defined within the last 10-12 years). We give general methods to construct new classes of structures that satisfy $\mathcal{L}\text{-}\mathsf{EBSP}(\cdot, k)$ from classes known to satisfy the latter property. Since $\mathsf{GLT}(k)$ for the case of $k = 0$ is exactly the Łoś-Tarski theorem, we get a whole array of new classes that were earlier not known to satisfy the Łoś-Tarski theorem. Interestingly, it turns out that all of these new classes satisfy the homomorphism preservation theorem as well. The above, and related results, are explained in more detail in Section .

We find it worth mentioning of $\mathcal{L}\text{-}\mathsf{EBSP}(\mathcal{S}, k)$ that it can be seen to be a finitary analogue of the property that the classical downward Löwenheim-Skolem theorem (one the first results of classical model theory) states of FO and arbitrary structures. We explicate this connection in Section . The importance of the downward Löwenheim-Skolem theorem in classical model theory can be gauged from the fact that this theorem, along with the compactness theorem, characterizes FO. It indeed is pleasantly surprising that while the downward Löwenheim-Skolem theorem is by itself meaningless over finite structures, a natural finitary analogue of the model theoretic property that this theorem talks about, is satisfied by a wide spectrum of classes of finite structures, that are of interest and importance in computer science and finite model theory.

In the remainder of this synopsis, we formally state our main results, techniques and some interesting open questions, first in the classical model theory setting, and subsequently all of those in the finite model theory setting.

Results in the classical model theory context

We define new dual preservation properties, that for a natural number k as a parameter, provide natural parameterized generalizations of the classical properties of preservation under substructures and preservation under extensions. Specifically, we define the property of *preservation under substructures modulo k-cruxes* as a generalization of the property of preservation under substructures, as follows.

Definition 1. A theory T is said to be *preserved under substructures modulo k-cruxes*, abbreviated as *T is $PSC(k)$*, if for every model $\mathfrak{A}$ of T, there exists a set C of at most k elements of $\mathfrak{A}$ such that every substructure of $\mathfrak{A}$, *that contains C*, is also a model of T. The set C is called a k-crux of $\mathfrak{A}$ with respect to T. A sentence ϕ is said to be $PSC(k)$ if the theory $\{\phi\}$ is $PSC(k)$.

It is easy to see that $PSC(0)$ is exactly the property of preservation under substructures. Likewise, on the dual front, we introduce the notion of *preservation under k-ary covered extensions*, denoted $PCE(k)$, as a natural generalization of the property of preservation under extensions, that is equivalent to the negation of the property of $PSC(k)$. The *generalized Łoś-Tarski theorem for sentences at level k*, denoted $\mathsf{GLT}(k)$, gives syntactic characterizations for $PSC(k)$ and $PCE(k)$, and is as stated below.

Theorem 2 ($\mathsf{GLT}(k)$)**.** *The following hold over arbitrary structures for each $k \in \mathbb{N}$.*

1. *An FO sentence is $PSC(k)$ if, and only if, it is equivalent to an $\exists^k\forall^*$ sentence.*
2. *An FO sentence is $PCE(k)$ if, and only if, it is equivalent to a $\forall^k\exists^*$ sentence.*

We call Theorem 2(1) the *substructural version* of $\mathsf{GLT}(k)$, and Theorem 2(2) the *extensional version* of $\mathsf{GLT}(k)$. To the best of our knowledge, this characterization is the first to relate natural quantitative properties of models of sentences in a semantic class to counts of leading quantifiers in equivalent $\exists^*\forall^*$ or $\forall^*\exists^*$ sentences.

Moving towards $\exists^k\forall^*$ and $\forall^k\exists^*$ theories, we show that the extensional version of $\mathsf{GLT}(k)$ extends to $\forall^k\exists^*$ theories as well. Intriguingly however, on the substructural front, even $\exists\forall^*$ theories, i.e. theories of Σ_2^0 sentences in which each sentence has exactly one existential quantifier, turn out to be too "powerful" for $PSC(k)$.

