

# Ontology Evolution Under Semantic Constraints

B. Cuenca Grau<sup>1</sup>, E. Jimenez-Ruiz<sup>1</sup>

<sup>1</sup>Department of Computer Science  
University of Oxford  
bernardo.cuenca.grau@cs.ox.ac.uk  
ernesto@cs.ox.ac.uk

E. Kharlamov<sup>1,2</sup>, D. Zheleznyakov<sup>2</sup>

<sup>2</sup>KRDB Research Centre  
Free University of Bozen-Bolzano  
kharlamov@inf.unibz.it  
zheleznyakov@inf.unibz.it

## Abstract

The dynamic nature of ontology development has motivated the formal study of ontology evolution problems. This paper presents a logical framework that enables fine-grained investigation of evolution problems at a deductive level. In our framework, the *optimal evolutions* of an ontology  $\mathcal{O}$  are those ontologies  $\mathcal{O}'$  that maximally preserve both the structure of  $\mathcal{O}$ , and its entailments in a given preservation language. We show that our framework is compatible with the postulates of Belief Revision, and we investigate the existence of optimal evolutions in various settings. In particular, we present first results on TBox-level revision and contraction in the  $\mathcal{EL}$  and  $\mathcal{FL}_0$  families of Description Logics.

## Introduction

Ontologies written in the Web Ontology Language (OWL) (Horrocks, Patel-Schneider, and van Harmelen 2003), and its revision OWL 2 (Cuenca Grau et al. 2008b) are becoming increasingly important for a wide range of applications. The formal underpinning of OWL is based on Description Logics (DLs) – knowledge representation formalisms with well-understood computational properties (Baader et al. 2003). A DL ontology  $\mathcal{O}$  typically consists of a TBox  $\mathcal{T}$ , which describes general (i.e., schema-level) domain knowledge, and an ABox  $\mathcal{A}$ , which provides data about specific individuals.

OWL ontologies are being extensively used in the clinical sciences, where large-scale ontologies have been developed (e.g., the NCI Thesaurus (NCI), the Foundational Model of Anatomy (FMA), and SNOMED). These ontologies are not static entities, but rather they are frequently modified when new information needs to be incorporated, or existing information is no longer considered valid (e.g., the developers of NCI perform over 900 monthly changes (Hartung, Kirsten, and Rahm 2008)). The impact of such changes on the semantics of the ontology, however, is difficult to predict and understand.

This dynamic nature of ontologies motivates the study of *ontology evolution* from both foundational and practical perspectives (Fridman Noy et al. 2004; Haase and Stojanovic 2005; Flouris et al. 2008; Qi and Du 2009;

Calvanese et al. 2010; Wang, Wang, and Topor 2010b; Jiménez-Ruiz et al. 2011; Konev, Walther, and Wolter 2008; Gonçalves, Parsia, and Sattler 2011).

In AI and Belief Revision, the process of “incorporating” new information into a knowledge base (KB) is called *revision*, whereas the process of “retracting” information that is no longer considered to hold is called *contraction* (Alchourrón, Gärdenfors, and Makinson 1985; Peppas 2007). The properties that revision and contraction operators need to satisfy are dictated by the *principle of minimal change* (Alchourrón, Gärdenfors, and Makinson 1985), according to which the semantics of a KB should change “as little as possible”, thus ensuring that modifications have the least possible impact. A distinction is often made between revision and *update*, where the purpose of the latter is to bring the KB up to date when the world changes (Katsuno and Mendelzon 1991; Kharlamov and Zheleznyakov 2011; Liu et al. 2011). In this paper, however, we use the term *evolution* to encompass revision and contraction, and we do not consider here the problem of update.

