Faculty of Computer Science, Free University of Bozen-Bolzano, Piazza Domenicani 3, 39100 Bolzano, Italy
Tel: +39 04710 16000, fax: +39 04710 16009, http://www.inf.unibz.it/krdb/

KRDB Research Centre Technical Report:

Expressive Power of DL-Lite (II)

Camilo Thorne

Affiliation	KRDB Research Centre, Faculty of Computer Science Free University of Bozen-Bolzano Piazza Domenicani 3, 39100, Bolzano, Italy
Corresponding author	Camilo Thorne cthorne@inf.unibz.it
Keywords	Description logics, DL-Lite, bi-simulations, absolute and relative expressive power
Number	KRDB08-07
Date	1-7-08
URL	http://www.inf.unibz.it/krdb/pub/

Expressive Power of DL-Lite$_{R,\sqcap}$ (II)

Camilo Thorne

July 2007

Contents

Introduction

When crafting a controlled language (CL) out of some natural language like English, care should be taken to clearly establish its expressivity bounds. CLs are fragments of natural language (NL) defined with the purpose of carrying out a particular data management of knowlegde representation without the ambiguity inherent to NL. Lite English purports to capture in NL query answering over description logic knowledge bases (QA) and in particular, over DL-Lite knowledge bases and hence ontology driven data access (cf. [1]). But then, how can we be sure we have attained this goal or if it is even possible? How can we know that a controlled language makes sense to a native speaker in the context of, say, accessing data from some structured knowlege source using a NL interface? One way, that we have explored to some extent is that of comparing Lite English (and *a fortiori* DL-Lite$_{R,\sqcap}$ to fragments of English that already exist, namely those of I. Pratt and A. Third (cf. [10]). Comparing them w.r.t. expressiver power provides a way of characterizing the expressive power of Lite English relatively to these fragments (cf. [1, 2, 12]). The question now is: can we characterize it in *absolute* terms? Is there some property capable of providing us sufficient and necessary considitions regarding the expressiveness (in FOL model-theoretical terms) of Lite English? Yes and no: yes at the level of formulae or concepts and no at the level of sentences and assertions. At the level of concepts, a notion of *simulation* can be defined, in a way analogous to other description logics (cf. [9]). However, when we move to assertions, such closure properties cease to provide sufficient conditions (only necessary ones). Moreover, we show that the entailement problem QA is **NP-Complete** in *combined complexity* (i.e., when both the query and the whole KB are taken into account): the same complexity class as that of COP+Rel. This without however being propositionally complete, given the controlled behaviour of negation, relatives and conjunction. But in any case showing that for a more fine-grained expressivity analysis, semantic techniques are essential.

The structure of this report is as follows. Section 1 will recall some basic results regarding the computational complexity of QA over DL-Lite$_{R,\sqcap}$ knowledge bases. Section 2 will recall some properties of DL-Lite$_{R,\sqcap}$ as a fragment of FOL. Section 3 will introduce the notion of DL-Lite$_{R,\sqcap}$ simulation. Section 4 will provide some negative results regarding an absolute characterization of DL-Lite$_{R,\sqcap}$'s expressive power (in model-theoretic terms). Section 5 will compare DL-Lite$_{R,\sqcap}$'s expressive power to that of Pratt's and Third's intractable fragment of English COP+Rel. Finally, the conclusions will sum up our results.

1 Complexity of QA over DL-Lite$_{R,\sqcap}$ KBs

In this section we recall some results regarding the computational complexity of conjunctive query answering (QA) over DL-Lite$_{R,\sqcap}$ knowledge bases (cf. [7, 6, 5]). This is important since the complexity is one of the two main properties that characterize a logic's expressive power – the other one being the classes of structures (or models) it can express. A we shall see, it is **NP-Complete** in *combined complexity*. That is, when we take as input for the decision problem (i) the size $|q|$ of a UCQ q (the number of its symbols), (ii) the *data complexity* of the KB (the number of its pairwise distinct constants) and (iii) the size $\#(\mathcal{T})$ of its TBox (the number of assertions the TBox contains). This implies that there is no way of distinguishing DL-Lite from, say, COP+Rel by computational complexity alone – we will distinguish them later on by purely semantic means.