Theorem 3. *The following hold over arbitrary structures for each $k \in \mathbb{N}$.*

1. *An FO theory is $PCE(k)$ if, and only if, it is equivalent to a theory of $\forall^k\exists^*$ sentences.*
2. *An FO theory that is $PSC(k)$ is always equivalent to a Σ_2^0 theory. The converse is not true in general: there exists a theory of $\exists\forall^*$ sentences that is not $PSC(l)$ for any $l \in \mathbb{N}$.*

Part (2) of Theorem 3, while showing that a $PSC(k)$ theory is always equivalent to a Σ_2^0 theory, does not tell us anything about the maximum number of existential quantifiers that can appear in any sentence of the Σ_2^0 theory. Analogous to part (1) of Theorem 2, it is natural to ask whether a $PSC(k)$ theory is equivalent to a theory of $\exists^k\forall^*$ sentences. We provide a affirmative answer to this question conditioned on a hypothesis that we describe below, thereby conditionally refining Part (2) of Theorem 3.

To state the hypothesis, we introduce a little bit of terminology. Given a structure $\mathfrak{A}$ and a k-tuple $\bar{a}$ of elements of $\mathfrak{A}$, the Π_1^0 *type of $\bar{a}$ in $\mathfrak{A}$* is the set of all Π_1^0 formulae having k free variables, that are true of $\bar{a}$ in $\mathfrak{A}$. For a $PSC(k)$ theory T and a model $\mathfrak{A}$ of T, let C be a k-crux of T and let $\bar{a}$ be a k-tuple formed out of C. We say that the Π_1^0 type of $\bar{a}$ *determines a k-crux*, if for any structure $\mathfrak{B}$, it is the case that if $\mathfrak{B}$ contains a k-tuple $\bar{b}$ that satisfies in $\mathfrak{B}$, all the formulae in the Π_1^0 type of $\bar{a}$ in $\mathfrak{A}$, then $\mathfrak{B}$ is a model of T and the elements of $\bar{b}$ form a k-crux in $\mathfrak{B}$. (Following the parlance used in the classical model theory literature, a tuple $\bar{b}$ of the kind just mentioned is said to *realize* the Π_1^0 type of $\bar{a}$ in $\mathfrak{A}$). We now make the following hypothesis which we argue is well-motivated and plausible.

Hypothesis 4. *Let T be a $PSC(k)$ theory. Then every model $\mathfrak{A}$ of T contains a k-crux C such that for any k-tuple $\bar{a}$ constructed from C, the Π_1^0 type of $\bar{a}$ in $\mathfrak{A}$ determines a k-crux.*

Theorem 5 (Conditional refinement of Theorem 3(2))**.** *Assuming Hypothesis 4 holds, a $PSC(k)$ theory is always equivalent to a theory of $\exists^k\forall^*$ sentences.*

However from Part (2) of Theorem 3, we know that $PSC(k)$ theories cannot characterize $\exists^k\forall^*$ theories, since the latter subsume $\exists\forall^*$ theories. This question of characterization of $\exists^k\forall^*$ theories remains open.

The above results give new semantic characterizations of the classes of Σ_2^0 and Π_2^0 sentences; the property that says that a sentence is $PSC(k)$, respectively $PCE(k)$, for some k, characterizes

Σ^0_2 and Π^0_2 sentences. For Σ^0_2 and Π^0_2 theories however, the situation is different. It turns out that Π^0_2 theories are strictly more general than $PCE(k)$ theories for each k. That Σ^0_2 theories, why in fact even $\exists\forall^*$ theories, cannot be subsumed by $PSC(k)$ theories for any k, has already been mentioned above. To still get a characterization of Σ^0_2 and Π^0_2 theories by staying within the ambit of the flavour of our preservation properties, we introduce the properties of $PSC(\lambda)$ and $PCE(\lambda)$ for an infinite cardinal λ. The property $PSC(\lambda)$ asserts the existence of a crux of size *less than* λ in any model, while $PCE(\lambda)$ is defined such that it is equivalent to the negation of $PSC(\lambda)$. We show that these properties indeed respectively characterize Σ^0_2 and Π^0_2 theories, thereby giving new characterizations of the latter.