Logic-based semantics derived from the principle of minimal change have been recently studied in the context of ontology evolution. These semantics are either *model-based* (MBS) or *formula-based* (FBS). Under both semantics, evolution of an  $\mathcal{LO}$ -ontology  $\mathcal{O}$  results in an  $\mathcal{LO}'$ -ontology  $\mathcal{O}'$  that incorporates (or retracts) the required information, and the difference lies in the way  $\mathcal{O}'$  is obtained. Under MBS the set of all models  $\mathcal{M}$  of  $\mathcal{O}$  is evolved into a new set  $\mathcal{M}'$  of models that are “as close as possible” to those in  $\mathcal{M}$  (w.r.t. some notion of distance between models); then,  $\mathcal{O}'$  is the ontology that axiomatises  $\mathcal{M}'$  (Qi and Du 2009; Giacomo et al. 2009; Calvanese et al. 2010; Kharlamov and Zheleznyakov 2011; Wang, Wang, and Topor 2010b; 2010a). MBSs, however, suffer from intrinsic inexpressibility problems, even for lightweight DLs such as *DL-Lite* (Calvanese et al. 2007), where axiomatisation of  $\mathcal{M}'$  requires a DL with disjunction and nominals (Kharlamov and Zheleznyakov 2011).

Under FBS,  $\mathcal{O}'$  is defined as a maximal subset of the deductive closure of  $\mathcal{O}$  (under  $\mathcal{LO}$ -consequences) that satisfies the evolution requirements. FBSs for DLs have been less studied. In particular, existing results

(Calvanese et al. 2010; Lenzerini and Savo 2011) are restricted to *DL-Lite*, where the deductive closure of  $\mathcal{O}$  is finite. It is unknown, however, how to compute  $\mathcal{O}'$  if the closure of  $\mathcal{O}$  is infinite, as is the case when  $\mathcal{O}$  is an  $\mathcal{EL}$  ontology (Baader, Brandt, and Lutz 2005).

Approaches to ontology evolution typically adopted in practice (especially when changes occur at the TBox level) are essentially *syntactic* (Haase and Stojanovic 2005; Kalyanpur et al. 2006; Jiménez-Ruiz et al. 2011). Many such approaches are based on the notion of a *justification*: a minimal subset of the ontology that entails a given consequence (Kalyanpur et al. 2005; Schlobach et al. 2007; Kalyanpur et al. 2007; Peñaloza and Sertkaya 2010). For example, to retract an axiom  $\alpha$  entailed by  $\mathcal{O}$ , it suffices to compute all justifications for  $\alpha$  in  $\mathcal{O}$ , find a minimal subset  $\mathcal{R}$  of  $\mathcal{O}$  with at least one axiom from each justification, and take  $\mathcal{O}' = \mathcal{O} \setminus \mathcal{R}$  as the result of the evolution. This solution complies with a “syntactical” notion of minimal change: retracting  $\alpha$  results in the deletion of a minimal set of axioms and hence the structure of  $\mathcal{O}$  is maximally preserved. Furthermore,  $\mathcal{O}'$  is guaranteed to exist for expressive DLs, and practical algorithms have been implemented in ontology development platforms (Kalyanpur et al. 2007; Suntisrivaraporn et al. 2008). By removing  $\mathcal{R}$  from  $\mathcal{O}$ , however, we may be inadvertently retracting consequences of  $\mathcal{O}$  other than  $\alpha$ , which are “intended”. Identifying and recovering such intended consequences is an important issue.

This paper presents a framework that bridges the gap between logic-based and syntactic approaches to ontology evolution. On the logic side we focus on FBSs, which deal with formulae rather than models, and hence are closer in spirit to syntactic approaches than MBSs. Roughly speaking, evolution of an  $\mathcal{LO}$ -ontology  $\mathcal{O}$  is “triggered” by *semantic constraints*  $\mathcal{C} = (\mathcal{C}^+, \mathcal{C}^-)$  – sets of formulae in a constraint language  $\mathcal{LC}$  specifying both the consequences  $\mathcal{C}^+$  that must hold in an evolved ontology  $\mathcal{O}'$ , and the consequences  $\mathcal{C}^-$  that cannot hold. The principle of minimal change is reflected in our framework along two dimensions. The first one is *structural*: ontologies are the result of a time-consuming modeling process and thus  $\mathcal{O}'$  should not change the structure of  $\mathcal{O}$  in a substantial way. The second dimension is *deductive*: in addition to satisfying the constraints  $\mathcal{C}$ ,  $\mathcal{O}'$  should, on the one hand, avoid introducing spurious consequences that do not follow from  $\mathcal{O} \cup \mathcal{C}^+$  and, on the other hand, it should maximally preserve the consequences of  $\mathcal{O}$  in a given *preservation language*  $\mathcal{LP}$ . Ontologies that comply with the principle of minimal change along these two dimensions are called *optimal evolutions*. To show that our semantics does not lead to unexpected results, we discuss instantiations of our framework and establish a connection with the revision and contraction postulates of Belief Revision (Alchourrón, Gärdenfors, and Makinson 1985; Peppas 2007).

