Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

First-Order Rewritability of Atomic Queries in Horn Description Logics

Meghyn Bienvenu
CNRS & Université Paris Sud
Orsay, France

Carsten Lutz
Fachbereich Informatik
Universität Bremen, Germany

Frank Wolter
Department of Computer Science
University of Liverpool, UK

Abstract

One of the most advanced approaches to querying data in the presence of ontologies is to make use of relational database systems, rewriting the original query and the ontology into a new query that is formulated in SQL or, equivalently, in first-order logic (FO). For ontologies written in many standard description logics (DLs), however, such FO-rewritings are not guaranteed to exist. We study FO-rewritings and their existence for a basic class of queries and for ontologies formulated in Horn DLs such as Horn-$\mathcal{SHI}$ and $\mathcal{EL}$. Our results include characterizations of the existence of FO-rewritings, tight complexity bounds for deciding whether an FO-rewriting exists (EXPTIME and PSPACE), and tight bounds on the (worst-case) size of FO-rewritings, when presented as a union of conjunctive queries.

1 Introduction

A prominent application of description logic (DL) ontologies is to facilitate access to data. Specifically, the ontology serves to assign a semantics to the relation symbols used in the data; it can also provide additional relation symbols that, although not explicitly occurring in the data, can be used in the query. Several approaches to querying data in the presence of ontologies utilize relational databases systems (RDBMSs), aiming to exploit their mature technology, advanced optimization techniques, and the general infrastructure that those systems offer. One of the most popular such approaches is to rewrite the original query and the DL ontology into an SQL query that is passed to the RDBMS for execution [Calvanese *et al.*, 2007; Pérez-Urbina *et al.*, 2009; Chortaras *et al.*, 2011; Gottlob *et al.*, 2011]. Based on the equivalence of first-order (FO) formulas and SQL queries, we call the rewritten query an *FO-rewriting*.

The FO-rewriting approach to ontology-based data access comes with its own set of DLs specifically designed for this purpose, the so-called DL-Lite family. To guarantee that an FO-rewriting of queries and ontologies always exists, DLs from this family are significantly restricted in expressive power. Unfortunately, this is not acceptable for all applications, in particular when the ontology is used to model the application domain in a detailed way instead of providing only more abstract, database-style constraints. Given that very expressive DLs of the $\mathcal{ALC}$ and $\mathcal{SHIQ}$ families do not admit tractable query answering (regarding data complexity), a good compromise between expressive power and computational complexity is provided by so-called Horn DLs such as $\mathcal{EL}$, $\mathcal{ELI}$, and Horn-$\mathcal{SHIQ}$. Already in the basic DLs of this kind, such as $\mathcal{EL}$, and for the simplest kind of queries, known as atomic queries (AQs) or instance queries, FO-rewritings are not guaranteed to always exist. To address this problem, other approaches to utilize RDBMSs and related database technology have been brought forward, including the combined approach [Lutz *et al.*, 2009] and rewritings into datalog [Pérez-Urbina *et al.*, 2009; Eiter *et al.*, 2012].

Depending on the application, however, there can still be good reasons to use the FO-rewriting approach for Horn DLs. First, an important feature of this approach is that it allows the ontology to be added on top of the query interface without any modifications to the underlying database. By contrast, the combined approach involves a data completion step, and thus can only be used if data manipulations are permitted. Second, a rewriting into FO queries rather than datalog programs means that one can exploit the comparatively more advanced optimization techniques available for SQL queries. For both reasons, when FO-rewritings happen to exist for the relevant queries and ontologies, the FO-rewriting approach might be very appropriate.

In this paper, we consider ontologies formulated in the Horn DLs $\mathcal{EL}$, $\mathcal{ELI}_\bot$, and Horn-$\mathcal{SHI}$. Notably, $\mathcal{EL}$ forms the basis of the OWL EL fragment of the OWL 2 web ontology language and is popular as a basic language for large-scale ontologies [Baader *et al.*, 2005]. $\mathcal{ELI}_\bot$ can be viewed as the smallest DL that contains as a fragment both $\mathcal{EL}$ and the core version of DL-Lite, and Horn-$\mathcal{SHI}$ is a generalization of $\mathcal{ELI}_\bot$ inspired by the well-known DL Horn-$\mathcal{SHIQ}$, but in contrast to the latter does not admit number restrictions [Hustadt *et al.*, 2007]. As an example for why the rewriting approach fails for these DLs, consider the AQ $A(x)$ and the $\mathcal{EL}$ ontology $\mathcal{T} = \{\exists r.A \sqsubseteq A\}$. The query $A(x)$ cannot be rewritten into an FO-query in the presence of $\mathcal{T}$, intuitively because $\mathcal{T}$ forces the concept name A to be *propagated unboundedly* along r-chains in the data and thus the rewritten query would have to express transitive closure of r. Of course, such an isolated example does not rule out the possibility that for some AQs and some $\mathcal{EL}$ (or Horn-$\mathcal{SHI}$) ontologies, including those that are used in applications, FO-rewritings do exist. For example, $A(x)$ is FO-rewritable relative to the $\mathcal{EL}$ ontology $\mathcal{T}' = \{A \sqsubseteq \exists r.A\}$, which has a lot of similarity with the aforementioned ontology $\mathcal{T}$. In fact, $\mathcal{T}'$ can simply be ignored when answering $A(x)$ without losing any answers. Inspired by these observations, the aim of this paper is to study FO-rewritings of AQs in the presence of ontologies formulated in $\mathcal{EL}$, $\mathcal{ELI}$, $\mathcal{ELI}_\bot$, and Horn-$\mathcal{SHI}$.

We primarily study the problem to decide, given an atomic query (AQ) q, an ontology $\mathcal{T}$, and a finite set Σ of symbols that are allowed to be used in the data (ABox), whether q is FO-rewritable relative to $\mathcal{T}$ over Σ-ABoxes. Note that the restriction of the data signature is natural in many applications of ontology-based data access, cf. [Baader *et al.*, 2010; Bienvenu *et al.*, 2012b]. We show that this problem is ExpTime-complete for ontologies formulated in Horn-$\mathcal{SHI}$, where the lower bound applies even to $\mathcal{ELI}$ ontologies and when the ABox signature is the full signature (that is, it must contain all concept and role names) rather than being an input. For ontologies formulated in $\mathcal{EL}$, the problem remains ExpTime-complete when the ABox signature is an input (though the lower bound is more difficult to establish), but is only PSpace-complete when the ABox signature is full.

Our analysis also yields characterizations of the existence of FO-rewritings in terms of the existence of certain tree-shaped ABoxes, which are interesting in their own right. Surprisingly, tree-shaped ABoxes can even be replaced with linear ABoxes (single role chains decorated with concept assertions) when the ontology is formulated in $\mathcal{EL}$ and the ABox signature is full. Our proofs also yield a way to effectively construct FO-rewritings when they exist. We use this observation to analyze the size of FO-rewritings, showing that they can always be represented by a union of conjunctive queries (UCQ) of at most triple exponential size, and that this bound is essentially optimal: there are families of AQs and $\mathcal{EL}$ ontologies for which FO-rewritings exist, but such that every presentation of the rewritings as a UCQ is necessarily triple-exponential in size.