Theorem 6. *The following hold over arbitrary structures.*

1. *For each $\lambda \geq \aleph_1$, an FO theory is $PSC(\lambda)$ if, and only if, it is equivalent to a Σ^0_2 theory.*
2. *For each $\lambda \geq \aleph_0$, an FO theory is $PCE(\lambda)$ if, and only if, it is equivalent to a Π^0_2 theory.*

This completes the description of our results in the classical model theory context. We present various directions for future work, and sketch how natural generalizations of the properties of $PSC(k)$ and $PCE(k)$ can be used to get finer characterizations of Σ^0_n and Π^0_n sentences/theories for $n > 2$, analogous to the finer characterizations of Σ^0_2 and Π^0_2 sentences/theories by $PSC(k)$ and $PCE(k)$.

We conclude this section of the introduction by describing the techniques used in proving our results. For $\mathsf{GLT}(k)$, we first show that this result holds over a special class of structures that are λ-*saturated*, where λ is an infinite cardinal. Then using the fact that for any structure $\mathfrak{A}$, there is always a λ-saturated structure for some suitable λ, that satisfies exactly the same FO sentences as $\mathfrak{A}$, we "transfer" the validity of $\mathsf{GLT}(k)$, from that over λ-saturated structures, to that over all structures. We give also a different proof of $\mathsf{GLT}(k)$ using ascending chains of structures. Similar proofs work for the characterization of $PCE(k)$ and $PCE(\lambda)$ theories.

To show that $PSC(k)$ and $PSC(\lambda)$ theories are equivalent to Σ^0_2 theories, we use Keisler's characterization of Σ^0_2 theories in terms of a preservation property defined in terms of 1-sandwiches, and show that any theory that is $PSC(k)$ or $PSC(\lambda)$ satisfies this preservation property, whereby its equivalence with a Σ^0_2 theory follows. The proof of Theorem 5 is the most involved of all our proofs. It introduces a novel technique of getting a syntactically defined FO theory equivalent to a given FO theory satisfying a semantic property, *by going outside of FO*. Specifically, for the case of $PSC(k)$ theories, under Hypothesis 4, we first "go up" into an *infinitary logic* and show that a $PSC(k)$ theory can be characterized by sentences of this logic. We then "come down" back to FO by providing a translation of sentences of the aforesaid infinitary logic, into their equivalent FO theories, whenever these sentences are known to be equivalent to FO theories. The FO theories are obtained from suitable *finite approximations* of the infinitary sentences, and turn out to be theories of $\exists^k\forall^*$ sentences. The "coming down" process can be seen as a "compilation" process (in the sense of compilers used in computer science) in which a "high level" description – via infinitary sentences that are known to be equivalent to FO theories – is translated into an equivalent "low level" description – via FO theories. We believe this technique of accessing the descriptive power of an infinitary logic followed by accessing the translation power of "compiler results" of the kind just mentioned, may have many applications.

Results in the finite model theory context

While the failure of the Łoś-Tarski theorem in the finite shows that universal sentences cannot capture in the finite, the property of preservation under substructures, we show below a stronger result.

Proposition 7. *There exists a vocabulary τ such that if $\mathcal{S}$ is the class of all finite τ-structures, then for each $k \geq 0$, there exists an FO(τ) sentence ψ_k that is preserved under substructures over $\mathcal{S}$, but that is not equivalent over $\mathcal{S}$ to any $\exists^k\forall^*$ sentence. It follows that there is a sentence that is $PSC(k)$ over $\mathcal{S}$ (ψ_k being one such sentence) but that is not equivalent over $\mathcal{S}$ to any $\exists^k\forall^*$ sentence.*

The above result therefore shows the failure of $\mathsf{GLT}(k)$ over all finite structures, for all $k \geq 0$. Furthermore, as already mentioned earlier, we show $\mathsf{GLT}(k)$ also fails in general for each $k \geq 2$, over the special classes of finite structures that are acyclic, of bounded degree or of bounded tree-width, that were shown in [4] to satisfy the Łoś-Tarski theorem. This is because of the following result. Below, hereditary means closed under induced subgraphs. Also a class $\mathcal{S}$ of directed graphs has bounded induced path lengths, if the class of undirected graphs underlying the graphs of $\mathcal{S}$ has bounded induced path lengths.

