$\mathcal{E}$-connections of Description Logics

Oliver Kutz
Department of Computer Science
University of Liverpool, U.K.
oliver@csc.liv.ac.uk

Carsten Lutz
Institute for Theoretical Computer Science
TU Dresden, Germany
lutz@tcs.inf.tu-dresden.de

Frank Wolter
Department of Computer Science
University of Liverpool, U.K.
frank@csc.liv.ac.uk

Michael Zakharyaschev
Department of Computer Science
King's College London, U.K.
mz@dcs.kcl.ac.uk

1 Motivation

Combining description logics (DLs) with other logical formalisms, including other DLs, is an important and challenging task. Recent examples include:

1. the combination of DLs with temporal logics to form *multi-dimensional temporal DLs*, cf. e.g. [1, 6, 10];

2. the *fusion* of multiple DLs into a single formalism that inherits decidability from its components, cf. [2];

3. the combination of different DLs in the context of loosely federated information systems resulted in *distributed description logics* (DDLs), cf. [4].

A relatively new technique of combining description logics, first proposed in [9], is the formation of so-called $\mathcal{E}$*-connections*. The general idea behind this combination method is that the interpretation domains of the connected logics are disjoint and interconnected by means of *link relations*. The language of the $\mathcal{E}$-connection is then the union of the original languages enriched with operators capable of talking about the link relations. To illustrate this idea, let us consider the $\mathcal{E}$-connection of the two description logics $\mathcal{ALCQI}$ and $\mathcal{ALCIO}$.

Assume that we have two knowledge bases: one deals with people and uses $\mathcal{ALCQI}$; the other deals with countries and employs $\mathcal{ALCIO}$. Note that such a scenario is quite natural: it may be the case that the two knowledge bases have been developed independently and are now required to interoperate—this situation is standard for loosely

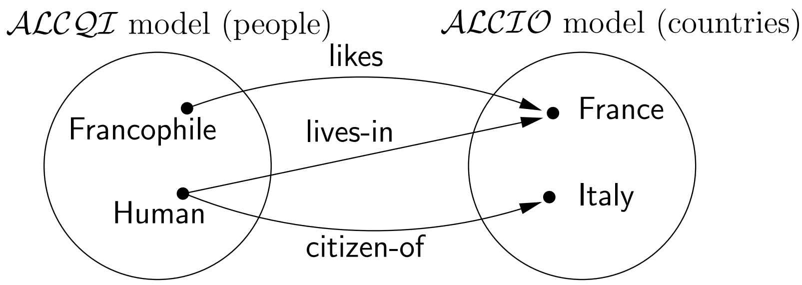

Figure 1: The connection of two knowledge bases.

federated information systems [4] and also for ontology integration in the semantic web [5]. Another reason for separating the two KBs could be that $\mathcal{ALCQI}$ is an appropriate language for representing people's affairs, $\mathcal{ALCIO}$ is appropriate for talking about countries—but their union $\mathcal{ALCQIO}$ is very hard to handle algorithmically.

The KBs can be integrated by using binary *link relations* such as $\mathsf{citizen\text{-}of}$, $\mathsf{lives\text{-}in}$, and likes, which relate domain objects from models of one KB with domain objects from models of the other KB. In the $\mathcal{E}$-connection of $\mathcal{ALCQI}$ and $\mathcal{ALCIO}$, we can, for example, say that francophile people like France:

$$\mathsf{Francophile} \doteq \langle \mathsf{likes} \rangle^1 \mathsf{France}$$

where $\mathsf{Francophile}$ is a concept name from the $\mathcal{ALCQI}$ component, the $\langle \mathsf{r} \rangle^1 C$ operator is one of the connection operators for talking about link relations, and France is a nominal of $\mathcal{ALCIO}$. Intuitively, the $\langle \mathsf{r} \rangle^1 C$ operator can be understood as an existential value restriction. The $\cdot^1$ indicates that this operator is applied to a concept of Logic 2 ($\mathcal{ALCIO}$) and returns a concept of Logic 1 ($\mathcal{ALCQI}$). In first-order logic, the above formula would read as

$$\forall x \in W_1(\mathsf{Francophile}(x) \leftrightarrow \exists y \in W_2(\mathsf{likes}(x, y) \wedge \mathsf{France}(y))),$$

where W_1 is the domain of an $\mathcal{ALCQI}$ model and W_2 is the domain of an $\mathcal{ALCIO}$ model. Of course, we can also use link relations in the other direction:

$$\mathsf{France} \sqsubseteq \langle \mathsf{lives\text{-}in} \rangle^2 (\mathsf{Human} \sqcap \neg \langle \mathsf{citizen\text{-}of} \rangle^1 \mathsf{France})$$

expresses that not all people living in France are French citizens; see Figure 1. Again, the reading of this formula in first-order logic would be

$$\begin{aligned}&\forall y \in W_2(\mathsf{France}(y) \rightarrow \\ &\quad \rightarrow \quad \exists x \in W_1(\mathsf{lives\text{-}in}(x, y) \wedge \mathsf{Human}(x) \wedge \neg(\exists z \in W_2(\mathsf{citizen\text{-}of}(x, z) \wedge \mathsf{France}(z))))).\end{aligned}$$