Some proof details are deferred to the appendix of the long version, http://www.informatik.uni-bremen.de/∼clu/papers/.

2 Preliminaries

Let $\mathsf{N_C}$ and $\mathsf{N_R}$ be disjoint and countably infinite sets of *concept* and *role names*. A *role* is a role name r or an *inverse role* r^-, with r a role name. A *Horn-$\mathcal{SHI}$ concept inclusion (CI)* is of the form $L \sqsubseteq R$, where L and R are concepts defined by the syntax rules

$$R, R' ::= \top \mid \bot \mid A \mid \neg A \mid R \sqcap R' \mid \neg L \sqcup R \mid \exists r.R \mid \forall r.R$$
$$L, L' ::= \top \mid \bot \mid A \mid L \sqcap L' \mid L \sqcup L' \mid \exists r.L$$

with A ranging over concept names and r over roles. In DLs, ontologies are formalized as TBoxes. A *Horn-$\mathcal{SHI}$ TBox* $\mathcal{T}$ is a finite set of Horn-$\mathcal{SHI}$ CIs, *transitivity assertions* $\mathsf{trans}(r)$, and *role inclusions (RI)* $r \sqsubseteq s$, with r and s roles. Note that different definitions of Horn-$\mathcal{SHI}$ can be found in the literature [Hustadt *et al.*, 2007; Eiter *et al.*, 2008; Kazakov, 2009]. As the original definition from [Hustadt *et al.*, 2007] based on polarity is rather technical, we prefer the above (equivalent) definition.

An $\mathcal{ELI}_\bot$ *TBox* is a finite set of inclusions of the form $L \sqsubseteq L'$ where L, L' are constructed by the rule above, but without using disjunction. An $\mathcal{ELI}_\bot$ TBox that does not use the $\bot$ concept is an $\mathcal{ELI}$ *TBox*, and an $\mathcal{ELI}$ TBox that does not use inverse roles is an $\mathcal{EL}$ *TBox*.

An *ABox* is a finite set of *concept assertions* $A(a)$ and *role assertions* $r(a,b)$ where A is a concept name, r a role name, and a, b individual names from a countably infinite set $\mathsf{N_I}$. We sometimes write $r^-(a,b)$ instead of $r(b,a)$ and use $\mathsf{Ind}(\mathcal{A})$ to denote the set of all individual names used in $\mathcal{A}$.

The semantics of DLs is given in terms of *interpretations* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I}$ is a non-empty set (the *domain*) and $\cdot^\mathcal{I}$ is the *interpretation function*, assigning to each $A \in \mathsf{N_C}$ a set $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, to each $r \in \mathsf{N_R}$ a relation $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$, and to each $a \in \mathsf{N_I}$ an element $a^\mathcal{I} \in \Delta^\mathcal{I}$ such that $a_1^\mathcal{I} \neq a_2^\mathcal{I}$ whenever $a_1 \neq a_2$ (the so-called unique name assumption). The interpretation $C^\mathcal{I} \subseteq \Delta^\mathcal{I}$ of a concept C in $\mathcal{I}$ is defined as usual, see [Baader *et al.*, 2003]. An interpretation $\mathcal{I}$ *satisfies* a CI $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$, a transitivity assertion $\mathsf{trans}(r)$ if $r^\mathcal{I}$ is transitive, an RI $r \sqsubseteq s$ if $r^\mathcal{I} \subseteq s^\mathcal{I}$, a concept assertion $A(a)$ if $a^\mathcal{I} \in A^\mathcal{I}$, and a role assertion $r(a,b)$ if $(a^\mathcal{I}, b^\mathcal{I}) \in r^\mathcal{I}$. We say that $\mathcal{I}$ is a *model of* a TBox or an ABox if it satisfies all inclusions and assertions in it. An ABox $\mathcal{A}$ is *consistent* w.r.t. a TBox $\mathcal{T}$ if $\mathcal{A}$ and $\mathcal{T}$ have a common model.

An *atomic query (AQ)* takes the form $A(x)$, with A a concept name and x a variable. We write $\mathcal{A}, \mathcal{T} \models A(a)$ if $a^\mathcal{I} \in A^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$. If $\mathcal{A}, \mathcal{T} \models A(a)$ and $a \in \mathsf{Ind}(\mathcal{A})$, then a is a *certain answer* to $A(x)$ given $\mathcal{A}$ and $\mathcal{T}$. We use $\mathsf{cert}_\mathcal{T}(A(x), \mathcal{A})$ to denote the set of all certain answers to $A(x)$ given $\mathcal{A}$ and $\mathcal{T}$. A first-order query (FOQ), is a first-order formula φ constructed from atoms $A(x)$, $r(x,y)$, and $x = y$; here, concept names are viewed as unary predicates, role names as binary predicates, and predicates of other arity, function symbols, and constant symbols are not permitted. As usual, we write $\varphi(\vec{x})$ to indicate that the free variables of φ are among $\vec{x}$ and call $\vec{x}$ the *answer variables* of φ. The number of answer variables is the *arity* of φ and a FOQ φ is *Boolean* if it has arity zero. We use $\mathsf{ans}(\mathcal{I}, \varphi)$ to denote the set of all answers to the FOQ φ in the interpretation $\mathcal{I}$; that is, if φ is n-ary, then $\mathsf{ans}(\mathcal{I}, \varphi)$ contains all tuples $\vec{d} \in (\Delta^\mathcal{I})^n$ such that the FO-sentence $\varphi[\vec{d}]$ is satisfied in $\mathcal{I}$ (written $\mathcal{I} \models \varphi[\vec{d}]$). To bridge the gap between certain answers and "normal" answers, we sometimes view an ABox $\mathcal{A}$ as an interpretation $\mathcal{I}_\mathcal{A}$, defined in the obvious way; see [Lutz and Wolter, 2012].

A *signature* is a set of concept and role names, which are uniformly called *symbols* in this context. We use $\mathsf{sig}(\mathcal{T})$ to denote the set of symbols used in the TBox $\mathcal{T}$. A Σ-*ABox* is an ABox that uses only concept and role names from Σ. We speak of an *ABox signature* if the purpose of the signature is to fix the symbols permitted in ABoxes.

Definition 1 (FO-rewriting). *Let $\mathcal{T}$ be a TBox and Σ an ABox signature. A FOQ $\varphi(x)$ is an* FO-rewriting of an AQ $A(x)$ relative to $\mathcal{T}$ and Σ *if* $\mathsf{cert}_\mathcal{T}(A(x), \mathcal{A}) = \mathsf{ans}(\mathcal{I}_\mathcal{A}, \varphi)$

for all Σ-ABoxes $\mathcal{A}$. If there is such a $\varphi(x)$, then $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ.