Proposition 8. *Let $\mathcal{V}$ be a hereditary class of undirected graphs. Let $\mathcal{S}$ be the class of all directed graphs whose underlying undirected graph belongs to $\mathcal{V}$. If $\mathsf{GLT}(k)$ holds over $\mathcal{S}$, then $\mathcal{S}$ has bounded induced path lengths.*

To “recover” $\mathsf{GLT}(k)$ in the face of the above failures, we define the *$\mathcal{L}$-equivalent bounded substructure property*, denoted $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, k)$, where $\mathcal{L}$ is either FO or MSO, $\mathcal{S}$ is a class of finite structures, and k is a natural number. In the definition below, structures $\mathfrak{B}$ and $\mathfrak{A}$ are said to be *$(m, \mathcal{L})$-similar* if $\mathfrak{B}$ and $\mathfrak{A}$ agree on all $\mathcal{L}$ sentences of quantifier nesting depth m.

Definition 9 ($\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, k)$)**.** Let $\mathcal{S}$ be a class of finite structures, k be a natural number and $\mathcal{L}$ be one of the logics FO or MSO. We say that $\mathcal{S}$ satisfies the *$\mathcal{L}$-equivalent bounded substructure property* for parameter k, abbreviated $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, k)$ *is true*, if given a structure $\mathfrak{A}$ in $\mathcal{S}$, a subset W of at most k elements of $\mathfrak{A}$, and a natural number m, there exists a bounded substructure $\mathfrak{B}$ of $\mathfrak{A}$ containing W, that is in $\mathcal{S}$ and that is $(m, \mathcal{L})$-similar to $\mathfrak{A}$. The bound on the size of $\mathfrak{B}$ depends only on m (if $\mathcal{S}$ and k are fixed). If this bound is computable, then we say $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, k)$ *holds with computable bounds*.

Thus if a structure in $\mathcal{S}$ satisfies a property that can be expressed using an $\mathcal{L}$ sentence, then there is a bounded substructure of it in $\mathcal{S}$ that also satisfies the same property, where the bound depends only on the quantifier nesting depth of the sentence.
A reader familiar with the downward Löwenheim-Skolem theorem will immediately recognize the close resemblance of $\mathcal{L}$-$\mathsf{EBSP}(\cdot, k)$ with the model-theoretic property that the former theorem talks about. The downward Löwenheim-Skolem theorem says that given an arbitrary structure $\mathfrak{A}$ over a countable vocabulary and a countable set W of elements of $\mathfrak{A}$, there exists a countable substructure $\mathfrak{B}$ of $\mathfrak{A}$ containing W, that is “FO-similar” to $\mathfrak{A}$, in that $\mathfrak{B}$ agrees with $\mathfrak{A}$ on all FO sentences. Thus if arbitrary structure satisfies an FO expressible property, then there is a countable substructure of it that also satisfies the same property. Indeed then, $\mathcal{L}$-$\mathsf{EBSP}(\cdot, k)$ can be well regarded as a finitary analogue of the downward Löwenheim-Skolem property.
It turns out that $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, k)$ entails not only $\mathsf{GLT}(k)$ but also the homomorphism preservation theorem (HPT).

Theorem 10. *Let $\mathcal{S}$ be a class of finite structures and $k \in \mathbb{N}$ be such that $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, k)$ holds. Then both $\mathsf{GLT}(k)$ and HPT hold over $\mathcal{S}$. Furthermore, if $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, k)$ holds with computable bounds, then effective versions of $\mathsf{GLT}(k)$ and HPT hold over $\mathcal{S}$.*

We then show that a variety of classes of finite structures, that are of interest in computer science and finite model theory, satisfy $\mathcal{L}$-$\mathsf{EBSP}(\cdot, k)$, and furthermore with computable bounds. The classes that we consider are broadly of two kinds: special kinds of labeled posets and special kinds of graphs. For the case of labeled posets, we have the following result.