The most important feature of $\mathcal{E}$-connections is that, just as DLs themselves, they offer an appealing compromise between expressive power and computational complexity: although powerful enough to express many interesting concepts, the coupling between the combined logics is sufficiently loose for proving general results about the transfer of decidability. Such transfer results state that if the connected logics are decidable, then their connection will also be decidable. Thus, $\mathcal{E}$-connections are closer in spirit to fusions than to multi-dimensional combinations: while there exist general transfer results for the former [2], the latter allow such a close interaction between the combined formalisms that general transfer results cannot be expected; see, e.g., [6].

The purpose of this paper is *to summarize the general transfer results for $\mathcal{E}$-connections that have recently been obtained in* [7]. The generality of the results is due to the fact that $\mathcal{E}$-connections are defined and investigated using the framework of so-called *abstract description systems* (ADSs), a common generalization of description logics, modal logics, logics of time and space, and many other logical formalisms [2]. Thus, we can connect not only DLs with DLs, but also, say, description logics with spatial logics [8]. A natural interpretation of link relations in this context would then be, for instance, to describe the spatial extension of abstract (DL) objects.

2 Basic $\mathcal{E}$-connections

In this section, we introduce the basic variant of $\mathcal{E}$-connections and formulate the fundamental transfer theorem first proved in [9]. We begin by introducing ADSs. For brevity, we give here a slightly trimmed-down version of ADSs that does not capture ABoxes. However, all results presented in this paper do also apply to ADSs with ABoxes as defined in [2, 7].

Definition 1 An *abstract description language* (ADL) $\mathcal{L}$ is determined by a countably infinite set $\mathcal{V}$ of *set variables* and a countable set $\mathcal{F}$ of *function symbols* f of arity m_f such that $\neg, \wedge \notin \mathcal{F}$. The *terms* t_j of $\mathcal{L}$ are built in the following way:

$$t_j ::= x \mid \neg t_1 \mid t_1 \wedge t_2 \mid f(t_1, \ldots, t_{m_f}),$$

where $x \in \mathcal{V}$ and $f \in \mathcal{F}$. The *term assertions* of $\mathcal{L}$ are of the form $t_1 \sqsubseteq t_2$. As usual, we use $t_1 \doteq t_2$ as an abbreviation for $t_1 \sqsubseteq t_2,\ t_2 \sqsubseteq t_1$.

An *abstract description model* (ADM) for an ADL $\mathcal{L} = \langle \mathcal{V}, \mathcal{F} \rangle$ is a structure of the form

$$\mathfrak{W} = \left\langle W, \mathcal{V}^{\mathfrak{W}} = (x^{\mathfrak{W}})_{x \in \mathcal{V}}, \mathcal{F}^{\mathfrak{W}} = (f^{\mathfrak{W}})_{f \in \mathcal{F}} \right\rangle,$$

where W is a non-empty set, $x^{\mathfrak{W}} \subseteq W$ and each $f^{\mathfrak{W}}$ is a function mapping m_f-tuples $\langle X_1, \ldots, X_{n_f} \rangle$ of subsets of W to a subset of W. The *value* $t^{\mathfrak{W}} \subseteq W$ of an $\mathcal{L}$-term t in $\mathfrak{W}$ is defined inductively by taking

$$(\neg t)^{\mathfrak{W}} = W \setminus t^{\mathfrak{W}}, \quad (t_1 \wedge t_2)^{\mathfrak{W}} = t_1^{\mathfrak{W}} \cap t_2^{\mathfrak{W}}, \quad (f(t_1, \ldots, t_{m_f}))^{\mathfrak{W}} = f^{\mathfrak{W}}(t_1^{\mathfrak{W}}, \ldots, t_{m_f}^{\mathfrak{W}}).$$

Intuitively, set variables correspond to concept names, function symbols to concept constructors, and term assertions to general concept inclusion axioms (GCIs). ADSs become a powerful tool in providing a choice of an appropriate class of ADMs in which the ADL is to be interpreted. In this way, we can ensure that function symbols have the desired semantics.

Definition 2 An *abstract description system* (ADS) is a pair $(\mathcal{L}, \mathcal{M})$, where $\mathcal{L}$ is an ADL and $\mathcal{M}$ is a class of ADMs for $\mathcal{L}$ that is closed under the following operation: if $\mathfrak{W} = \langle W, \mathcal{V}^{\mathfrak{W}}, \mathcal{F}^{\mathfrak{W}} \rangle$ and $\mathcal{V}^{\mathfrak{W}'} = (x^{\mathfrak{W}'})_{x \in \mathcal{V}}$ is a new assignment of set variables in W, then $\mathfrak{W}' = \left\langle W, \mathcal{V}^{\mathfrak{W}'}, \mathcal{F}^{\mathfrak{W}} \right\rangle \in \mathcal{M}$.