We begin by recalling the notion of FOL-*reducibility*. A DL is said to be FOL-*reducible* whenever, given a UCQ q and a KB we can "compile" the TBox into the query and store the ABox in a DB engine in such a way that the computational complexity of full FOL queries is preserved in this new setting: it has to be logarithmic in the number of the DB's tuples. That is, whenever a *perfect reformulation* exists for this logic. DL-Lite$_{R,\sqcap}$ happens to verify this property.

Definition 1.1. A *perfect reformulation* is a reduction algorithm denoted $\mathsf{PerfectRef}(\mathcal{T}, q)$ that takes as input a description logic TBox $\mathcal{T}$ and a UCQ q and outputs in time polynomial on the size $\#(\mathcal{T})$ of $\mathcal{T}$ a UCQ $q_{\mathcal{T}}$ such that, for every description logic KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, it holds that:

$$\begin{array}{lll} \mathcal{T}, \mathcal{A} \models q(\vec{c}) & \text{iff} & \vec{c} \in \mathsf{PerfectRef}(\mathcal{T}, q)^{\mathcal{A}} \\ & \text{iff} & \vec{c} \in q_{\mathcal{T}}^{\mathcal{A}}. \end{array}$$

That is, such that the description logic is FOL-reducible.

Proposition 1.1. **(Calvanese et. al)** *A* PerfectRef *exists for* DL-Lite$_{R,\sqcap}$.

Proposition 1.2. **(Calvanese et. al)** QA *over* DL-Lite$_{R,\sqcap}$ KBs *is:*

- **LOGSPACE** *in data complexity.*
- **P-Hard** *on the size of the KB.*
- **NP-Complete** *in query complexity.*

Lemma 1.1. QA *is in* **NP**.

Proof. Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a KB and let $q(\vec{c})$ be the grounding of a CQ over the signature $\mathcal{L}$ of $\mathcal{K}$. First, consider:

$$\mathcal{T}, \mathcal{A} \models q(\vec{c}). \tag{1}$$

Let $\#(\mathcal{T})$ denote the number of assertions in $\mathcal{T}$. We know that $\mathcal{T}$ can be compiled into q by PerfectRef in time polynomial on $\#(\mathcal{T})$, such that:

$$\mathcal{A} \models q_{\mathcal{T}}(\vec{c}) \tag{2}$$

is equivalent to (1). Now, $q_{\mathcal{T}}(\vec{c})$ is of the form $\exists\vec{y}_1\psi_1(\vec{y}_1,\vec{c}_1)\vee ... \vee\exists\vec{y}_k\psi_k(\vec{y}_k,\vec{c}_k)$. Hence, (2) holds if, for some $i \in [1,k]$, there is an assigment $v\colon Var(\psi_i) \to Con(\mathcal{A})$, where $Var(\psi_i)$ denotes the set of variables of ψ_i and $Con(\mathcal{A})$ the set of constants of $\mathcal{A}$, such that:

(3) $$\mathcal{A} \models \exists\vec{y}_i\psi_i(\vec{y}_i,\vec{c}_i)[v].$$

Remark that the formula $\psi_i(\vec{y}_i,\vec{c}_i)$ for $i \in [1,k]$ is quantifier-free. Next, choose a random assigment v and scan the k disjuncts of the UCQ. Denote by $\#(\mathcal{A})$ the number of assertions in $\mathcal{A}$. Suppose moreover w.l.g. that the $\psi_i(\vec{y}_i,\vec{c}_i)$, for $i \in [1,k]$ contain at most p atoms. Then we can check whether (3) holds in time polynomial on $k \times \#(\mathcal{A}) \times p$. □

Lemma 1.2. QA *is* **NP-Hard** *in combined complexity.*

Proof. By reduction of the graph homeomorphism problem. We will consider KBs with empty TBoxes. Let $G_1 = \langle V_1, E_1\rangle$ and $G_2 = \langle V_2, E_2\rangle$ be two graphs. Encode G_1 and G_2 as follows:

- For each $\langle u, v\rangle \in E_1$, add the fact $R(c_u, c_v)$ to the ABox $\mathcal{A}_{G_1}$.
- For each $\langle u', v'\rangle \in E_2$, add the ground atom $R(c_{u'}, c_{v'})$ to the UCQ q_{G_2}, which is a conjunction of such atoms.