An important issue in our framework is *expressibility* (Can an optimal evolution  $\mathcal{O}'$  of  $\mathcal{O}$  be expressed in a

given  $\mathcal{LO}'$  for given  $\mathcal{C}$  and  $\mathcal{LP}$ ?). If such  $\mathcal{O}'$  exists, we aim at establishing bounds to its size (Cadoli et al. 1999). We show that if  $\mathcal{LP}$  is finite (i.e., it has only finitely many non-equivalent sentences for any finite signature) an optimal  $\mathcal{O}'$  exists provided that the constraints themselves can be satisfied; this case already extends all syntactic evolution approaches known to us and also captures the scenario where  $\mathcal{LP}$  is *DL-Lite*. We then address the challenging problem of expressibility when  $\mathcal{LP}$  is an “infinite” language and focus on the retraction of axioms in TBoxes expressed in the lightweight DLs  $\mathcal{EL}$  and its “dual” logic  $\mathcal{FL}_0$  (Baader et al. 2003). We show inexpressibility of optimal contractions in both cases when  $\mathcal{LP}$  coincides with the ontology language, even when retracting a single subsumption between atomic concepts. Our negative results provide insights in the causes of inexpressibility, so we then focus on  $\mathcal{EL}$  (which is especially relevant for bio-medical ontology modeling), investigate sufficient conditions for expressibility, and study the size of the resulting optimal evolutions. To the best of our knowledge, ours are the first results on TBox-level revision and contraction beyond *DL-Lite*. Finally, we report on experiments in which we study contraction in SNOMED. This paper is accompanied by an online technical report containing all proofs [www.inf.unibz.it/~zheleznyakov/krfull.pdf](http://www.inf.unibz.it/~zheleznyakov/krfull.pdf).

## Preliminaries

Our evolution framework is applicable to first-order logic (FOL) rather than Description Logics. Our work, however, is motivated by DL ontologies, so we will use DL terminology throughout the paper. We assume standard definitions of (function-free) FOL signature, predicates, sentences, interpretations, satisfiability and entailment.

An ontology  $\mathcal{O} = \mathcal{T} \cup \mathcal{A}$  consists of a finite set of sentences  $\mathcal{T}$  (the TBox) and a finite set of ground atoms  $\mathcal{A}$  (the ABox). A DL is a recursive set of ontologies closed under renaming of constants and the subset relation. Predicates in DL signatures are either unary (called *atomic concepts*) or binary (*atomic roles*). DLs use a specialised syntax, where variables are omitted, and which provides operators for constructing complex concepts (formulae with one free variable) and roles (formulae with two free variables) from simpler ones, as well as a set of axioms. For  $\xi$  a concept, role, or (set of) axiom(s),  $\text{sig}(\xi)$  denotes the set of atomic concepts, roles and constants in  $\xi$ .

An *interpretation*  $\mathcal{I}$  for a DL signature  $\Sigma$  is a pair  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ , where  $\Delta^{\mathcal{I}}$  is a non-empty *domain set* and the *interpretation function*  $\cdot^{\mathcal{I}}$  maps each constant  $a$  to an element  $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ , each atomic concept  $A$  to a set  $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ , and each atomic role  $R$  to a relation  $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ . Many DLs provide the  $\top$  and  $\perp$  concepts, which are interpreted as  $\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$  and  $\perp^{\mathcal{I}} = \emptyset$ , respectively.