Theorem 11. *Given a finite alphabet Σ and a function $\rho : \Sigma \to \mathbb{N}$, let* $\mathsf{Words}(\Sigma)$*,* $\mathsf{Unordered\text{-}trees}(\Sigma)$*,* $\mathsf{Ordered\text{-}trees}(\Sigma)$*,* $\mathsf{Ordered\text{-}ranked\text{-}trees}(\Sigma, \rho)$ *and* $\mathsf{Nested\text{-}words}(\Sigma)$ *denote respectively, the classes of all Σ-words, all unordered Σ-trees, all ordered Σ-trees, all ordered Σ-trees ranked by ρ, and all nested Σ-words. Let $\mathcal{S}$ be a regular subclass of any of these classes. Then $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, k)$ holds with computable bounds for each $k \in \mathbb{N}$.*

While words and trees have had a long history of studies in the literature, nested words are much recent [2], and have attracted a lot of attention as they admit a seamless generalization of the theory of regular languages and are also closely connected with visibly pushdown languages [1]. For the case of graphs, we show the following result.

Theorem 12. *Given $n, k \in \mathbb{N}$, let* $\mathsf{Labeled\text{-}n\text{-}partite\text{-}cographs}(\Sigma)$ *be the class of all Σ-labeled n-partite cographs. Let $\mathcal{S}$ be any subclass of* $\mathsf{Labeled\text{-}n\text{-}partite\text{-}cographs}(\Sigma)$*, that is hereditary over the latter. Then $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, k)$ holds with computable bounds. Consequently, each of the following classes of graphs satisfies $\mathcal{L}$-$\mathsf{EBSP}(\cdot, k)$ with computable bounds for each $k \geq 0$.*

1. *Any hereditary class of n-partite cographs, for each $n \in \mathbb{N}$.*
2. *Any hereditary class of graphs of bounded shrub-depth.*
3. *Any hereditary class of graphs of bounded $\mathcal{SC}$-depth.*
4. *Any hereditary class of graphs of bounded tree-depth.*
5. *Any hereditary class of cographs.*

The class of n-partite cographs, introduced in [17], jointly generalizes the classes of cographs, graph classes of bounded tree-depth, those of bounded shrub-depth and those of bounded $\mathcal{SC}$-depth. The importance of the latter graph classes is that they have various interesting finiteness properties, and have become very prominent in the context of fixed parameter tractability of MSO model checking, and in the context of investigating when FO equals MSO in its expressive power [11, 15, 16, 26]. The notion of tree-depth plays a very important role in the proof of HPT over all finite structures, and the latter theorem was a long standing open problem in finite model theory.

We next give ways to construct new classes of structures satisfying $\mathcal{L}$-$\mathsf{EBSP}(\cdot, \cdot)$ from known ones by showing various closure properties of $\mathcal{L}$-$\mathsf{EBSP}(\cdot, \cdot)$. Firstly, we show that $\mathcal{L}$-$\mathsf{EBSP}(\cdot, \cdot)$ is closed under taking subclasses that are hereditary or $\mathcal{L}$-definable, and is also closed under finite intersections and finite unions. We next consider operations on structures that are implementable using quantifier-free translation schemes; examples of these include disjoint union and various products (cartesian, tensor, strong, lexicographic), and show the following result. Call an n-ary operation "sum-like", respectively "product-like", if it can be implemented using a quantifier-free translation scheme that "acts on" the disjoint union, respectively cartesian product, of the inputs of the operation [28].

Theorem 13. *Let $\mathcal{S}$ be a class of structures. Let O be an n-ary operation and $\mathsf{O}(\mathcal{S})$ be the class of structures obtained by applying O to the structures of $\mathcal{S}$. Then the following hold for each $k \in \mathbb{N}$.*

1. *If MSO-$\mathsf{EBSP}(\mathcal{S}, k)$ is true, then so is MSO-$\mathsf{EBSP}(\mathsf{O}(\mathcal{S}), k)$, whenever O is sum-like.*
2. *If FO-$\mathsf{EBSP}(\mathcal{S}, k \cdot n)$ is true, then so is FO-$\mathsf{EBSP}(\mathsf{O}(\mathcal{S}), k)$, whenever O is product-like.*

In each of the implications above, if the antecedent holds with computable bounds, then so does the consequent.

It follows that finite unions of classes obtained by finite compositions of the aforesaid operations also satisfies $\mathcal{L}$-$\mathsf{EBSP}(\cdot, \cdot)$. However, many interesting classes of structures can be obtained only by taking infinite unions of the kind just described, a notable example being the class of hamming graphs of the n-clique [21]. We show that if the aforementioned infinite unions are "regular" in a sense that we make precise below, then $\mathcal{L}$-$\mathsf{EBSP}(\cdot, 0)$ is preserved under these unions, under reasonable assumptions on the operations. Whereby, the class of hamming graphs of the n-clique satisfies FO-$\mathsf{EBSP}(\cdot, 0)$, as does the class of p-dimensional grid posets, where p belongs to any MSO definable (using a linear order) class of natural numbers (like, even numbers).