Both steps can be done in time polynomial on $\#(E_1)$ and $\#(E_2)$. Now we claim that:

(4) There is an homeomorphism h from G_1 to G_2 iff $\mathcal{A}_{G_1} \models q_{G_2}$.

Recall that q_{G_2} is of the form $q_{G_2} := q_{G_2}(\vec{c})$. Now, since there is a PerfectRef algorithm for DL-Lite:

$$\mathcal{A}_{G_1} \models q_{G_2}(\vec{c}) \text{ iff } \vec{c} \in q_{G_2}^{G_1}.$$

Now, the interpretation function $.^{G_1}$ can be seen as an homeomorphism mapping q_{G_2} onto G_1. Since q_{G_2} encodes G_2, the claim (4) follows immediately. □

Theorem 1.1. QA *is* **NP-Complete** *in combined complexity.*

2 DL-Lite$_{R,\sqcap}$ as a fragment of FOL

We can extend the language $\mathcal{L}$ of DL-Lite to a language $\mathcal{L}'$ with some new constants and concept constructors such that every TBox $\mathcal{T}'$ over $\mathcal{L}'$ is a conservative extension of a TBox $\mathcal{T}$ over $\mathcal{L}$. A TBox $\mathcal{T}'$ over $\mathcal{L}'$ is said to be a *conservative extension* of a TBox $\mathcal{T}$ over $\mathcal{L}$ when, and only when, for every assertion α over $\mathcal{L}$, $\mathcal{T}' \models \alpha$ iff $\mathcal{T} \models \alpha$. This can be trivially achieved with the following definition:

Definition 2.1. We put:

$$
\begin{array}{rcl}
\{A \sqsubseteq B \sqcap C\} & =_{df} & \{A \sqsubseteq B, \\
 & & A \sqsubseteq C\}. \\
\{A \sqcup B \sqsubseteq C\} & =_{df} & \{A \sqsubseteq C, \\
 & & B \sqsubseteq C\}. \\
\{A \sqsubseteq \exists R\colon C\} & =_{df} & \{A \sqsubseteq \exists R, \\
 & & R \sqsubseteq R', \\
 & & \exists R'^{-} \sqsubseteq C\}. \\
\{A \sqsubseteq \bot\} & =_{df} & \{A \sqsubseteq B \sqcap \neg B\}.
\end{array}
$$

Where C and D are arbitrary right hand side concepts, A and B arbitrary left hand side concepts and R, R' basic roles. The right hand side concepts are called, respectively, *qualified existential role*, *concept conjunction* and *bottom*. The left hand side concept is called *concept disjunction*.

DL-Lite, as other description logics is a fragment of FOL (cf. [3]), belonging to HORN, i.e., the class of FOL horn clauses (cf. [4]). This means that they are closed under the following properties: (i) finite intersections and (ii) ultraproducts (cf. [8]). Property (i) in particular implies the existence of minimal models up to elementary equivalence of structures as well as the existence of a least Herbrand model w.r.t. inclusion. Furthermore, DL-Lite is included in FO^2 the two-variable fragment of FOL and, as most description logics, in the guarded fragment – hence it satisfies the finite model property. Finally, as it belongs to the $\forall\exists^*$ prefix class, it is included in the Gödel class.

3 DL-Lite$_{R,\sqcap}$ Simulations

Our purpose in this section is that of characterizing the absolute expressive power of DL-Lite as a logic. This is possible only at the level of concepts or formulas (in FOL) and not of assertions (or sentences) by means of the notion of DL-Lite$_{R,\sqcap}$ simulations. This notion is adapted from de Rijke's work on classifying DLs (cf. [9]). The intuition behind is that a concept (or a formula) cannot distinguish between structures, models or interpretations in the same way a sentence or assertion does. Simulations are satisfaction-preserving equivalence relations on structures based on the notion of bisimulations for modal logic (and for DLs such as $\mathcal{ALC}$). Their nice feature is that whenever a FOL formula is closed under DL-Lite$_{R,\sqcap}$ simulations, it is equivalent to some DL-Lite$_{R,\sqcap}$ right or left hand side concept. Furthermore, this condition is both necessary and sufficient. When we move on to sentences or assertions this changes. Only necessary conditions are possible: we can prove that DL-Lite's expressive power (at the level of assertions) cannot be characterized (in the technical sense of the word) through semantic means.