Let  $\mathcal{L}$  and  $\mathcal{L}'$  be DLs s.t.  $\mathcal{L}' \subseteq \mathcal{L}$  and  $\mathcal{O}$  be an  $\mathcal{L}$ -ontology. The *closure* of  $\mathcal{O}$  w.r.t.  $\mathcal{L}'$ , written  $\text{Cl}_{\mathcal{L}'}(\mathcal{O})$  (or  $\text{Cl}(\mathcal{O})$  when  $\mathcal{L}'$  is clear from the context), is the set of all  $\mathcal{L}'$ -axioms  $\alpha$  entailed by  $\mathcal{O}$ . Let  $\alpha \in \text{Cl}_{\mathcal{L}'}(\mathcal{O})$ , then  $[\alpha] = \{\beta \in \text{Cl}_{\mathcal{L}'}(\mathcal{O}) \mid \beta \equiv \alpha\}$ . Clearly,  $[\alpha]$  is an

equivalence class in the quotient set of  $\text{Cl}_{\mathcal{L}'}(\mathcal{O})$  modulo logical equivalence.

## The Description Logics $\mathcal{EL}$ and $\mathcal{FL}_0$

We next describe the specific DLs mentioned in this paper, namely  $\mathcal{EL}$  (Baader, Brandt, and Lutz 2005) and its “dual” logic  $\mathcal{FL}_0$  (Baader et al. 2003). Since we only mention these logics in the context of TBox-level evolution, we omit the definition of ABoxes.

Let  $\mathcal{L}$  be either  $\mathcal{EL}$  or  $\mathcal{FL}_0$  and let  $\Sigma$  be a DL signature, which we consider implicit in all definitions. The set of  $\mathcal{L}$ -concepts is the smallest set containing  $\top$ ,  $A$ ,  $C_1 \sqcap C_2$ , and  $\text{QR}.C$  with  $\text{Q} = \exists$  if  $\mathcal{L} = \mathcal{EL}$  and  $\text{Q} = \forall$  if  $\mathcal{L} = \mathcal{FL}_0$ , for  $A$  an atomic concept,  $C$ ,  $C_1$  and  $C_2$   $\mathcal{L}$ -concepts, and  $R$  an atomic role. For  $w = R_1 \dots R_n$  a word of atomic roles,  $\text{Q} \in \{\exists, \forall\}$  and  $C$  a concept, we denote with  $\text{Q}w.C$  the concept  $\text{QR}_1 \dots \text{QR}_n.C$ .

The *quantifier depth* of an  $\mathcal{L}$ -concept is inductively defined as follows, for  $A$  atomic, and  $C$ ,  $C_1$ ,  $C_2$   $\mathcal{L}$ -concepts: (i)  $\text{depth}(A) = 0$ ; (ii)  $\text{depth}(\text{QR}.C) = 1 + \text{depth}(C)$ ; and (iii)  $\text{depth}(C_1 \sqcap C_2) = \max(\text{depth}(C_1), \text{depth}(C_2))$ .

An  $\mathcal{L}$ -axiom is of the form  $C_1 \sqsubseteq C_2$ , where  $C_1$  and  $C_2$  are  $\mathcal{L}$ -concepts ( $C_1 \equiv C_2$  is a shorthand for  $C_1 \sqsubseteq C_2$  and  $C_2 \sqsubseteq C_1$ ). An  $\mathcal{L}$ -TBox is a finite set of  $\mathcal{L}$ -axioms. The semantics is standard. We denote with  $\mathcal{L}^{\min}$  the DL that only allows for axioms  $A \sqsubseteq B$  with  $A, B$  atomic.<sup>1</sup>

An  $\mathcal{EL}$ -TBox  $\mathcal{T}$  is normalised if it has only axioms of the following forms, where  $A, B, A_1, A_2$  are atomic concepts or  $\top$ :  $A \sqsubseteq B$ ,  $A_1 \sqcap A_2 \sqsubseteq B$ ,  $\exists R.A \sqsubseteq B$ , or  $A \sqsubseteq \exists R.B$ . Each  $\mathcal{EL}$ -TBox  $\mathcal{T}$  can be normalised into a TBox  $\mathcal{T}'$  that is a conservative extension of  $\mathcal{T}$  (Baader, Brandt, and Lutz 2005). The *canonical model*  $\mathcal{I}_{\mathcal{T}}$  of a normalised  $\mathcal{EL}$ -TBox  $\mathcal{T}$  is defined as follows:

- $\Delta^{\mathcal{I}_{\mathcal{T}}} = \{v_A \mid A \in \text{sig}(\mathcal{T}) \cup \{\top\} \text{ is a concept}\}$ ;
- $A^{\mathcal{I}_{\mathcal{T}}} = \{v_B \mid \mathcal{T} \models B \sqsubseteq A\}$ ; and
- $R^{\mathcal{I}_{\mathcal{T}}} = \{(v_A, v_B) \mid \mathcal{T} \models A \sqsubseteq \exists R.B\}$ .

Deductively,  $\mathcal{I}_{\mathcal{T}}$  is described by the *basic closure*  $\text{BCl}(\mathcal{T})$  of  $\mathcal{T}$ , i.e., the subset of  $\text{Cl}(\mathcal{T})$  with all axioms of the form  $A \sqsubseteq B$ , or  $A \sqsubseteq \exists R.B$ , with  $A, B, R \in \text{sig}(\mathcal{T}) \cup \{\top\}$ . The set  $\text{BCl}(\mathcal{T})$  is of size polynomial in the size of  $\mathcal{T}$ .

## Ontology Evolution Under Constraints

We next introduce our ontology evolution framework. Since our framework is not restricted to any particular Description Logic, we adopt the general definition of a DL in the preliminaries.

### Semantic Constraints

Ontologies are dynamic entities, which are subject to frequent modifications. Consider, for example, the development of an ontology  $\mathcal{O}_{\text{ex}}$  about disorders of the skeletal system, which entails the following axioms:

$$\begin{aligned} \beta_1 &= \text{Arthropathy} && \sqsubseteq \text{JointDisorder}, \\ \beta_2 &= \text{ArthropathyTest} && \sqsubseteq \text{JointFinding}, \\ \gamma_1 &= \text{Arthropathy} && \sqsubseteq \text{JointFinding}. \end{aligned}$$

<sup>1</sup> $\mathcal{L}^{\min}$  is arguably the smallest DL and  $\mathcal{L}^{\min} \subseteq \mathcal{EL} \cap \mathcal{FL}_0$ .

After close inspection, the developers of  $\mathcal{O}_{\text{ex}}$  notice that  $\gamma_1$  is due to a modeling error and should be retracted, whereas the other entailments are intended and hence the retraction of  $\gamma_1$  should not invalidate them. These requirements are formalised in our framework using a *constraint*  $\mathcal{C} = (\mathcal{C}^+, \mathcal{C}^-)$ , where  $\mathcal{C}^+$  specifies the entailments that must hold in the evolved ontology and  $\mathcal{C}^-$  specifies those that cannot hold. In our example, we have  $\mathcal{C}_{\text{ex}}^+ = \{\beta_1, \beta_2\}$ ,  $\mathcal{C}_{\text{ex}}^- = \{\gamma_1\}$ , and  $\mathcal{C}_{\text{ex}} = (\mathcal{C}_{\text{ex}}^+, \mathcal{C}_{\text{ex}}^-)$ .

**Definition 1.** Let  $\mathcal{LC}$  and  $\mathcal{LO}'$  be DLs. An  $\mathcal{LC}$ -constraint is a pair  $\mathcal{C} = (\mathcal{C}^+, \mathcal{C}^-)$ , with  $\mathcal{C}^+, \mathcal{C}^- \in \mathcal{LC}$ . We say that  $\mathcal{O}' \in \mathcal{LO}'$  conforms to

- $\mathcal{C}^+$  (written  $\mathcal{O}' \times \mathcal{C}^+$ ) if  $\mathcal{O}' \models \mathcal{C}^+$ ;
- $\mathcal{C}^-$  (written  $\mathcal{O}' \times \mathcal{C}^-$ ) if  $\mathcal{O}' \not\models \alpha$  for all  $\alpha \in \mathcal{C}^-$ ; and
- $\mathcal{C}$  (written  $\mathcal{O}' \times \mathcal{C}$ ) if  $\mathcal{O}' \times \mathcal{C}^+$  and  $\mathcal{O}' \times \mathcal{C}^-$ .