Given a finite set Op of sum-like or product-like operations on structures, a finite composition of such operations can be represented as an *operation tree over* Op in which the internal nodes are labeled with operations from Op, the number of children of any internal node equals the arity of the operation labeling the node, and all leaf nodes are labeled with a symbol, say $\diamond$, which is a place-holder for an "input" structure. When the operation tree is "applied" to a class of structures, it produces another class of structures. Generalizing, a class $\mathcal{V}$ of operation trees over Op, applied to a class $\mathcal{S}$ of structures, produces a class $\mathcal{V}(\mathcal{S})$ of structures which is the union of the classes of structures produced by each operation tree in $\mathcal{V}$. If $\mathcal{V}$ is finite, then it follows from Theorem 13 and the closure of $\mathcal{L}$-$\mathsf{EBSP}(\cdot, \cdot)$ under finite unions, that $\mathcal{L}$-$\mathsf{EBSP}(\cdot, \cdot)$ is also closed under $\mathcal{V}$ (for appropriate $\mathcal{L}$). However, important classes of structures like the aforementioned class of hamming graphs of the n-clique can be produced only by considering infinite $\mathcal{V}$. Our result below shows the closure of $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, 0)$ under infinite $\mathcal{V}$ provided that $\mathcal{V}$, seen as a language of trees, is regular, and that the operations of Op are *monotone* (inputs embed into the output) and $\equiv_{m,\mathcal{L}}$-*preserving* ($(m, \mathcal{L})$-similar inputs produce $(m, \mathcal{L})$-similar outputs).

Theorem 14. *Let* Op *be a finite set of sum-like or product-like operations, where each operation in* Op *is monotone and* $\equiv_{m,\mathcal{L}}$*-preserving. Let* $\mathcal{V}$ *be a class of operation trees over* Op*, that is regular. Let* $\mathcal{S}$ *be a class of structures. If* $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, 0)$ *is true, then so is* $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{V}(\mathcal{S}), 0)$*. Further, if* $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, 0)$ *holds with computable bounds, then so does* $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{V}(\mathcal{S}), 0)$*.*

The proofs of the above results rely on tree-representations of structures, and proceed by performing appropriate "prunings" of, and "graftings" within, these trees, while preserving the substructure and "$(m, \mathcal{L})$-similarity" relations between the structures represented by these trees. The process eventually yields small subtrees that represent bounded $(m, \mathcal{L})$-equivalent substructures of the original structure. The aforementioned prunings and graftings make use of *composition lemmas* of the Feferman-Vaught kind, or composition-like lemmas, in all the cases [28].

Finally, we present two additional observations about the $\mathcal{L}$-$\mathsf{EBSP}(\cdot, k)$ property. Firstly, since $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, k)$ entails the "small model property" for $\mathcal{L}$ over $\mathcal{S}$, the satisfiability problem for $\mathcal{L}$ is decidable over $\mathcal{S}$. Secondly, we show the following surprising connection between well-quasi-ordering (w.q.o.) under embedding and $\mathcal{L}$-$\mathsf{EBSP}(\cdot, 0)$.

Theorem 15. *If* $\mathcal{S}$ *is w.q.o. under embedding, then* $\mathcal{L}$-$\mathsf{EBSP}(\mathcal{S}, 0)$ *is true.*

The notion of w.q.o. is very important in the literature. For instance, one of the many features of a w.q.o. class is that checking membership in any hereditary subclass can be done efficiently (i.e. in polynomial time). Theorem 15 gives a technique to show the $\mathcal{L}$-$\mathsf{EBSP}(\cdot, 0)$ property for a class, namely by showing that the class is w.q.o. under embedding. And flipped around, it also gives a "logic-based" tool to show that a class of structures is not w.q.o. under embedding, namely by showing that the class does not satisfy $\mathcal{L}$-$\mathsf{EBSP}(\cdot, 0)$.