The closure condition on the class of models $\mathcal{M}$ demands that set variables (i.e., concept names) can be interpreted as arbitrary subsets of the interpretation domain—a property that all DLs comply with. It should be noted that ADSs can capture all standard expressive means such as number restrictions, transitive closure of roles, and concrete domains. A very detailed description of how standard DLs can be conceived as ADSs can be found in [2]. Here we will only briefly describe the translations of the basic description logic $\mathcal{ALC}$ into an ADS, as well as its extension by nominals, $\mathcal{ALCO}$. Again, we omit the discussion of ABox assertions for brevity.

The language of $\mathcal{ALC}$ is based on concept names $A_1, A_2, \ldots$, role names $R_1, R_2, \ldots$, the Boolean constructors $\neg$ and $\sqcap$, and the existential restriction $\exists$. $\mathcal{ALC}$-concepts are built according to the following rule:

$$C ::= A_i \mid \neg C \mid C \sqcap D \mid \exists R.C$$

An $\mathcal{ALC}$-model is a structure

$$\mathcal{I} = \left\langle \Delta, A_1^{\mathcal{I}}, \ldots, R_1^{\mathcal{I}}, \ldots \right\rangle,$$

where Δ is a non-empty set, the $A_i^{\mathcal{I}}$ are subsets of Δ and the $R_i^{\mathcal{I}}$ are binary relations on Δ. The interpretation of complex concepts is defined by setting:

$$\begin{array}{c} (\neg C)^{\mathcal{I}} \;=\; \Delta \setminus C^{\mathcal{I}} \qquad (C \sqcap D)^{\mathcal{I}} \;=\; C^{\mathcal{I}} \sqcap D^{\mathcal{I}} \\ (\exists R.C)^{\mathcal{I}} = \{w \in \Delta \mid \exists v \; ((w, v) \in R^{\mathcal{I}} \wedge v \in C^{\mathcal{I}})\} \end{array}$$

The concepts of $\mathcal{ALC}$ can be regarded as terms $C^\sharp$ of an ADS $\mathcal{ALC}^\sharp$: associate with each concept name A_i a set variable $A_i^\sharp$, and with each role name R_i a unary function symbols $f_{\exists R_i}$. Then set inductively:

$$\begin{array}{c} (\neg C)^\sharp \;=\; \neg C^\sharp \qquad (C \sqcap D)^\sharp \;=\; C^\sharp \wedge D^\sharp \\ (\exists R_i.C)^\sharp = f_{\exists R_i}(C^\sharp) \end{array}$$

Thus, $\mathcal{ALC}^\sharp$-term assertions correspond to concept inclusion statements. The class $\mathcal{M}$ of ADMs for $\mathcal{ALC}^\sharp$ is defined as follows. For every $\mathcal{ALC}$-model $\mathcal{I} = \left\langle \Delta, A_1^{\mathcal{I}}, \ldots, R_1^{\mathcal{I}}, \ldots \right\rangle$, the class $\mathcal{M}$ contains the model

$$\mathfrak{M} = \left\langle \Delta, \mathcal{V}^{\mathfrak{M}}, \mathcal{F}^{\mathfrak{M}} \right\rangle,$$

where, for every concept name A and role name R, we have

$$\begin{array}{rcl} (A^\sharp)^{\mathfrak{M}} & = & A^{\mathcal{I}} \\ f_{\exists R}^{\mathfrak{M}}(X) & = & \{w \in \Delta \mid \exists v \; ((w, v) \in R^{\mathcal{I}} \wedge v \in X)\} \end{array}$$

Observe that the semantics of the function symbol $f_{\exists R}$ is obtained in a straightforward way from the semantics of the DL constructor $\exists R.C$.

Next, we discuss the addition of nominals. The description logic $\mathcal{ALCO}$ extends $\mathcal{ALC}$ with nominals n_i [12] that are always interpreted by singleton subsets of the interpretation domain, but syntactically treated as concepts. Due to this special property of nominals and the closure property on set variables in ADS (Definition 2), nominals cannot be translated as set variables. Rather, the corresponding ADS

$\mathcal{ALCO}^\sharp$ is obtained from the above translation of $\mathcal{ALC}$ by additionally introducing, for every nominal n_i of $\mathcal{ALCO}$, the *nullary* function symbol f_{n_i} with $f_{n_i}^{\mathfrak{W}} = n_i^{\mathcal{I}}$ and by setting $n_i^\sharp = f_{n_i}$.