Thus, FO-rewritings reduce the computation of certain answers (which is a form of deduction) to standard query answering on structures (which is a form of model checking).

Example 2. *Recall from the introduction that $A(x)$ is not FO-rewritable relative to $\mathcal{T} = \{\exists r.A \sqsubseteq A\}$ and the signature $\Sigma = \{r, A\}$. If we add $\exists r.\top \sqsubseteq A$ to $\mathcal{T}$, then $A(x)$ is FO-rewritable relative to the resulting TBox and Σ, and $\varphi(x) = A(x) \vee \exists y\, r(x, y)$ is an FO-rewriting. If we choose $\Sigma = \{A\}$, then $A(x)$ becomes FO-rewritable also relative to the original $\mathcal{T}$, with the trivial FO-rewriting $A(x)$.*

In some applications, the signature of the ABox is not restricted at all and thus, in principle, infinite. However, FO-rewritability of an AQ $A(x)$ relative to $\mathcal{T}$ and any (potentially infinite) signature $\Sigma \subseteq \mathsf{N_C} \cup \mathsf{N_R}$ coincides with FO-rewritability of $A(x)$ relative to $\mathcal{T}$ and the (finite) ABox signature $\mathsf{sig}(\mathcal{T}) \cap \Sigma$. In fact, any FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ is trivially also an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and $\mathsf{sig}(\mathcal{T}) \cap \Sigma$, and when $\varphi(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and $\mathsf{sig}(\mathcal{T}) \cap \Sigma$, then (i) $\varphi(x)$ is also an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ if $A \in \mathsf{sig}(\mathcal{T})$ and (ii) $\varphi(x) \vee A(x)$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ otherwise. Consequently, we from now on restrict our attention to ABox signatures Σ with $\Sigma \subseteq \mathsf{sig}(\mathcal{T})$ and speak of the *full signature* Σ when $\Sigma = \mathsf{sig}(\mathcal{T})$.

Atomic queries are closely related to queries of the more general form $C(x)$ with C an $\mathcal{EL}$ concept or an $\mathcal{ELI}$ concept. Note that such queries can be viewed as tree-shaped conjunctive queries where the root is the only answer variable. The results presented in this paper also capture these more general queries since $\varphi(x)$ is an FO-rewriting of $C(x)$ relative to $\mathcal{T}$ and an ABox signature Σ iff it is an FO-rewriting of $A(x)$ relative to $\mathcal{T} \cup \{A \equiv C\}$ and Σ, with A a fresh concept name.

The reasoning problem studied in this paper is as follows: given an AQ $A(x)$, a TBox $\mathcal{T}$, and an ABox signature Σ (with $\Sigma \subseteq \mathsf{sig}(\mathcal{T})$), decide whether $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ and if this is the case, produce an FO-rewriting. We obtain different versions of this problem by varying the language in which the TBox $\mathcal{T}$ can be formulated, and by admitting a finite ABox signature Σ as input or fixing it to be the full signature.

We also consider restricted forms of FOQs for rewriting atomic queries, and restricted kinds of ABoxes. A FOQ is a *conjunctive query (CQ)* if it has the form $\exists \vec{y} \varphi(\vec{x}, \vec{y})$ with φ a conjunction of atoms; it is a *union of conjunctive queries (UCQ)* if it is a disjunction of CQs. For simplicity, we disallow equality in CQs and UCQs. If an FO-rewriting is a UCQ, we speak of a *UCQ-rewriting*. With every ABox $\mathcal{A}$, we associate the undirected graph $G_\mathcal{A}$ with nodes $\mathsf{Ind}(\mathcal{A})$ and edges $\{\{a, b\} \mid r(a, b) \in \mathcal{A} \text{ or } r(b, a) \in \mathcal{A}\}$. An ABox $\mathcal{A}$ is *acyclic* if the corresponding graph $G_\mathcal{A}$ is acyclic and $r(a, b) \in \mathcal{A}$ implies that (i) $s(a, b) \notin \mathcal{A}$ for all $s \neq r$ and (ii) $s(b, a) \notin \mathcal{A}$ for all role names s; $\mathcal{A}$ is *tree-shaped* if it is acyclic and $G_\mathcal{A}$ is connected. In tree-shaped ABoxes $\mathcal{A}$, we often distinguish one individual $\rho_\mathcal{A} \in \mathsf{Ind}(\mathcal{A})$ as the *root* of $\mathcal{A}$.

We often identify a CQ q with the set of its atoms and regard q as an ABox whose individual names are the variables of q. Then q is called *acyclic* or *tree-shaped* if the ABox corresponding to q has the same property.

3 FO-rewritability in Horn-$\mathcal{SHI}$ and $\mathcal{ELI}_\bot$

We show that deciding FO-rewritability in Horn-$\mathcal{SHI}$ and $\mathcal{ELI}_\bot$ is ExpTime-complete. To achieve this, we provide a characterization of FO-rewritability in terms of the existence of certain ABoxes, which is of independent interest. We also show how to compute FO-rewritings if they exist and give upper and lower bounds on their size.

We start by observing that it suffices to concentrate on TBoxes that are formulated in $\mathcal{ELI}_\bot$ and in *normal form*, that is, all CIs are of one of the forms

$$A \sqsubseteq \bot \quad A \sqsubseteq \exists r.B \quad \top \sqsubseteq A \quad B_1 \sqcap B_2 \sqsubseteq A \quad \exists r.B \sqsubseteq A$$

with A, B, B_1, B_2 concept names and r a role.

Theorem 3. *For every Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$ and ABox signature Σ, one can construct in polynomial time an $\mathcal{ELI}_\bot$ TBox $\mathcal{T}'$ such that for all AQs $A(x)$ with $A \notin \mathsf{sig}(\mathcal{T}') \setminus \mathsf{sig}(\mathcal{T})$, every FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}'$ and Σ, and vice versa.*

The proof of Theorem 3 is similar to reductions in [Hustadt *et al.*, 2007; Kazakov, 2009]. For the elimination of value restrictions $\forall r.B$, observe that the CI $A \sqsubseteq \forall r.B$ is logically equivalent to the CI $\exists r^-.A \sqsubseteq B$. In the following, we will generally work with $\mathcal{ELI}_\bot$ TBoxes and assume normal form whenever this is more convenient.

The following is a direct consequence of a result by Rossman and the fact that the class of finite pointed structures $(\mathcal{I}_\mathcal{A}, a)$, with $\mathcal{A}, \mathcal{T} \models A(a)$, is preserved under homomorphisms [Rossman, 2008].

Proposition 4. *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox and Σ an ABox signature. If an AQ $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ, then there is a UCQ-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ.*

Next, we observe that ABox inconsistency plays a central role for FO-rewritability of AQs relative to $\mathcal{ELI}_\bot$ TBoxes.