Definition 3.1. Given to interpretations $\mathcal{I}$ and $\mathcal{J}$, a DL-Lite$_{R,\sqcap}$ *left simulation* is a relation $\mathcal{B} \subseteq \mathcal{P}(\Delta^{\mathcal{I}}) \times \Delta^{\mathcal{J}}$ s.t., for every $X_1 \subseteq \Delta^{\mathcal{I}}$ and every $d_2 \in \Delta^{fJ}$ and any basic concept A:

1. $X_1 \mathcal{B} d_2$ and $X_1 \subseteq A^{\mathcal{I}}$ imply $d_2 \in \Delta^{\mathcal{J}}$ (A).
2. $X_1 \mathcal{B} d_2$ and forall $d_1 \in X_1$ exists $e_1 \in y_1 \subseteq \Delta^{\mathcal{I}}$ such that $d_1 R^{\mathcal{I}} e_1$ imply exists $e_2 \in \Delta^{\mathcal{J}}$ such that $d_2 R^{\mathcal{J}} e_2$ $(\exists R)$.

The clause for concept conjunction follows implicitly from the definition. We can extend the notion of simulation to right-hand side concepts as follows:

Definition 3.2. A DL-Lite$_{R,\sqcap}$ *right simulation* is a relation as above. We just add new clauses to the definition to cover right hand side concepts. C is an arbitrary right hand side concept and B a left hand side concept:

1. $X_1 \mathcal{B} d_2$ and $X_1 \subseteq \neg B^{\mathcal{I}}$ imply $d_2 \notin \Delta^{\mathcal{J}}$ $(\neg B)$.
2. $X_1 \mathcal{B} d_2$ and forall $d_1 \in X_1$ there exists no $e_1 \in y_1 \subseteq \Delta^{\mathcal{I}}$ such that $d_1 R^{\mathcal{I}} e_1$ implies that there is no $e_2 \in \Delta^{\mathcal{J}}$ such that $d_2 R^{\mathcal{J}} e_2$ $(\neg\exists R)$.
3. $X_1 \mathcal{B} d_2$ and forall $d_1 \in X_1$ exists $e_1 \in y_1 \subseteq \Delta^{\mathcal{I}}$ such that $d_1 R^{\mathcal{I}} e_1$ imply exists $e_2 \in \Delta^{\mathcal{J}}$ such that $d_2 R^{\mathcal{J}} e_2$ and $Y_1 B e_2$ $(\exists R \colon C)$.

Definition 3.3. A DL-Lite$_{R,\sqcap}$ *simulation* is a left or a right simulation. If there exists a DL-Lite$_{R,\sqcap}$ simulation B among two interpretations $\mathcal{I}$ and $\mathcal{J}$ we say that they are DL-Lite$_{R,\sqcap}$ *similar* and write $\mathcal{I} \sim_{DL} \mathcal{J}$.

DL-Lite simulations preserve concept satisfiability. It is trivial to show that for any arbitrary interpretations $\mathcal{I}$ and $\mathcal{J}$ such that $\mathcal{I} \sim_{DL} \mathcal{J}$ and any concept C, there exist $d \in \Delta^{\mathcal{I}}$, $d' \in \Delta^{\mathcal{J}}$ such that $d \in C^{\mathcal{I}}$ iff $d' \in C^{\mathcal{J}}$.

They are equivalence relations on interpretations (reflexive, transitive and symmetric).

Definition 3.4. We say that a FOL formula ϕ is *closed under* DL-Lite *simulations* iff for every two interpretations $\mathcal{I}$ and $\mathcal{J}$, and any DL-Lite simulation $\mathcal{B} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{J}}$ every $X \subseteq \Delta^{\mathcal{I}}$ and every $d' \in \Delta^{\mathcal{J}}$ such that $X\mathcal{B}d'$ it holds that, for every $d \in X$,:

$$\mathcal{I} \models \phi[d] \text{ implies } \mathcal{J} \models \phi[d'].$$

Lemma 3.1. *If a* FOL *formula ϕ is equivalent to a* DL-Lite *right hand or left hand side concept, then it is closed under* DL-Lite$_{R,\sqcap}$ *simulations.*