In the following, we will not distinguish between a description logic and the corresponding ADS. We are interested in the following satisfiability problem.

Definition 3 Let $\mathcal{S} = (\mathcal{L}, \mathcal{M})$ be an ADS, t an $\mathcal{S}$-term, and Γ a finite set of term assertions. Then t is called *satisfiable* relative to Γ if there exists an ADM $\mathfrak{W} \in \mathcal{M}$ such that $t^{\mathfrak{W}} \neq \emptyset$ and $t_1^{\mathfrak{W}} \subseteq t_2^{\mathfrak{W}}$ for all $t_1 \sqsubseteq t_2 \in \Gamma$.

It is not hard to see that this corresponds to the satisfiability of concepts with respect to general TBoxes. Indeed, the presented transfer results do only apply to DLs for which reasoning with respect to general TBoxes is decidable.

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two ADSs that are to be connected.[1] We assume that the set variables and non-Boolean functions symbols of $\mathcal{S}_1$ and $\mathcal{S}_2$ are pairwise disjoint. To form a connection, fix a non-empty set $\mathcal{E} = \{E_j \mid j \in J\}$ of binary relation symbols. The set of terms of the resulting $\mathcal{E}$-connection $\mathcal{C}^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$ is partitioned into a set of 1-terms and a set of 2-terms. Intuitively, i-terms are the terms of $\mathcal{L}_i$ enriched with new function symbols for talking about link relations. For the following definition, we set $\overline{1} = 2$ and $\overline{2} = 1$.

Definition 4 The sets of 1-*terms* and 2-*terms* of $\mathcal{C}^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$ are defined by simultaneous induction: for $i \in \{1, 2\}$,

- every set variable of $\mathcal{L}_i$ is an i-term;
- the set of i-terms is closed under $\neg$, $\wedge$, and the function symbols of $\mathcal{L}_i$;
- if t is an $\overline{i}$-term, then the expression $\langle E_j \rangle^i t$ is an i-term, for every $j \in J$.

The set of *terms* of $\mathcal{C}^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$ is the union of the set of 1-terms and the set of 2-terms. The *term assertions* of $\mathcal{C}^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$ are of the form $t_1 \sqsubseteq t_2$, where both t_1 and t_2 are i-terms, for $i \in \{1, 2\}$.

As expected, a model for the $\mathcal{E}$-connection $\mathcal{C}^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$ consists of a model for $\mathcal{S}_1$, a model for $\mathcal{S}_2$, and an interpretation of the link relations.

Definition 5 A structure $\mathfrak{M} = \left\langle \mathfrak{W}_1, \mathfrak{W}_2, \mathcal{E}^{\mathfrak{M}} = (E_j^{\mathfrak{M}})_{j \in J} \right\rangle$, where $\mathfrak{W}_i \in \mathcal{M}_i$ for $i \in \{1, 2\}$ and $E_j^{\mathfrak{M}} \subseteq W_1 \times W_2$ for each $j \in J$, is called a *model for* $\mathcal{C}^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$. The *value* $t^{\mathfrak{M}} \subseteq W_i$ of an i-term t is defined by simultaneous induction. For set variables X of $\mathcal{L}_i$, we put $X^{\mathfrak{M}} = X^{\mathfrak{W}_i}$; the inductive steps for the Booleans and function symbols of $\mathcal{L}_i$ are the same as in Definition 1; finally,

$$
\begin{aligned}
(\langle E_j \rangle^1 t)^{\mathfrak{M}} &= \{x \in W_1 \mid \exists y \in t^{\mathfrak{M}}\ (x, y) \in E_j^{\mathfrak{M}}\}, \\
(\langle E_j \rangle^2 t)^{\mathfrak{M}} &= \{x \in W_2 \mid \exists y \in t^{\mathfrak{M}}\ (y, x) \in E_j^{\mathfrak{M}}\}.
\end{aligned}
$$

We are now ready to formulate the fundamental transfer result mentioned above:

[1] In general, $\mathcal{E}$-connections can connect $n < \omega$ ADSs [7], and all the formulated results apply to the n-dimensional case as well.

Theorem 6 *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be ADSs with decidable satisfiability problems. Then the satisfiability problem for every $\mathcal{E}$-connection $\mathcal{C}^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$ is decidable as well.*

It is of interest to note that this transfer theorem is more general than the corresponding theorem for fusions obtained in [3]. The transfer result for fusions only applies to ADSs whose class of models is closed under disjoint unions—thus ruling out description logics with nominals. This is not the case for the above result: it means, in particular, that the connection of $\mathcal{ALCQI}$ and $\mathcal{ALCIO}$ mentioned in the introduction is decidable.

Theorem 6 is proved by a reduction to the satisfiability problems for the component ADSs. Since this reduction is non-deterministic and involves an exponential blow-up, we obtain an upper complexity bound for the $\mathcal{E}$-connection that is one non-deterministic exponential higher than the complexity of the component logics. It is currently unknown whether this complexity is optimal in the general case. However, it seems that in many natural cases the increase in complexity will be less dramatic.