Example 5. *Let $\mathcal{T} = \{\exists r.A \sqsubseteq A, A \sqcap B \sqsubseteq \bot\}$. Then $B(x)$ is not FO-rewritable since $\mathcal{A}, \mathcal{T} \models B(a)$ whenever $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$, which is the case iff in $\mathcal{A}$, there are individuals a and b such that $A(a) \in \mathcal{A}$, a is reachable from b on an r-path, and $B(b) \in \mathcal{A}$. Clearly, this condition cannot be expressed by an FO-formula. If we only admit ABoxes that are consistent w.r.t. $\mathcal{T}$, then $B(x)$ is trivially rewritable (it is a rewriting itself).*

To make precise the interplay between FO-rewritability of AQs and of ABox inconsistency, we require some further notions. We say that *ABox inconsistency is FO-rewritable relative to $\mathcal{T}$ and Σ* if there is a FOQ $\varphi()$ such that for every Σ-ABox $\mathcal{A}$, $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$ iff $\mathcal{I}_\mathcal{A} \models \varphi()$. We call an AQ $A(x)$ *FO-rewritable relative to $\mathcal{T}$ and consistent Σ-ABoxes* if there exists a FOQ $\varphi(x)$ such that $\mathsf{cert}_\mathcal{T}(q, \mathcal{A}) = \mathsf{ans}_{\mathcal{I}_\mathcal{A}}(\varphi)$ for all Σ-ABoxes $\mathcal{A}$ that are consistent w.r.t. $\mathcal{T}$. Finally, we an AQ $A(x)$ is *Σ-trivial relative to $\mathcal{T}$* if $\mathcal{A}, \mathcal{T} \models A(a)$ for all Σ-ABoxes $\mathcal{A}$ and $a \in \mathsf{Ind}(\mathcal{A})$. Note that Σ-trivial AQs are FO-rewritable relative to Σ (with $x = x$ a rewriting) and that $A(x)$ is Σ-trivial relative to $\mathcal{T}$ iff $\mathcal{T} \models C \sqsubseteq A$ for all concept names $C \in \Sigma$ and all C of the form $\exists r.\top$ and $\exists r^-.\top$ with $r \in \Sigma$. Thus, it is straightforward to check Σ-triviality of AQs.

Proposition 6. *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox, Σ an ABox signature, and $A(x)$ an AQ that is not Σ-trivial relative to $\mathcal{T}$. Then $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and Σ iff*

1. *$A(x)$ is FO-rewritable relative to $\mathcal{T}$ and consistent Σ-ABoxes, and*
2. *ABox inconsistency is FO-rewritable relative to $\mathcal{T}$, Σ.*

Proof. Assume first that Points 1 and 2 hold. Let $\varphi_1()$ be an FO-rewriting of ABox inconsistency relative to $\mathcal{T}$ and Σ, and let $\varphi_2(x)$ be an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and consistent Σ-ABoxes. Then $(\varphi_1 \wedge x = x) \vee \varphi_2$ is an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ.

Conversely, assume that there is an FO-rewriting $\varphi(x)$ of $A(x)$ relative to $\mathcal{T}$ and Σ, and that $A(x)$ is not Σ-trivial relative to $\mathcal{T}$. Point 1 is trivial since $\varphi(x)$ is an FO-rewriting of $A(x)$ also relative to $\mathcal{T}$ and consistent Σ-ABoxes. For Point 2, Proposition 4 implies that we may assume $\varphi(x)$ to be a UCQ $q_1 \vee \cdots \vee q_n$. Let $\psi()$ be the union of all Boolean CQs p such that

(i) $p \subseteq q_i$ (p is a subset of q_i) for some $1 \leq i \leq n$;

(ii) for all Σ-ABoxes $\mathcal{A}$: if $\mathcal{I}_\mathcal{A} \models p$, then $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$.

We show that $\psi()$ is an FO-rewriting of ABox inconsistency relative to $\mathcal{T}$ and Σ. By Point (ii), $\mathcal{I}_\mathcal{A} \models \psi()$ implies that $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$. Conversely, assume that $\mathcal{A}$ is a Σ-ABox that is inconsistent w.r.t. $\mathcal{T}$. Since $A(x)$ is not Σ-trivial relative to $\mathcal{T}$, there is a Σ-ABox $\mathcal{A}_0$ and $a_0 \in \mathsf{Ind}(\mathcal{A}_0)$ such that $\mathcal{A}_0, \mathcal{T} \not\models A(a_0)$. Let $\mathsf{Ind}(\mathcal{A}) \cap \mathsf{Ind}(\mathcal{A}_0) = \emptyset$. We have $\mathcal{A}_0 \cup \mathcal{A}, \mathcal{T} \models A(a_0)$ since $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$. Hence $\mathcal{I}_{\mathcal{A} \cup \mathcal{A}_0} \models \varphi[a_0]$ and there is a q_i such that $\mathcal{I}_{\mathcal{A} \cup \mathcal{A}_0} \models q_i(a_0)$. Let π be a match of q_i in $\mathcal{I}_{\mathcal{A} \cup \mathcal{A}_0}$ and let p be the Boolean CQ that consists of all atoms in q_i whose variables are mapped by π to $\mathsf{Ind}(\mathcal{A})$. It is readily checked that p is a disjunct of $\psi()$ and so $\mathcal{I}_\mathcal{A} \models \psi()$, as required. □

Proposition 6 suggests a decomposition of the test for FO-rewritability of AQs: first check FO-rewritability of ABox inconsistency and then check FO-rewritability of the AQ relative to consistent ABoxes. In the following, we pursue this approach.

We now develop characterizations of FO-rewritability in terms of the existence of certain ABoxes. For a tree-shaped ABox $\mathcal{A}$ with distinguished root $\rho_\mathcal{A}$ and $k \geq 0$, we use $\mathcal{A}|_k$ to denote the restriction of $\mathcal{A}$ to all $a \in \mathsf{Ind}(\mathcal{A})$ with distance from $\rho_\mathcal{A}$ less than or equal to k. Moreover, $\mathcal{A} \setminus \{\rho_\mathcal{A}\}$ denotes the (acyclic) ABox obtained from $\mathcal{A}$ by removing from $\mathcal{A}$ all assertions that involve $\rho_\mathcal{A}$.