Proof. Let ϕ be a FOL formula closed under DL-Lite$_{R,\sqcap}$ simulations. Let $Con(\phi)$ denote the set of consequences in DL-Lite of a FOL formula ϕ. If we can prove that ϕ and $Con(\phi)$ are equivalent we will be done. Now by compactness $Con(\phi)$ has a model iff every finite subset Σ has a model, whence the concept $\bigsqcap_{C\in\Sigma} C$ should have a model too. Clearly, $\phi \models \bigsqcap_{C\in\Sigma} C$, and hence every model of ϕ is a model of $\bigsqcap_{C\in\Sigma} C$. It is more lengthy to prove that:

$$Con(\phi) \models \phi. \tag{5}$$

Let $\mathcal{I} \models Con(\phi)[d]$, for an arbitrary intrepretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$. Now we need to see whether $\mathcal{I} \models \phi[d]$ holds too. Let $\Gamma = \{\neg C | d \notin C^{\mathcal{I}}\}$. Then,

$$\text{for every } \neg C \in \Gamma, \{\phi, \neg C\} \text{ is consistent.}$$

Otherwise $\phi \models C$ and hence $\mathcal{I} \models C[d]$, i.e., $\neg C \notin \Gamma$. Hence for every $\neg C \in \Gamma$ there exists an interpretation $\mathcal{I}_C$ and $d_C \in \Delta^{\mathcal{I}_C}$ such that $\mathcal{I}_C \models \phi[d_C]$ and $d_C \in C^{\mathcal{I}_C}$. The idea now is to build an interpretation (from which to build a DL-Lite$_{R,\sqcap}$ simulation) by picking the union of all such interpretations; *modulo* this bisimulation we will be able to prove claim (5). For this put, whenever $\neg C \in \Gamma$,:

$$\mathcal{J} = \bigcup_{d_C \in C^{\mathcal{I}_C}} (\mathcal{I}_C, d_C).$$

The structure $\mathcal{J}$ is a DL-Lite$_{R,\sqcap}$ interpretation. Then, for every $\neg C \in \Gamma$ there is a DL-Lite simulation $\mathcal{B}$ such that $\{d_C\}\mathcal{B}d_C$, for $\{d_C\} \subseteq \Delta^{\mathcal{J}}$ and $d_C \in \Delta^{\mathcal{I}_C}$. Moreover it holds that:

$$\text{for all } d_C, d_C \in D^{\mathcal{J}} \text{ implies } d \in D^{\mathcal{I}}.$$

For, indeed, let $d_C \in D^{\mathcal{J}}$ and suppose that $d \notin D^{\mathcal{I}}$. Then $\neg D \in \Gamma$ and by the same token there is a $d_D \in \Delta^{\mathcal{J}}$ such that $d_D \notin D^{\mathcal{J}}$. It is enough to put $d_C = d_D$ to get a contradiction. Now, define a DL-Lite simulation $\mathcal{B} \subseteq \mathcal{P}(\Delta^{\mathcal{J}}) \times \Delta^{\mathcal{I}}$ by putting:

$$X_1 \mathcal{B} d_2 \text{ iff for every concept } D, X_1 \subseteq D^{\mathcal{J}} \text{ implies } d_2 \in D^{\mathcal{I}}.$$

B is a DL-Lite simulation. To prove this we reason by cases:

- The property trivially holds for basic concepts.
- Consider $C = \neg D$. Let $X_1 \subsetneq \neg D^{\mathcal{J}}$, $X_1 \mathcal{B} d_2$. Now, by definition of $\mathcal{B}$, $d_2 \in C^{\mathcal{I}}$, that is, $d_2 \in \neg D^{\mathcal{I}} = \Delta^{\mathcal{I}} - D^{\mathcal{I}}$ iff $d_2 \notin D^{\mathcal{I}}$.
- Consider $C = \exists R$. Let $X_1 \mathcal{B} d_2$ and $d_1 \in X_1$ such that there is some $e_1 \in Y_1 \subseteq \Delta^{\mathcal{J}}$ such that $d_1 R^{\mathcal{J}} e_1$. Now, $X_1 \subseteq (\exists R)^{\mathcal{J}}$, so $d_2 \in (\exists R)^{\mathcal{I}}$ and hence there is some $e_2 \in \Delta^{\mathcal{I}}$ such that $d_2 R^{\mathcal{J}} e_2$.
- Consider $C = \neg \exists R$. This is proven by combining the two previous cases.
- Consider $C = \exists R\colon D$. Let $X_1 B d_2$ s.t. exists $e_1 \in Y_1 \subseteq \Delta^{\mathcal{J}}$ and $d_1 R^{\mathcal{J}} e_1$. $X_1 \subsetneq (\exists R\colon D)$, therefore, $d_2 \in (\exists R\colon D)^{\mathcal{I}}$ by definition and so there is an $e_2 \in \Delta^{\mathcal{I}}$ such that $d_2 R^{\mathcal{I}} e_2$, $e_2 \in D^{I}$ and $Y_1 \mathcal{B} e_2$. Let E be an arbitrary concept such that $Y_1 \subseteq D^{\mathcal{J}}$ and suppose that $e_2 \notin D^{\mathcal{I}}$. This should hold in particular for $D := C$, which leads to an absurdity. Thus $e_2 \in D^{\mathcal{I}}$.