3 Extensions

The basic idea of connecting logics by means of link relations can be extended in various directions. For example, in the distributed KB example given in the introduction we may want to describe people living in a country of which they are not citizens, or people who like all countries. To do this, basic $\mathcal{E}$-connections are not enough, since Boolean operations on link relations are required:

$$\begin{aligned} \mathsf{Expat} &\doteq \langle \mathsf{lives\text{-}in} \cap \neg \mathsf{citizen\text{-}of} \rangle^1 \, \mathsf{Country} \\ \mathsf{Internationalist} &\doteq \neg \langle \neg \mathsf{likes} \rangle^1 \mathsf{Country} \end{aligned}$$

The $\mathcal{E}$-connection of two ADSs $\mathcal{S}_1$ and $\mathcal{S}_2$ that admits the Boolean operators on link relations is denoted by $\mathcal{C}^{\mathcal{E}}_B(\mathcal{S}_1, \mathcal{S}_2)$. Since we deal with Boolean connections in some more detail, let us give the precise definition:

Definition 7 Suppose that $\mathcal{S}_i = (\mathcal{L}_i, \mathcal{M}_i)$, $i \in \{1, 2\}$, are abstract description systems and $\mathcal{E} = \{E_j \mid j \in J\}$ is a set of binary relation symbols. Denote by $\mathcal{C}^{\mathcal{E}}_B(\mathcal{S}_1, \mathcal{S}_2)$ the $\mathcal{E}$-connection with the smallest set $\overline{\mathcal{E}}$ of links such that

- $\mathcal{E} \subseteq \overline{\mathcal{E}}$;
- if $F \in \overline{\mathcal{E}}$, then $\neg F \in \overline{\mathcal{E}}$;
- if $F, G \in \overline{\mathcal{E}}$, then $F \wedge G \in \overline{\mathcal{E}}$.

Given an ADM $\mathfrak{M} = \left\langle (\mathfrak{W}_i)_{i \in \{1,2\}}, \mathcal{E}^{\mathfrak{M}} \right\rangle$ we interpret the links $F \in \overline{\mathcal{E}}$ as relations $F^{\mathfrak{M}} \subseteq W_1 \times W_2$ (with W_i being the domain of $\mathfrak{W}_i$) inductively in the obvious way:

$$(F \wedge G)^{\mathfrak{M}} = F^{\mathfrak{M}} \cap G^{\mathfrak{M}}, \quad (\neg F)^{\mathfrak{M}} = (W_1 \times W_2) \setminus F^{\mathfrak{M}}.$$

Observe that role hierarchies on link relations can be expressed by writing, e.g., $\top_i \sqsubseteq \neg \langle F \cap \neg G \rangle^1 \top_2$ for $F \sqsubseteq G$, where $\top_i = x_i \vee \neg x_i$ for some set variable x_i of $\mathcal{L}_i$. Fortunately, our general transfer result carries over to the Boolean case:

Theorem 8 *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be ADSs with decidable satisfiability problems. Then the satisfiability problem for every $\mathcal{E}$-connection $\mathcal{C}^{\mathcal{E}}_B(\mathcal{S}_1, \mathcal{S}_2)$ is decidable as well.*

The proof is similar to the basic case, although much more involved. The complexity of the obtained algorithm is also as in the basic case. Interestingly, in the Boolean case we are able to prove that the obtained complexity bound is optimal. Let $\mathcal{B}$ be the ADS that has no function symbols apart from the Booleans and whose class of ADMs is not restricted in any way. It is not hard to see that, for this simple ADS, satisfiability is NP-complete. The Boolean connection $\mathcal{C}_B(\mathcal{B}, \mathcal{B})$ of $\mathcal{B}$ with itself, however, is much more complex: it is possible to reduce the NEXPTIME-complete satisfiability problem for the modal logic $\mathbf{S5} \times \mathbf{S5}$ [11] to the satisfiability problem for $\mathcal{C}_B(\mathcal{B}, \mathcal{B})$, which yields the following result:

Theorem 9 *The satisfiability problem for $\mathcal{C}^{\mathcal{E}}_B(\mathcal{B}, \mathcal{B})$ is* NEXPTIME*-hard, for any infinite $\mathcal{E}$.*

To illustrate the expressive power of Boolean connections, we sketch the proof of this theorem.