We say that  $\mathcal{C}$  is  $\mathcal{LO}'$ -conformant if there exists an ontology  $\mathcal{O}' \in \mathcal{LO}'$  such that  $\mathcal{O}' \times \mathcal{C}$ .

In general,  $\mathcal{C}^+$  may contain new information to be incorporated in  $\mathcal{O}$ , or information already entailed by  $\mathcal{O}$  to be preserved in the evolution; conversely,  $\mathcal{C}^-$  may contain information in  $\mathcal{O}$  to be retracted, or information not in  $\mathcal{O}$  that must not be introduced by the evolution. A constraint  $\mathcal{C}$  such that  $\mathcal{C}^-$  contains a tautology  $\alpha$ , (i.e.,  $\emptyset \models \alpha$ ) cannot be  $\mathcal{LO}'$ -conformant, regardless of  $\mathcal{LO}'$ .

Let us assume that the following axioms about skeletal disorders are also contained in  $\mathcal{O}_{\text{ex}}$ :

$$\begin{aligned} \delta_1 &= \text{JointDisorder} && \sqsubseteq \text{SkeletDisorder} \sqcap \text{JointFinding}, \\ \delta_2 &= \text{JointDisorder} && \sqsubseteq \exists \text{located.Joint}, \\ \delta_3 &= \exists \text{located.Joint} && \sqsubseteq \exists \text{located.Skeleton}, \\ \delta_4 &= \text{SkeletDisorder} && \equiv \text{Disorder} \sqcap \exists \text{located.Skeleton}, \\ \delta_5 &= \text{Cortisone} && \sqsubseteq \text{Steroid} \sqcap \exists \text{treats.JointDisorder}. \end{aligned}$$

Clearly, an ontology containing  $\delta_1$  cannot conform to  $\mathcal{C}_{\text{ex}}$ ; in contrast, axioms  $\delta_2$ - $\delta_5$  do not preclude conformance. The following reasoning tasks are thus of interest.

(T1) Check if  $\mathcal{O}'$  conforms to  $\mathcal{C}$ ;

(T2) Check if  $\mathcal{C}$  is  $\mathcal{LO}'$ -conformant.

Task T1 amounts to checking entailment. Furthermore, as shown next, task T2 also reduces to entailment checking provided that  $\mathcal{LO}'$  can express the constraints.

**Proposition 2.** Let  $\mathcal{C} = (\mathcal{C}^+, \mathcal{C}^-)$  be an  $\mathcal{LC}$ -constraint with  $\mathcal{LC} \subseteq \mathcal{LO}'$ . Then,  $\mathcal{C}$  is  $\mathcal{LO}'$ -conformant iff either  $\mathcal{C}^+$  is satisfiable and  $\mathcal{C}^+ \times \mathcal{C}^-$ ; or  $\mathcal{C}^- = \emptyset$ .

Essentially, a constraint is conformant if its two components  $\mathcal{C}^+$  and  $\mathcal{C}^-$  do not contradict each other. The constraints in our running example are clearly  $\mathcal{EL}$ -conformant since  $\beta_1, \beta_2 \in \mathcal{EL}$  where every ontology is satisfiable, and  $\{\beta_1, \beta_2\} \not\models \gamma_1$ .

## The Notion of an Evolution

The notion of conformance, however, does not yet establish a connection between the original ontology  $\mathcal{O}$  and the evolved ontology  $\mathcal{O}'$ . The required connection is established by the notion of an *evolution*.

**Definition 3.** Let  $\mathcal{LO}$  and  $\mathcal{LO}'$  be DLs. An ontology  $\mathcal{O}'$  is an  $\mathcal{LO}'$ -evolution of a satisfiable  $\mathcal{LO}$ -ontology  $\mathcal{O}$  under constraint  $\mathcal{C} = (\mathcal{C}^+, \mathcal{C}^-)$  if the following holds:

1.  $\mathcal{O}' \in \mathcal{LO}'$ ;
2.  $\mathcal{O}' \propto \mathcal{C}$ ; and
3. if  $\mathcal{C}^+$  is satisfiable, then an ontology  $\mathcal{O}_1 \in \mathcal{LO}$  exists s.t.  $\mathcal{O} \models \mathcal{O}_1$ ,  $\mathcal{O}_1 \cup \mathcal{C}^+$  is satisfiable, and  $\mathcal{O}_1 \cup \mathcal{C}^+ \models \mathcal{O}'$ .