Furthermore, $\mathcal{B}$ is a DL-Lite$_{R,\sqcap}$ simulation between $\{d_C \in \Delta^{\mathcal{J}} | \neg C \in \Gamma\}$ and d, hence for every $d_C \in \Delta^{\mathcal{J}}$, $\mathcal{J} \models \phi[d]$. Finally, given that, by assumption, ϕ is closed under DL-Lite simulations, $\mathcal{I} \models \phi[d]$. □

Lemma 3.2. *If A* FOL *formula ϕ is closed under* DL-Lite *simulations, then it is equivalent to a* DL-Lite *right hand or left hand side concept.*

Proof. We prove the lemma by induction on C:

- (Basis)
  - $C := A$ (basic concept). Let $\mathcal{I}, \mathcal{J}$ be two interpretations, $X_1 \subseteq \Delta^{\mathcal{I}}, d_2 \in \Delta^{\mathcal{J}}, \mathcal{B} \subseteq \mathcal{P}(\Delta^{\mathcal{I}}) \times \Delta^{\mathcal{J}}$ and assume that $X_1 \mathcal{B} d_2$. Let $d_1 \in X_1$ such that $d_1 \in C^{\mathcal{I}}$. Therefore it holds that $X_1 \subseteq A^{\mathcal{I}}$, whence (by definition) $d_2 \in A^{\mathcal{J}}$.
  - $C := \neg A$ (analogous argument).
  - $C := \exists R$ (unqualified existential). Make the same assumptions as before and suppose that $d_1 \in (\exists R)^{\mathcal{I}}$. Then there exists $e_1 \in \Delta^{\mathcal{I}}$ such that $d_1 R^{\mathcal{I}} e_1$, whence, by definition of DL-Lite simulations $\mathcal{B}$, there is an $e_2 \in \Delta^{\mathcal{J}}$ such that $d_2 R^{\mathcal{J}} e_2$, that is, such that $d_2 \in (\exists R)^{\mathcal{J}}$.
  - $C := \neg \exists R$ (analogous argument).

- (Inductive step)
  - $C := \exists R\colon D$ (qualified existential). Suppose that $d_1 \in (\exists R\colon D)^{\mathcal{I}}$ and that therefore there is some $e_1 \in \Delta^{\mathcal{I}}$ such that $e_1 \in D^{\mathcal{I}}$ and $d_1 R^{\mathcal{I}} e_2$. By induction hypothesis this impies that $e_2 \in D^{\mathcal{J}}$. Therefore $d_2 \in (\exists R\colon D)^{\mathcal{J}}$ as well.
  - $C := D \sqcap E$ (concept conjunction). By induction hypothesis the property holds for D and E. Now:

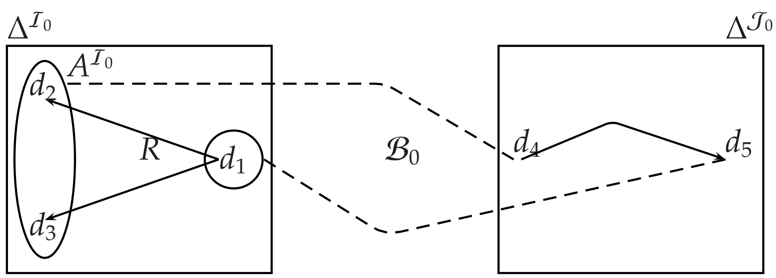
$$\begin{array}{rcl} d_1 \in (D \sqcap E)^{\mathcal{I}} & \text{iff} & d_1 \in D^{\mathcal{I}} \text{ and } d_1 \in E^{\mathcal{I}} \\ & \text{implies} & d_2 \in D^{\mathcal{J}} \text{ and } d_2 \in E^{\mathcal{J}} \\ & \text{iff} & d_2 \in (D \sqcap E)^{\mathcal{J}}, \end{array}$$

    which closes the proof. □

From this two lemmas, we immediately derive the result we wanted, namely:

Theorem 3.1. *A* FOL *formula ϕ is equivalent to a* DL-Lite$_{R,\sqcap}$ *right hand or left hand side concept iff it is closed under* DL-Lite *simulations.*

Example 3.1. As an example, we will again prove that COP+TV is not as expressive as DL-Lite$_{R,\sqcap}$ by restricting ourselves this time to formulas. Consider the following COP+TV formula: $\forall y R(x, y) \wedge A(y)$. This formula constitutes, intuitively, the MR of a universally quantified COP+TV VP and as we will soon see, it is not closed under DL-Lite$_{R,\sqcap}$ simulations.



As the reader may see, $\mathcal{B}_0$ is a DL-Lite simulation, $\mathcal{I}_0$ and $\mathcal{J}_0$ are interpretations, $\{d_1\} \subseteq \{d \in \Delta^{\mathcal{I}_0} | \mathcal{I}_0 \models \forall y R(x, y) \wedge A(y)[d]\}$, but $d_2 \notin \{d' \in \Delta^{\mathcal{J}_0} | \mathcal{J}_0 \models \forall y R(x, y) \wedge A(y)[d']\} = \emptyset$. ‡

4 Some Negative Results

As we said before, it is impossibly to characterize exactly the expressive power of DL-Lite when we move up to assertions, ABoxes and TBoxes. We can, at most, as we have done elsewhere, provide necessary conditions by way of closure properties that hold for its assertions (like closure under union of chains). This result is analogous to those obtained by Allan Third (cf. [11]).

Proposition 4.1. *Disjunction is not expressible in* DL-Lite

Proof. Suppose it is. Consider the FOL sentence $\phi := P(a) \vee Q(a)$. Let $\mathcal{H} = \langle \{a\}, \{P(a)\} \rangle$ and $\mathcal{H}' = \langle \{a\}, \{Q(a)\} \rangle$ be two Herbard models. Clearly $\mathcal{H} \models \phi$, $\mathcal{H}' \models \phi$ and $\mathcal{H} \neq \mathcal{H}'$. Futhermore, they are two minimal models of ϕ. Now, DL-Lite$_{R,\sqcap}$ is HORN, which means that it has a least Herbrand model (w.r.t. inclusion) and so should ϕ. But this is absurd. □

Theorem 4.1. *There is no invariance relation $\sim$ on interpretations such that, for any* FOL *sentence ϕ, ϕ is equivalent to a* DL-Lite$_{R,\sqcap}$ *assertion iff it is closed under the relation $\sim$.*

Proof. Suppose the contrary and consider the ABox assertion $P(a)$. Let $\mathcal{I}$ and $\mathcal{J}$ be two structures s.t. $\mathcal{I} \sim \mathcal{J}$ and suppose that $\mathcal{I} \models P(a)$. Then, obviously, $\mathcal{J} \models P(a)$ too. But then:

$$\mathcal{I} \models P(a) \text{ implies } \mathcal{I} \models P(a) \vee Q(a) \text{ and}$$
$$\mathcal{J} \models P(a) \text{ implies } \mathcal{J} \models P(a) \vee Q(a).$$

This makes sense because interpretations are nothing but FOL models. In other words, $P(a) \vee Q(a)$ is closed under $\sim$. But this is impossible, because disjunction is not expressible in DL-Lite$_{R,\sqcap}$. □

5 DL-Lite$_{R,\sqcap}$ and COP+Rel

As expected, DL-Lite$_{R,\sqcap}$ and COP+Rel MRs only overlap w.r.t. expressive power, since they both contain the MRs of COP. This is shown using, again, model-theoretic techniques. This makes sense because COP+Rel allows reducing entailment to satisfiability, which is **NP-Complete** – i.e. the same complexity class as for QA for DL-Lite$_{R,\sqcap}$, which is, esentially, that of deciding the entailment by a KB of a UCQ. For, ideed, QA can be captured to a certain extent by COP+Rel.