We conclude by presenting various directions for future work, notable amongst these being two questions, one concerning $\mathcal{L}$-$\mathsf{EBSP}(\cdot, k)$, and the other concerning a variant of $\mathsf{GLT}(k)$. The former asks for an investigation of a structural characterization of $\mathcal{L}$-$\mathsf{EBSP}(\cdot, k)$, motivated by the observation that any hereditary class of graphs satisfying $\mathcal{L}$-$\mathsf{EBSP}(\cdot, k)$ has bounded induced path lengths. The latter asks for an investigation of whether the property $\bigcup_{k \geq 0} PSC(k)$ characterizes Σ^0_2 over all finite structures. Indeed, the failure of $\mathsf{GLT}(k)$ over all finite structures does not rule out this possibility. Since $\bigcup_{k \geq 0} PSC(k)$ characterizes Σ^0_2 over arbitrary structures, the truth of this characterization over all finite structures would give a new preservation theorem, that is different from HPT, and that holds in the finite.

Conclusion

In summary, the properties introduced in this thesis are interesting in both the classical and finite model theory contexts, and yield in both these contexts, a new and natural generalization of the classical Łoś-Tarski preservation theorem.

List of Publications

I. Classical model theory

1. Abhisekh Sankaran, Bharat Adsul, and Supratik Chakraborty. *A generalization of the Łoś-Tarski preservation theorem*. Annals of Pure and Applied Logic, 167(3):189 - 210, 2016.

II. Finite model theory

1. Abhisekh Sankaran, Bharat Adsul, and Supratik Chakraborty. *A generalization of the Łoś-Tarski preservation theorem over classes of finite structures*. In Mathematical Foundations of Computer Science 2014 - 39th International Symposium, MFCS 2014, Budapest, Hungary, August 25-29, 2014. Proceedings, Part I, pages 474 - 485, 2014.

2. A. Sankaran, B. Adsul, V. Madan, P. Kamath, and S. Chakraborty. *Preservation under substructures modulo bounded cores*. In proceedings of WoLLIC, pages 291 - 305, 2012.

3. Abhisekh Sankaran. *A finitary analogue of the downward Löwenheim-Skolem property*. submitted to ICALP 2016.

Bibliography

[1] Rajeev Alur and P. Madhusudan. Visibly pushdown languages. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 202–211, 2004. 11

[2] Rajeev Alur and P. Madhusudan. Adding nesting structure to words. *J. ACM*, 56(3), 2009. 11

[3] A. Atserias, A. Dawar, and P. G. Kolaitis. On preservation under homomorphisms and unions of conjunctive queries. *J. ACM*, 53(2):208–237, 2006. 4

[4] A. Atserias, A. Dawar, and M. Grohe. Preservation under extensions on well-behaved finite structures. *SIAM J. Comput.*, 38(4):1364–1381, 2008. 4, 9

[5] Egon Börger, Erich Grädel, and Yuri Gurevich. *The Classical Decision Problem*. Perspectives in Mathematical Logic. Springer, 1997. 3

[6] Andrea Calì, Georg Gottlob, and Andreas Pieris. New expressive languages for ontological query answering. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011*, 2011. 3

[7] C. C. Chang and H. J. Keisler. *Model Theory*. Elsevier Science Publishers, 3 edition, 1990. 1, 2

[8] Anuj Dawar. Homomorphism preservation on quasi-wide classes. *J. Comput. Syst. Sci.*, 76(5):324–332, 2010. 4

[9] Rodney G. Downey and Michael R. Fellows. *Fundamentals of Parameterized Complexity*. Texts in Computer Science. Springer, 2013. ISBN 978-1-4471-5558-4. 3, 4

[10] D. Duris. Extension preservation theorems on classes of acyclic finite structures. *SIAM J. Comput.*, 39(8):3670–3681, 2010. 4

[11] Michael Elberfeld, Martin Grohe, and Till Tantau. Where first-order and monadic second-order logic coincide. In *Proceedings of the 27th Annual IEEE Symposium on Logic in Computer Science, LICS 2012, Dubrovnik, Croatia, June 25-28, 2012*, pages 265–274, 2012. 11