Theorem 7. *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox, Σ an ABox signature, and $A(x)$ an AQ.*

1. *$A(x)$ is FO-rewritable relative to $\mathcal{T}$ and consistent Σ-ABoxes iff there is a $k \geq 0$ such that for all tree-shaped Σ-ABoxes $\mathcal{A}$ with root $\rho_\mathcal{A}$ that are consistent w.r.t. $\mathcal{T}$: if $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$, then $\mathcal{A}|_k, \mathcal{T} \models A(\rho_\mathcal{A})$;*
2. *ABox inconsistency is FO-rewritable relative to $\mathcal{T}$ and Σ iff there is a $k \geq 0$ such that for all tree-shaped Σ-ABox $\mathcal{A}$ with root $\rho_\mathcal{A}$: if $\mathcal{A}$ is inconsistent w.r.t. $\mathcal{T}$ and $\mathcal{A} \setminus \{\rho_\mathcal{A}\}$ is consistent w.r.t. $\mathcal{T}$, then $\mathcal{A}|_k$ is inconsistent w.r.t. $\mathcal{T}$.*

Proof (sketch). For the "only if" direction of Point 1, let φ be an FO-rewriting of $A(x)$ relative to $\mathcal{T}$ and consistent Σ-ABoxes. By Proposition 4, we may assume φ to be a UCQ. Let k be the maximum number of atoms in any CQ in φ. It can be shown that k is the required bound, that is, for all tree-shaped Σ-ABoxes $\mathcal{A}$ with root $\rho_\mathcal{A}$ that are consistent w.r.t. $\mathcal{T}$, $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$ implies $\mathcal{A}|_k, \mathcal{T} \models A(\rho_\mathcal{A})$.

For the "if" direction, assume that k satisfies the conditions in Point 1. We consider all minimal tree-shaped Σ-ABoxes $\mathcal{A}$ of depth at most k that are consistent w.r.t. $\mathcal{T}$ and such that $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$ and $\mathcal{A}|_k, \mathcal{T} \not\models A(\rho_\mathcal{A})$. View each such ABox $\mathcal{A}$ as a (tree-shaped) CQ $q_\mathcal{A}$ in the obvious way with the root $\rho_\mathcal{A}$ translated into the answer variable x, and define $\varphi(x)$ to be the UCQ obtained as the disjunction of all the CQs $q_\mathcal{A}$ (it is not hard to see that there are only finitely many such queries, up to equivalence). Then $\varphi(x)$ is a UCQ-rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ on acyclic Σ-ABoxes. We use a result on 'unraveling tolerance' from [Lutz and Wolter, 2012] to show that such a $\varphi(x)$ must also be a rewriting of $A(x)$ relative to $\mathcal{T}$ and Σ (for this argument, it is not sufficient to know that $\varphi(x)$ is a UCQ-rewriting on tree-shaped Σ-ABoxes). The proof of Point 2 is similar. □

The following examples illustrate Theorem 7.

Example 8. *(1) We use Point 1 of Theorem 7 to show that $A(x)$ is not FO-rewritable relative to $\mathcal{T} = \{\exists r.A \sqsubseteq A\}$ and the full ABox signature. In fact, it suffices to observe that the ABoxes $\mathcal{A}_k = \{r(a_0, a_1), \ldots, r(a_k, a_{k+1}), A(a_{k+1})\}$ with $\rho_{\mathcal{A}_k} = a_0$ are consistent w.r.t. $\mathcal{T}$ and satisfy $\mathcal{A}_k, \mathcal{T} \models A(\rho_{\mathcal{A}_k})$ and $\mathcal{A}_k|_k, \mathcal{T} \not\models A(\rho_{\mathcal{A}_k})$.*

(2) In Example 5, it was claimed that ABox inconsistency is not FO-rewritable relative to $\mathcal{T} = \{\exists r.A \sqsubseteq A, A \sqcap B \sqsubseteq \bot\}$ and the full ABox signature. This is a consequence of Point 2 of Theorem 7 and the facts that the ABoxes $\mathcal{A}'_k = \mathcal{A}_k \cup \{B(a_0)\}$ are not consistent w.r.t. $\mathcal{T}$, but with $\rho_{\mathcal{A}'_k} = a_0$, both $\mathcal{A}'_k|_k$ and $\mathcal{A}'_k \setminus \{\rho_{\mathcal{A}'_k}\}$ are consistent w.r.t. $\mathcal{T}$.

(3) In Point 2 of Theorem 7, the precondition that $\mathcal{A} \setminus \{\rho_\mathcal{A}\}$ has to be consistent w.r.t. $\mathcal{T}$ cannot be dropped. To show this, let $\mathcal{T} = \{A \sqsubseteq \bot\}$. Then ABox inconsistency is FO-rewritable relative to $\mathcal{T}$ and $\Sigma = \{A, r\}$ (with rewriting $\exists x A(x)$), but $\mathcal{A}_k$ is inconsistent w.r.t. $\mathcal{T}$ and $\mathcal{A}_k|_k$ is consistent w.r.t. $\mathcal{T}$.

To exploit Theorem 7 for developing a decision procedure for FO-rewritability, we prove that the depth and outdegree of the tree-shaped ABoxes considered in that theorem can be bounded. We use $|\mathcal{T}|$ to denote the size of the TBox $\mathcal{T}$, that is, the number of symbols needed to write $\mathcal{T}$.

Theorem 9. *Let $\mathcal{T}$ be an $\mathcal{ELI}_\bot$ TBox in normal form, Σ an ABox signature, $A(x)$ an AQ, and $n = |\mathcal{T}|$. Then Points 1 and 2 of Theorem 7 still hold when*

1. *"there is a $k \geq 0$" is replaced with "for $k = 2^{3n^2}$" and*

2. *"tree-shaped Σ-ABox $\mathcal{A}$" is replaced with "tree-shaped Σ-ABox $\mathcal{A}$ of outdegree at most n".*

Point 1 is established by a very careful pumping argument (here, the presence of inverse roles complicates matters significantly), and Point 2 relies on a selection of the relevant individuals in tree-shaped ABoxes.

Theorem 9 immediately suggests a naïve decision procedure for deciding FO-rewritability: simply enumerate all tree-shaped Σ-ABoxes up to the relevant bounds and check that they have the required properties. To obtain better complexity, we construct a tree automaton that accepts precisely those ABoxes which violate the required properties, and use a subsequent emptiness test. Specifically, we work with alternating two-way Büchi automata on finite trees, using the two-way feature to handle inverse roles. We obtain EXPTIME upper bounds and establish matching lower bounds via a reduction from subsumption in $\mathcal{ELI}$, which is EXPTIME-hard [Baader *et al.*, 2008]. Theorem 3 lifts the upper bounds to Horn-$\mathcal{SHI}$.