First, recall that $\mathbf{S5} \times \mathbf{S5}$–formulas are composed from propositional variables $p_1, p_2 \ldots$ by means of the Booleans and the modal operators $\Box_1$ and $\Box_2$. $\mathbf{S5} \times \mathbf{S5}$-models $\mathfrak{N} = \langle W_1 \times W_2, \mathfrak{V} \rangle$ consist of the Cartesian product of two non-empty sets W_1 and W_2 and a valuation $\mathfrak{V}$ which maps any propositional variable to a subset of $W_1 \times W_2$. The extension $\varphi^{\mathfrak{N}}$ of an $\mathbf{S5} \times \mathbf{S5}$-formula φ in $\mathfrak{N}$ is computed inductively by setting

$$
\begin{array}{rcl}
p_i^{\mathfrak{N}} & = & \mathfrak{V}(p_i) \\
(\psi_1 \wedge \psi_2)^{\mathfrak{N}} & = & \psi_1^{\mathfrak{N}} \cap \psi_2^{\mathfrak{N}} \\
(\neg\psi)^{\mathfrak{N}} & = & (W_1 \times W_2) - \psi^{\mathfrak{N}} \\
(\Box_1\psi)^{\mathfrak{N}} & = & \{(w_1, w_2) \mid \forall v \in W_1 \ (v, w_2) \in \psi^{\mathfrak{N}}\} \\
(\Box_2\psi)^{\mathfrak{N}} & = & \{(w_1, w_2) \mid \forall v \in W_2 \ (w_1, v) \in \psi^{\mathfrak{N}}\}.
\end{array}
$$

A formula φ is $\mathbf{S5} \times \mathbf{S5}$-satisfiable if there exists an $\mathbf{S5} \times \mathbf{S5}$-model in which φ has a non-empty extension.

Suppose now that φ is an $\mathbf{S5} \times \mathbf{S5}$-formula. Denote by $sub(\varphi)$ the set of all subformulas of φ. For any $\psi \in sub(\varphi)$ take a link $E_\psi \in \mathcal{E}$ and let the $\mathcal{C}^{\mathcal{E}}_B(\mathcal{B}, \mathcal{B})$-knowledge base Γ consist of:

(1) $E_{\psi_1 \wedge \psi_2} = E_{\psi_1} \wedge E_{\psi_2}$, for $\psi_1 \wedge \psi_2 \in sub(\varphi)$,

(2) $E_{\neg\psi} = \neg E_\psi$, for $\neg\psi \in sub(\varphi)$;

(3) $\langle \neg E_\psi \rangle^2 \top_1 = [E_{\Box_1\psi}]^2 \bot_1$ and $[E_{\Box_1\psi}]^2 \bot_1 = \left\langle \neg E_{\Box_1\psi} \right\rangle^2 \top_1$, for $\Box_1\psi \in sub(\varphi)$;

(4) $\langle \neg E_\psi \rangle^1 \top_2 = [E_{\Box_2\psi}]^1 \bot_2$ and $[E_{\Box_2\psi}]^1 \bot_2 = \left\langle \neg E_{\Box_2\psi} \right\rangle^1 \top_2$, for $\Box_2\psi \in sub(\varphi)$.

As was mentioned above, such equations can be added to the vocabulary when working in connections with Boolean closures of links. More precisely, an equation of the form $F = G$ is a shorthand for the conjunction of the two link inclusions $F \sqsubseteq G$ and $G \sqsubseteq F$. We now claim that

($\clubsuit$) φ is $\mathbf{S5} \times \mathbf{S5}$-satisfiable iff $\langle E_\varphi \rangle^1 \top_2$ is satisfiable relative to Γ in $\mathcal{C}^{\mathcal{E}}_B(\mathcal{B}, \mathcal{B})$.

To prove (♣), assume first that φ is satisfied in a model $\mathfrak{N} = \langle W_1 \times W_2, \mathfrak{V}\rangle$. We construct a model $\mathfrak{M} = \left\langle \mathfrak{M}_1, \mathfrak{M}_2, \{E_\psi^{\mathfrak{M}}\}_{\psi \in sub(\varphi)} \right\rangle$ that satisfies $\langle E_\varphi \rangle^1 \top_2$ relative to Γ. Let $\mathfrak{M}_2$ be any model for $\mathcal{B}$ with domain W_2. By assumption, $\varphi^{\mathfrak{N}} \neq \emptyset$, so we can pick some $(u, v) \in \varphi^{\mathfrak{N}}$ and choose $\mathfrak{M}_1$ to be any model for $\mathcal{B}$ with domain W_1. Finally, we can define $E_\psi^{\mathfrak{M}} := \psi^{\mathfrak{N}} \subseteq W_1 \times W_2$, for every $\psi \in sub(\varphi)$. By construction, $(\langle E_\varphi \rangle^1 \top_2)^{\mathfrak{M}} \neq \emptyset$, so it suffices to show that the equations (1)–(4) hold in $\mathfrak{M}$, which can easily be shown by structural induction; details can be found in [7].