[12] M. Emmer, Z. Khasidashvili, K. Korovin, and A. Voronkov. Encoding industrial hardware verification problems into effectively propositional logic. In *Proc. of FMCAD*, pages 137–144, 2010. 3

[13] Ronald Fagin. Generalized first-order spectra and polynomial-time recognizable sets. *Complexity of Computation*, pages 27–41, 1974. 3

[14] Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: semantics and query answering. *Theor. Comput. Sci.*, 336(1):89–124, 2005. 3

[15] J. Gajarsky and P. Hlinený. Kernelizing MSO properties of trees of fixed height, and some consequences. *LMCS*, 11(19):1–26, 2015. 11

[16] Jakub Gajarský and Petr Hlinený. Faster deciding MSO properties of trees of fixed height, and some consequences. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2012, December 15-17, 2012, Hyderabad, India*, pages 112–123, 2012. 11

[17] Robert Ganian, Petr Hlinený, Jaroslav Nesetril, Jan Obdrzálek, Patrice Ossona de Mendez, and Reshma Ramadurai. When trees grow low: Shrubs and fast MSO1. In *Mathematical Foundations of Computer Science 2012 - 37th International Symposium, MFCS 2012, Bratislava, Slovakia, August 27-31, 2012. Proceedings*, pages 419–430, 2012. 11

[18] Sumit Gulwani. Dimensions in program synthesis. In *Proc. of PPDP*, pages 13–24, 2010. 3

[19] Frederik Harwath, Lucas Heimberg, and Nicole Schweikardt. Preservation and decomposition theorems for bounded degree structures. In *Joint Meeting of the Twenty-Third*

EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), CSL-LICS '14, Vienna, Austria, July 14 - 18, 2014, pages 49:1–49:10, 2014. 4

[20] Wilfrid Hodges. *A Shorter Model Theory*. Cambridge University Press, 1997. 1

[21] Wilfried Imrich and Sandi Klavzar. Recognizing hamming graphs in linear time and space. *Inf. Process. Lett.*, 63(2):91–95, 1997. 12

[22] Mathew Joseph, Gabriel M. Kuper, and Luciano Serafini. Query answering over contextualized RDF knowledge with forall-existential bridge rules: Attaining decidability using acyclicity. In *Proceedings of the 29th Italian Conference on Computational Logic, Torino, Italy, June 16-18, 2014.*, pages 210–224, 2014. 3

[23] Mathew Joseph, Gabriel M. Kuper, Till Mossakowski, and Luciano Serafini. Query answering over contextualized RDF/OWL knowledge with forall-existential bridge rules: Decidable finite extension classes. *Semantic Web*, 7(1):25–61, 2016. 3

[24] H. J. Keisler. Theory of models with generalized atomic formulas. *Journal of Symbolic Logic*, 25:1–26, 1960. 2

[25] Markus Krötzsch and Sebastian Rudolph. Extending decidable existential rules by joining acyclicity and guardedness. In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 963–968, 2011. 3

[26] Michael Lampis. Model checking lower bounds for simple graphs. *Logical Methods in Computer Science*, 10(1), 2014. 11

[27] Maurizio Lenzerini. Data integration: A theoretical perspective. In *Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '02, pages 233–246. ACM, 2002. ISBN 1-58113-507-6. 3

[28] J. A. Makowsky. Algorithmic uses of the Feferman-Vaught theorem. *Ann. Pure Appl. Logic*, 126(1-3):159–213, 2004. 12, 13

[29] Ruzica Piskac, Leonardo Mendonça de Moura, and Nikolaj Bjørner. Deciding effectively propositional logic using DPLL and substitution sets. *J. Autom. Reasoning*, 44(4):401–424, 2010. 3

[30] F. P. Ramsey. On a problem of formal logic. *Proc. of The London Mathematical Society*, s2-30:264–286, 1930. 3

[31] B. Rossman. Homomorphism preservation theorems. *J. ACM*, 55(3), 2008. 4

[32] Saurabh Srivastava and Sumit Gulwani. Program verification using templates over predicate abstraction. In *Proc. of PLDI*, pages 223–234, 2009. 3

[33] W. W. Tait. A counterexample to a conjecture of Scott and Suppes. *Journal of Symbolic Logic*, 24(1):15–16, 1959. 4