Theorem 10. *The following problems are* EXPTIME-*complete, with the lower bounds already applying to $\mathcal{ELI}$ (Points 1 and 3) and $\mathcal{ELI}_\bot$ (Point 2), and to the full ABox signature:*

1. *Given a Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$, an ABox signature Σ, and an AQ $A(x)$, is $A(x)$ FO-rewritable relative to $\mathcal{T}$ and Σ-ABoxes that are consistent w.r.t. $\mathcal{T}$?*
2. *Given a Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$ and an ABox signature Σ, is inconsistency of Σ-ABoxes FO-rewritable relative to $\mathcal{T}$ and Σ?*
3. *Given a Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$, an ABox signature Σ, and an AQ $A(x)$, is $A(x)$ FO-rewritable relative to $\mathcal{T}$ and Σ?*

We now discuss the actual computation of FO-rewritings. A method for computing FO-rewritings of AQs relative to consistent ABoxes and FO-rewritings of ABox inconsistency is implicit in the proof (sketch) of Theorem 7. As explained in the proof of Proposition 6, these rewritings can be combined into FO-rewritings of AQs, without the restriction to consistent ABoxes. By Theorem 3, this can be lifted from $\mathcal{ELI}_\bot$ to Horn-$\mathcal{SHI}$. The constructed rewritings are UCQs whose disjuncts are tree-shaped CQs and whose size is at most triple exponential in the size of the TBox. Using constructions from [Lutz and Wolter, 2010; Nikitina and Rudolph, 2012], one can show that this is essentially optimal. The lower bound already applies to $\mathcal{EL}$-TBoxes and the full ABox signature.

Theorem 11.

1. *For every Horn-$\mathcal{SHI}$ TBox $\mathcal{T}$, signature Σ, and AQ $A(x)$ that is rewritable relative to $\mathcal{T}$ and Σ, one can effectively construct a UCQ-rewriting $\varphi(x)$ of size at most $2^{2^{2^{\mathcal{O}(|\mathcal{T}|^2)}}}$, in time polynomial in the size of $\varphi(x)$.*
2. *There is a family of $\mathcal{EL}$ TBoxes $\mathcal{T}_0, \mathcal{T}_1, \ldots$ such that for all $i \geq 0$, $|\mathcal{T}_i| \in \mathcal{O}(i^2)$ and $A(x)$ is FO-rewritable relative to $\mathcal{T}_i$ and the full ABox signature $\mathsf{sig}(\mathcal{T}_i)$, but the smallest UCQ-rewriting is of size at least $2^{2^{2^i}}$.*

4 FO-rewritability in $\mathcal{EL}$

We next consider FO-rewritability relative to TBoxes formulated in the popular lightweight DL $\mathcal{EL}$. Unlike for Horn-$\mathcal{SHI}$ and $\mathcal{ELI}_\bot$, we obtain different complexities depending on whether we admit the ABox signature as an input or fix it to be the full signature. Note that, in $\mathcal{EL}$, ABox consistency is not an issue as every ABox is consistent w.r.t. every TBox. The results in this section can be extended to the extension $\mathcal{EL}_\bot$ of $\mathcal{EL}$ with the $\bot$ concept by dealing with inconsistent ABoxes in essentially the same way as was done in Section 3.

Again, we can work with TBoxes in normal form.

Theorem 12. *FO-rewritability of AQs relative to $\mathcal{EL}$ TBoxes and the full ABox signature can be polynomially reduced to FO-rewritability of AQs relative to $\mathcal{EL}$ TBoxes in normal form and the full ABox signature.*

Note that Theorem 12 differs from Theorem 3 in that we are interested in FO-rewritability of $A(x)$ relative to the normalized TBox $\mathcal{T}'$ *and the full signature* $\mathsf{sig}(\mathcal{T}')$, rather than the original signature $\mathsf{sig}(\mathcal{T})$ as in Theorem 3. In fact, proving Theorem 12 requires a more careful construction.

We now show the surprising result that, when the TBox is formulated in $\mathcal{EL}$ and the ABox signature is full, the tree-shaped ABoxes from Theorem 9 can be replaced with linear ones. Formally, an ABox $\mathcal{A}$ is *linear* if it consists of role assertions $r_0(a_0, a_1), \ldots, r_{n-1}(a_{n-1}, a_n)$ with $a_i \neq a_j$ for $i \neq j$ and concept assertions $A(a)$ with $a \in \{a_0, \ldots, a_n\}$.

Theorem 13. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox in normal form, $A(x)$ an AQ, $n = |\mathcal{T}|$, and $k = 2^{3n^2}$. Then $A(x)$ is FO-rewritable relative to $\mathcal{T}$ and the full ABox signature iff for all linear ABoxes $\mathcal{A}$ with root $\rho_\mathcal{A}$, $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$ implies $\mathcal{A}|_k, \mathcal{T} \models A(\rho_\mathcal{A})$.*

Proof. (sketch) In light of Theorem 9, it is sufficient to show that if there is a tree-shaped ABox $\mathcal{A}$ with root $\rho_\mathcal{A}$ such that $\mathcal{A}, \mathcal{T} \models A(\rho_\mathcal{A})$ and $\mathcal{A}|_k, \mathcal{T} \not\models A(\rho_\mathcal{A})$ with $k = 2^{3n^2}$, then there is a linear ABox $\mathcal{A}'$ that satisfies the same properties. Since $\mathcal{EL}$ does not allow inverse roles, we can assume w.l.o.g. that $\mathcal{A}$ has the shape of a *directed* tree, that is, whenever $r(a, b) \in \mathcal{A}$, then b is further away from the root b than a. Note that the depth of $\mathcal{A}$ must exceed k. We can further assume that $B(b) \in \mathcal{A}$ whenever $\mathcal{A}|_k, \mathcal{T} \models B(b)$, since these assertions can be added without changing the relevant properties of $\mathcal{A}$. By replacing subtrees with the concept assertions that they entail, we can further ensure that there is only a single individual on level $k+1$, and no individuals on any level $> k+1$. The desired linear ABox $\mathcal{A}'$ is then defined as the restriction of $\mathcal{A}$ to assertions that involve only the individuals that appear on the unique path in $\mathcal{A}$ of length $k+1$. Since $\mathcal{A}'|_k \subseteq \mathcal{A}|_k$, we have $\mathcal{A}'|_k, \mathcal{T} \not\models A(\rho_\mathcal{A})$. Since all assertions entailed by $\mathcal{A}|_k$ and $\mathcal{T}$ appear in $\mathcal{A}$ and all individuals in $\mathcal{A} \setminus \mathcal{A}'$ are on level at most k, we have $\mathcal{A}', \mathcal{T} \models A(\rho_\mathcal{A})$. The latter argument relies on $\mathcal{A}$ being a directed tree. □

The following example shows that, even for $\mathcal{EL}$, it is not possible to replace tree-shaped ABoxes with linear ones if we are interested in signatures other than the full signature.