Theorem 5.1. *We have that:*

1. DL-Lite$_{R,\sqcap}$ *is not as expressive as* COP+Rel.

2. COP+Rel *is not as expressive as* DL-Lite$_{R,\sqcap}$.

Proof. For (1) Consider the COP+Rel sentence: "It is not the case that John is not a policeman who is not a man" whose MR is:

$$\neg(\neg Policeman(John) \wedge \neg Man(John)))$$

This FOL sentence is equivalent to $Policeman(John) \vee Man(John)$ which cannot be expressed in DL-Lite$_{R,\sqcap}$.

For (2) we recall that have shown elsewhere that COP is not as expressive as DL-Lite. The result follows immediately. □

Remark 5.1. As we said before QA can be captured to a certain extent by COP+Rel. For consider the following entailment:

$$\Gamma, \Delta \models \phi$$

where:

- Γ is a set of universally quantified FOL sentence i.e. belonging to the $\forall^*$ class.
- Δ is a set of ground atoms.
- ϕ is a positive existential FOL sentence.

These sentences correspond to COP+Rel MRs and can be expressed in DL-Lite$_{R,\sqcap}$ too. The complexity upper bound for reasoning in such a fragment of COP+Rel would thus be **LOGSPACE** too. †

Conclusions

As we have seen, an notion of simulation can be defined over DL-Lite$_{R,\sqcap}$ concepts, characterizing, so to speak, *modulo* this closure property, the classes of interpretations (models or FOL interpretation structures) that satisfy DL-Lite$_{R,\sqcap}$ concepts. It provides necessary and sufficient conditions. However, when we move to assertions, this is no longer the case and only necessary conditions can be achieved, due to DL-Lite$_{R,\sqcap}$'s properties as a fragment of FOL. In any case, semantics, model-theoretic properties of these are essential, we believe, for providing a more fine-grained analysis of the expressive power of DL-Lite$_{R,\sqcap}$ w.r.t. to Pratt's and Third's fragments, given that combined complexity for QA is untractable, which blurs the differences between this entailement problem and those of the intractable fragments of English. Furthermore, many of these results hold too for other members of the DL-Lite family.

References

[1] Raffaella Bernardi, Diego Calvanese, and Camilo Thorne. Expressing DL-Lite Ontologies with Controlled English. 2007.

[2] Raffaella Bernardi, Diego Calvanese, and Camilo Thorne. Lite Natural Language. In *Proceedings of the 7th International Workshop on Computational Semantics (IWCS-7)*, 2007.

[3] Alex Borgida. On the Relative Expressiveness of Description Logics and Predicate Logics. *Artificial Intelligence*, (82), 1996.

[4] Diego Calvanese, Alessandro Artale, Roman Kontchakov, and Michael Zakharyaschev. DL-Lite in the Light of First Order Logic. 2007.

[5] 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. *JAR*, 2007.

[6] Diego Calvanese, Giusseppe de Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Data Complexity of Query Answering in Description Logics. In *Proceedings of the 10th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2006)*, 2006.

[7] Diego Calvanese, Giusseppe de Giacomo, and Maurizio Lenzerini. Conjunctive Query Containment Under Description Logic Constraints. http://www.inf.unibz.it/čalvanese/publications.shtml, 2006.

[8] René Cori and Daniel Lascar. *Logique mathématique (2 vols)*. Dunod, 2003.

[9] Natasha Kurtonina and Marteen de Rijke. Expressiveness of First-Order Description Logics. Technical report, Warwick University, 1997. http://citeseer.ist.psu.edu/kurtonina97expressiveness.html.

[10] Ian PRATT and Allan THIRD. More Fragments of Language. *Notre Dame Journal of Formal Logic*, 2005.

[11] Allan THIRD. *Logical Analysis of Fragments of Natural Language*. PhD thesis, Faculty of Engineering and Physical Sciences, University of Manchester, 2006.

[12] Camilo THORNE. Controlled English for DL-Lite. Technical report, KRDB Research Centre, Free University of Bozen-Bolzano, 2007. http://www.inf.unibz.it/krdb/pub/index.php.