Conversely, assume that $\langle E_\varphi \rangle^1 \top_2$ is satisfied relative to Γ in a model $\mathfrak{M}$, where $\mathfrak{M} = \left\langle \mathfrak{M}_1, \mathfrak{M}_2, \{E_\psi^{\mathfrak{M}}\}_{\psi \in sub(\varphi)} \right\rangle$ is based on the domains W_1 and W_2. We define a model $\mathfrak{N}$ for $\mathbf{S5} \times \mathbf{S5}$ based on the domain $W_1 \times W_2$ by letting $p_i^{\mathfrak{N}} := E_{p_i}^{\mathfrak{M}}$, for $p_i \in sub(\varphi)$, and arbitrary otherwise. It can now be shown by induction that

$$(\heartsuit) \quad E_\psi^{\mathfrak{M}} = \psi^{\mathfrak{N}}, \quad \text{for all} \quad \psi \in sub(\varphi).$$

Again, the details of this induction can be found in [7].

Since $\langle E_\varphi \rangle^1 \top_2$ is satisfiable in $\mathfrak{M}$, there exists a $v \in W_1$ and a $w \in W_2$ such that $(v, w) \in E_\varphi^{\mathfrak{M}} = \varphi^{\mathfrak{N}} \neq \emptyset$. It follows that φ is satisfied in $\mathfrak{N}$ and hence proves (♣).

The reduction shows that the satisfiability problem of $\mathcal{C}_B^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$ is at least NEXPTIME-hard for most interesting ADSs $\mathcal{S}_1$ and $\mathcal{S}_2$.

Another interesting way of extending basic $\mathcal{E}$-connections is to add qualified number restrictions on link relations. Suppose, for example, that we want to describe persons who are citizens of exactly one country. Then it would obviously be convenient to write

$$\mathsf{Uni\text{-}National} \doteq \langle \mathsf{citizen\text{-}of} \rangle^1 \, \mathsf{Country} \wedge \langle \leq 1 \ \mathsf{citizen\text{-}of} \rangle^1 \top_2$$

where the semantics of $\langle \leq r\ E \rangle^i C$ and its counterpart $\langle \geq r\ E \rangle^i C$ are defined as for standard qualified number restrictions in DL. The $\mathcal{E}$-connection of two ADSs $\mathcal{S}_1$ and $\mathcal{S}_2$ that allows qualified number restrictions (but not the Boolean operators on link relations) is denoted by $\mathcal{C}_Q^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$. Unfortunately, it turns out that, in general, decidability does not transfer from two ADSs $\mathcal{S}_1, \mathcal{S}_2$ to their $\mathcal{E}$-connection $\mathcal{C}_Q^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$.

Theorem 10 *There exist ADSs $\mathcal{S}_1$ and $\mathcal{S}_2$ with decidable satisfiability problems such that the satisfiability problem for $\mathcal{C}_Q^{\mathcal{E}}(\mathcal{S}_1, \mathcal{S}_2)$ is undecidable even if $\mathcal{E}$ is a singleton.*

Although to prove this theorem we use rather artificial ADSs, there is an intuitive reason for this 'negative' result: number restrictions on links allow the transfer of 'counting capabilities' from one component to another. For example, in the connection $\mathcal{C}_Q^{\mathcal{E}}(\mathcal{ALCQI}, \mathcal{ALCIO})$, we can 'export' the nominals of $\mathcal{ALCIO}$ to $\mathcal{ALCQI}$: the assertions

$$\top_2 = \langle \leq 1 E \rangle^2 \top_1, \quad \top_2 = \langle \geq 1 E \rangle^2 \top_1, \quad \top_1 = \langle \leq 1 E \rangle^1 \top_2, \quad \top_1 = \langle \geq 1 E \rangle^1 \top_2$$

state that E is a bijective function, and so we can use $\langle E \rangle^1 N$, N a nominal of $\mathcal{ALCIO}$, as a nominal in $\mathcal{ALCQI}$. To obtain a general transfer result, we thus have to restrict the class of ADSs we are working with. For a set of term assertions Γ, we use $term(\Gamma)$ to denote the set of (sub)terms occurring in Γ.

Definition 11 An ADS $\mathcal{S} = (\mathcal{L}, \mathcal{M})$ is called *number tolerant* if there is a cardinal κ such that, for every $\kappa' \geq \kappa$ and every satisfiable finite set Γ of term assertions, there exists a model $\mathfrak{W} \in \mathcal{M}$ satisfying Γ and such that, for each $d \in W$, there are precisely κ' elements $d' \in W$ for which

$$\{t \in term(\Gamma) \mid d \in t^{\mathfrak{W}}\} = \{t \in term(\Gamma) \mid d' \in t^{\mathfrak{W}}\}.$$

Intuitively, DLs that provide means for 'global counting' such as nominals are not number tolerant, whereas those that can only 'locally count' are: for example, $\mathcal{ALCQI}$ is number tolerant, while $\mathcal{ALCIO}$ is not. More details can be found in [7].