Example 14. *Let*

$$\mathcal{T} = \{A_i \sqsubseteq X_i,\ B_i \sqcap X_i \sqsubseteq Y_i,\ \exists r.Y_i \sqsubseteq X_i \mid i \in \{1,2\}\} \cup \{X_1 \sqcap X_2 \sqsubseteq X,\ B_1 \sqcap B_2 \sqsubseteq Z,\ \exists r.Z \sqsubseteq X\},$$

choose $\Sigma = \{A_1, A_2, B_1, B_2, r\}$*, and take the AQ* $X(x)$*. The tree-shaped ABox* $\mathcal{A}$ *composed of the assertions*

$$\{r(a_0, a_{i,0}), r(a_{i,0}, a_{i,1}), ..., r(a_{i,2^{3n^2}}, a_{i,2^{3n^2}+1}) \mid i \in \{1,2\}\}$$
$$\cup \{B_i(a_{i,0}), \ldots, B_i(a_{i,2^{3n^2}+1}), A_i(a_{i,2^{3n^2}+1}) \mid i \in \{1,2\}\},$$

with n *as in Theorem 9, is of depth exceeding* 2^{3n^2}*, and we can show that* $\mathcal{A}, \mathcal{T} \models X(a_0)$*, but* $\mathcal{A}|_{2^{3n^2}}, \mathcal{T} \not\models X(a_0)$*. However, for all linear* Σ*-ABoxes* $\mathcal{A}$*, we have* $\mathcal{A}, \mathcal{T} \models X(a_0)$ *iff* $\mathcal{A}|_1, \mathcal{T} \models X(a_0)$*: since* $X, X_1, X_2 \notin \Sigma$*, we can only have* $\mathcal{A}, \mathcal{T} \models X(a_0)$ *if there is an* r*-successor* b *of* a_0 *in* $\mathcal{A}$ *where* Z *is entailed, or where* Y_1 *and* Y_2 *are entailed. Since* $Y_1, Y_2, Z \notin \Sigma$*, this in turn can only be the case when* $B_1(b), B_2(b) \in \mathcal{A}$*. But then,* $\mathcal{A}|_1, \mathcal{T} \models X(a_0)$*.*

Theorem 13 allows us to replace the alternating tree automata in the proof of Theorem 10 with alternating word automata, improving the upper bound to PSPACE.

Theorem 15. *Deciding FO-rewritability of an AQ relative to an* $\mathcal{EL}$ *TBox and the full ABox signature is in* PSPACE.

We establish matching lower bounds.

Theorem 16. *Deciding FO-rewritability of an AQ relative to an* $\mathcal{EL}$ *TBox and an ABox signature* Σ *is (1)* PSPACE*-hard when* Σ *is full and (2)* EXPTIME*-hard when* Σ *is an input.*

Point 1 is proved by a reduction from the word problem for polynomially space-bounded deterministic Turing machines. For Point 2, we use polynomially space-bounded alternating Turing machines. As both reductions are lengthy and somewhat subtle, we present instead a proof of CONP-hardness, which illustrates some general ideas that are also used in the other reductions.

We reduce propositional tautology to FO-rewritability of AQs relative to $\mathcal{EL}$ TBoxes and the full signature. Let ϑ be a propositional formula in negation normal form with variables $p_1, \ldots, p_n$, and let $\mathsf{sub}(\vartheta)$ be the set of subformulas of ϑ. Define a TBox $\mathcal{T}$ with the CIs:

$$\begin{array}{rcll}
\exists r.(L_i \sqcap V_{i,i}) & \sqsubseteq & L_{i-1} & V \in \{T,F\}, 1 \leq i \leq n \\
\exists r.V_{i,j} & \sqsubseteq & V_{i,j-1} & V \in \{T,F\}, 1 \leq j \leq i \leq n \\
T_{i,0} & \sqsubseteq & A_{p_i} & p_i \in \mathsf{sub}(\vartheta) \\
F_{i,0} & \sqsubseteq & A_{\neg p_i} & \neg p_i \in \mathsf{sub}(\vartheta) \\
A_\varphi \sqcap A_\psi & \sqsubseteq & A_{\varphi \wedge \psi} & \varphi \wedge \psi \in \mathsf{sub}(\vartheta) \\
A_\rho & \sqsubseteq & A_{\varphi \vee \psi} & \rho \in \{\varphi, \psi\}, \varphi \vee \psi \in \mathsf{sub}(\vartheta) \\
A_\vartheta & \sqsubseteq & L_0 & \\
\exists r.L_0 & \sqsubseteq & L_n &
\end{array}$$

Lemma 17. ϑ *is a tautology iff* $L_0(x)$ *is FO-rewritable relative to* $\mathcal{T}$ *and the full signature.*

Proof. First assume that ϑ is not a tautology. Then there is a truth assignment t such that $t \not\models \vartheta$. Let $k = 2^{3|\mathcal{T}|^2}$ and define a linear ABox $\mathcal{A}$ as the union of

$$\{\, r(a_{1,0}, a_{1,1}), \ldots, r(a_{1,n}, a_{2,0}), \ldots, r(a_{k,n-1}, a_{k,n})\}$$
$$\{r(a_{k,n}, a), L_0(a)\,\}$$
$$\{F_{i,j}(a_{\ell,j}) \mid 1 \leq i \leq n,\ 0 \leq j \leq i,\ 1 \leq \ell \leq k,\ t(p_i) = \mathsf{f}\}$$
$$\{T_{i,j}(a_{\ell,j}) \mid 1 \leq i \leq n,\ 0 \leq j \leq i,\ 1 \leq \ell \leq k,\ t(p_i) = \mathsf{t}\}.$$

Starting from the assertion $L_0(a)$, one can derive $L_n(a_{k,n})$, then $L_{n-1}(a_{k,n-1})$, and so on, until one obtains $L_0(a_{1,0})$. Note that the generation of $L_0(a_{1,0})$ cannot be 'shortcut' using the inclusion $A_\vartheta \sqsubseteq L_0$: since $t \not\models \vartheta$, A_ϑ is not derived anywhere on the chain. We thus have $\mathcal{A}, \mathcal{T} \models L_0(a_{1,0})$, but $\mathcal{A}|_k, \mathcal{T} \not\models L_0(a_{1,0})$. By Theorem 13, L_0 is not FO-rewritable relative to $\mathcal{T}$ and the full ABox signature.

Now assume that ϑ is a tautology and that $\mathcal{A}, \mathcal{T} \models L_0(a_0)$ with $\mathcal{A}$ linear and $r(a_0, a_1), \ldots, r(a_{m-1}, a_m)$ the role assertions in $\mathcal{A}$. Assume to the contrary of what is to be shown that $\mathcal{A}|_k, \mathcal{T} \not\models L_0(a_0)$, with k as above. Then $m > k$. By analyzing $\mathcal{T}$, it can be verified that we must have assertions $V_{i,i}(a_i) \in \mathcal{A}$ with $V \in \{T, F\}$, for $1 \leq i \leq n$. These assertions represent a truth assignment t. Since ϑ is valid, we have $t \models \vartheta$. Again analyzing $\mathcal{T}$, this can be used to show that $\mathcal{A}|_k, \mathcal{T} \models L_0(a_0)$, a contradiction. □