Theorem 12 *Let $\mathcal{S}_1, \mathcal{S}_2$ be number-tolerant ADSs with decidable satisfiability problems. Then the satisfiability problem for any $\mathcal{E}$-connection $\mathcal{C}^{\mathcal{E}}_{Q}(\mathcal{S}_1, \mathcal{S}_2)$ is decidable as well.*

Again, the proof is a variation on the initial idea of the proof of Theorem 6, though much more complex. It is now a natural question whether we can combine the Boolean operators with qualified number restrictions on link relations and, at least for number tolerant ADSs, obtain a general transfer result. Unfortunately, the answer to this question is negative:

Theorem 13 *There exist number tolerant ADSs $\mathcal{S}_1, \mathcal{S}_2$ with decidable satisfiability problems such that the satisfiability problem for $\mathcal{C}^{\mathcal{E}}_{QB}(\mathcal{S}_1, \mathcal{S}_2)$ is undecidable even if $\mathcal{E}$ is a singleton.*

4 Further results

In this paper, we have presented a brief overview of our recent results on $\mathcal{E}$-connections. More details and full proofs can be found in [7], where also several additional results are proved. Here we mention only two of them:

(1) As already noted, the results in [7] are more general than those presented here in that they take into account ADSs with ABoxes. Moreover, another extension of basic $\mathcal{E}$-connections is considered, in which ABox individuals may occur as arguments of a connection operator even if nominals are *not* provided by the connected logics. Quite surprisingly, we can still prove a general transfer result in the spirit of Theorem 6. The combination of this extension with Boolean operators on link relations poses no problems, whereas the combination with qualified number restrictions leads to undecidability.

(2) There exists a close connection between $\mathcal{E}$-connections and distributed description logics (DDLs) considered by Borgida and Serafini in [4]. Indeed, the extension of $\mathcal{E}$-connections mentioned in (1) can almost be viewed as a generalization of DDLs. We say 'almost' because DDLs are able to express that an ABox individual of one logic is connected via a certain link relation to *exactly* the objects $b_1, \ldots, b_k$ of another logic. In [7], we extend $\mathcal{E}$-connections with this expressive means, prove a general transfer result for the case when all connected DLs are equipped with nominals, and show that, in the general case, decidability does not transfer.

References

[1] A. Artale, E. Franconi, F. Wolter, and M. Zakharyaschev. A temporal description logic for reasoning about conceptual schemas and queries. In *Proceedings of JELIA-02*, volume 2424 of *LNAI*, pages 98–110. Springer, 2002.

[2] F. Baader, C. Lutz, H. Sturm, and F. Wolter. Fusions of description logics. In F. Baader and U. Sattler, editors, *Proceedings of DL2000*, number 33 in CEUR-WS (http://ceur-ws.org/), pages 21–30, 2000.

[3] F. Baader, C. Lutz, H. Sturm, and F. Wolter. Fusions of description logics and abstract description systems. *Journal of Artificial Intelligence Research (JAIR)*, 16:1–58, 2002.

[4] A. Borgida and L. Serafini. Distributed description logics: Directed domain correspondences in federated information sources. In *On The Move to Meaningful Internet Systems 2002: CoopIS, Doa, and ODBase*, volume 2519 of *LNCS*, pages 36–53. Springer, 2002.

[5] D. Calvanese, G. Giacomo, and M. Lenzerini. A framework for ontology integration. In *The Emerging Semantic Web — Selected Papers from the First Semantic Web Working Symposium*, pages 201–214. IOS Press, 2002.

[6] D. Gabbay, A. Kurucz, F. Wolter, and M. Zakharyaschev. *Many-Dimensional Modal Logics: Theory and Applications*. Elsevier, North-Holland, 2003.

[7] O. Kutz, C. Lutz, F. Wolter, and M. Zakharyaschev. $\mathcal{E}$-connections of abstract description systems. Submitted, 2003. Available at `http://www.csc.liv.ac.uk/~frank/publ/publ.html`.

[8] O. Kutz, F. Wolter, and M. Zakharyaschev. A note on concepts and distances. In *Proceedings of DL2001*, number 49 in CEUR-WS (`http://ceur-ws.org/`), pages 113–121, 2001.

[9] O. Kutz, F. Wolter, and M. Zakharyaschev. Connecting abstract description systems. In *Proceedings of KR 2002*, pages 215–226. Morgan Kaufmann, 2002.

[10] C. Lutz, H. Sturm, F. Wolter, and M. Zakharyaschev. Tableaux for temporal description logic with constant domains. In *Proceedings of IJCAR'01*, number 2083 in LNAI, pages 121–136. Springer-Verlag, 2001.

[11] M. Marx. Complexity of products of modal logics. *Journal of Logic and Computation*, 9:197–214, 1999.

[12] A. Schaerf. Reasoning with individuals in concept languages. *Data and Knowledge Engineering*, 13(2):141–176, 1994.