5 Related Work

As observed in [Lutz and Wolter, 2011], there is a close connection between FO-rewritability of AQs relative to DL TBoxes and the *boundedness problem* for datalog programs. In fact, the known 2EXPTIME upper bound for predicate boundedness of connected monadic datalog programs [Cosmadakis *et al.*, 1988] can be used to obtain a 3EXPTIME upper bound for FO-rewritability of an AQ relative to an $\mathcal{ELI}$ TBox; via our Proposition 6, this can be extended to $\mathcal{ELI}_\bot$ TBoxes. Boundedness was studied also in the context of the μ-calculus, for which it is EXPTIME-complete [Otto, 1999], and for monadic second order logic [Blumensath *et al.*, 2009]. A different approach to FO-rewritability is suggested in [Bienvenu *et al.*, 2013], based on a connection between query answering in DLs and constraint satisfaction problems (CSPs). This approach is different in spirit from ours and tailored towards expressive DLs of the $\mathcal{ALC}$ family. However, it also yields a NEXPTIME upper bound for FO-rewritability relative to Horn-$\mathcal{SHI}$ TBoxes. Finally, it is shown in [Bienvenu *et al.*, 2012a] that FO-rewritability of AQs relative to a restricted form of $\mathcal{EL}$ TBoxes called *classical TBoxes* is PTIME-complete; that work also analyzes *acyclic TBoxes*, for which FO-rewritings always exist.

6 Future Work

It would be interesting to generalize our approach both regarding the query language and the ontology language covered. Regarding the latter, it would be particularly interesting to generalize our results from Horn-$\mathcal{SHI}$ to Horn-$\mathcal{SHIQ}$, which we conjecture to be possible using slight extensions of the techniques introduced in this paper. Regarding the query language, it would be interesting to analyze FO-rewriting of conjunctive queries. We believe that a mix of techniques from this paper and those in [Bienvenu *et al.*, 2012b] might provide a good starting point. Finally, existing ontologies should be investigated regarding FO-rewritability. Important questions are: How many atomic queries are FO-rewritable w.r.t. natural ABox signatures? How difficult is it to find FO-rewritings if they exist, and how large are they? Interesting ontologies to consider are GALEN and non-acyclic versions of NCI.

Acknowledgements. M. Bienvenu was supported by ANR project PAGODA (ANR-12-JS02-007-01). C. Lutz was supported by the DFG SFB/TR 8 'Spatial Cognition'.

References

[Baader *et al.*, 2003] Franz Baader, Deborah L. McGuiness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook*. Cambridge University Press, 2003.

[Baader *et al.*, 2005] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. of IJCAI*, pages 364–369, 2005.

[Baader *et al.*, 2008] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope further (long version). In *Proc. of OWLED workshop*, 2008.

[Baader *et al.*, 2010] Franz Baader, Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. Query and predicate emptiness in description logics. In *Proc. of KR*, pages 192–202, 2010.

[Bienvenu *et al.*, 2012a] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. Deciding FO-rewritability in $\mathcal{EL}$. In *Proc. of DL workshop*, 2012.

[Bienvenu *et al.*, 2012b] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. Query containment in description logics reconsidered. In *Proc. of KR*, 2012.

[Bienvenu *et al.*, 2013] Meghyn Bienvenu, Balder ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: A study through disjunctive datalog, csp, and mmsnp. In *Proc. of PODS*, 2013.

[Blumensath *et al.*, 2009] Achim Blumensath, Martin Otto, and Mark Weyer. Boundedness of monadic second-order formulae over finite words. In *Proc. of ICALP*, pages 67–78, 2009.

[Calvanese *et al.*, 2007] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *J. of Automated Reasoning*, 39(3):385–429, 2007.

[Chortaras *et al.*, 2011] Alexandros Chortaras, Despoina Trivela, and Giorgos B. Stamou. Optimized query rewriting for OWL 2 QL. In *Proc. of CADE*, pages 192–206, 2011.

[Cosmadakis *et al.*, 1988] Stavros S. Cosmadakis, Haim Gaifman, Paris C. Kanellakis, and Moshe Y. Vardi. Decidable optimization problems for database logic programs (preliminary report). In *Proc. of STOC*, pages 477–490, 1988.

[Eiter *et al.*, 2008] Thomas Eiter, Georg Gottlob, Magdalena Ortiz, and Mantas Simkus. Query answering in the description logic Horn-$\mathcal{SHIQ}$. In *Proc. of JELIA*, pages 166–179, 2008.

[Eiter *et al.*, 2012] Thomas Eiter, Magdalena Ortiz, Mantas Simkus, Trung-Kien Tran, and Guohui Xiao. Query rewriting for Horn-$\mathcal{SHIQ}$ plus rules. In *Proc. of AAAI*, 2012.

[Gottlob *et al.*, 2011] Georg Gottlob, Giorgio Orsi, and Andreas Pieris. Ontological queries: Rewriting and optimization. In *Proc. of ICDE*, pages 2–13, 2011.

[Hustadt *et al.*, 2007] Ullrich Hustadt, Boris Motik, and Ulrike Sattler. Reasoning in description logics by a reduction to disjunctive datalog. *J. of Automated Reasoning*, 39(3):351–384, 2007.

[Kazakov, 2009] Yevgeny Kazakov. Consequence-driven reasoning for Horn-$\mathcal{SHIQ}$ ontologies. In Craig Boutilier, editor, *Proc. of IJCAI*, pages 2040–2045, 2009.

[Lutz and Wolter, 2010] Carsten Lutz and Frank Wolter. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. *J. of Symbolic Computation*, 45(2):194–228, 2010.

[Lutz and Wolter, 2011] Carsten Lutz and Frank Wolter. Non-uniform data complexity of query answering in description logics. In *Proc. of DL workshop*, 2011.

[Lutz and Wolter, 2012] Carsten Lutz and Frank Wolter. Non-uniform data complexity of query answering in description logics. In *Proc. of KR*, 2012.

[Lutz *et al.*, 2009] Carsten Lutz, David Toman, and Frank Wolter. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In *Proc. of IJCAI*, pages 2070–2075, 2009.

[Nikitina and Rudolph, 2012] Nadeschda Nikitina and Sebastian Rudolph. Expexpexplosion: Uniform interpolation in general $\mathcal{EL}$ terminologies. In *Proc. of ECAI*, pages 618–623, 2012.

[Otto, 1999] Martin Otto. Eliminating recursion in the μ-calculus. In *Proc. of STACS*, pages 531–540, 1999.

[Pérez-Urbina *et al.*, 2009] Héctor Pérez-Urbina, Ian Horrocks, and Boris Motik. Efficient query answering for OWL 2. In *Proc. of ISWC*, pages 489–504, 2009.

[Rossman, 2008] Benjamin Rossman. Homomorphism preservation theorems. *J. of the ACM*, 55(3), 2008.

[Serre, 2006] Olivier Serre. Parity games played on transition graphs of one-counter processes. In *Proc. of FoSSaCS*, pages 337–351, 2006.

[Vardi, 1998] Moshe Y. Vardi. Reasoning about the past with two-way automata. In *Proc. of ICALP*, pages 628–641, 1998.