With  $\text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$  we denote the class of all  $\mathcal{LO}'$ -evolutions of  $\mathcal{O}$  under  $\mathcal{C}$ .

The last condition in Definition 3 essentially ensures that  $\mathcal{O}'$  only entails “genuine” information that follows from  $\mathcal{O} \cup \mathcal{C}^+$ , thus preventing the introduction of logical consequences unrelated to  $\mathcal{O}$  and  $\mathcal{C}^+$ ; furthermore, the main role of ontology  $\mathcal{O}_1$  in the definition is to preserve satisfiability in the evolution whenever possible: if  $\mathcal{C}^+$  is satisfiable, then every evolution  $\mathcal{O}'$  is guaranteed to be satisfiable as well. As discussed later on, Condition 3 in Definition 3 makes our notion of evolution compatible with the postulates of Belief Revision, according to which unsatisfiable revisions are only acceptable when the new information itself is unsatisfiable.

Definition 3 motivates the following reasoning tasks.

- (T3) Check if  $\mathcal{O}'$  is an  $\mathcal{LO}'$ -evolution of  $\mathcal{O}$  under  $\mathcal{C}$ ;  
(T4) Check if some  $\mathcal{LO}'$ -evolution of  $\mathcal{O}$  under  $\mathcal{C}$  exists.

Task T3 amounts to entailment checking provided that  $\mathcal{O} \cup \mathcal{C}^+$  is satisfiable. The following proposition shows that tasks T4 and T2 are inter-reducible, thus establishing a strong connection between constraint conformance and existence of an evolution.

**Proposition 4.** Let  $\mathcal{C}$  be an  $\mathcal{LC}$ -constraint where  $\mathcal{LC} \subseteq \mathcal{LO}'$ . Then,  $\text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$  is non-empty iff  $\mathcal{C}$  is  $\mathcal{LO}'$ -conformant.

In our example,  $\mathcal{O}'_{\text{ex}} = \mathcal{C}_{\text{ex}}^+ = \{\beta_1, \beta_2\}$  is an evolution of  $\mathcal{O}_{\text{ex}}$ . Although  $\mathcal{O}'_{\text{ex}}$  conforms to  $\mathcal{C}_{\text{ex}}$  and does not introduce spurious entailments (in fact, it is the simplest ontology with these properties), it is arguably not compliant with the *principle of minimal change*, as it loses all information in  $\mathcal{O}_{\text{ex}}$  that is not in  $\mathcal{C}_{\text{ex}}^+$  (e.g., everything that follows from axioms  $\delta_2$ - $\delta_5$ ).

## Optimal Evolutions

The principle of minimal change is reflected in our framework both *structurally* and *deductively*. On the one hand,  $\mathcal{O}'$  should minimize alterations in the structure of  $\mathcal{O}$ ; on the other hand,  $\mathcal{O}'$  should maximally preserve the entailments of  $\mathcal{O}$  in a given *preservation language*  $\mathcal{LP}$ . Formally, our framework defines a preorder  $\succeq_{\mathcal{LP}}$  over the class  $\text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$ , which establishes a “preference” relation between evolutions based on the aforementioned structural and deductive criteria. Our definition of  $\succeq_{\mathcal{LP}}$  uses the notion of entailment w.r.t. a DL introduced in (Konev, Walther, and Wolter 2008).

**Definition 5.** Let  $\mathcal{LP}$  be a DL. We say that an ontology  $\mathcal{O}_1$   $\mathcal{LP}$ -entails an ontology  $\mathcal{O}_2$  if  $\mathcal{O}_2 \models \alpha$  implies  $\mathcal{O}_1 \models \alpha$  for each  $\alpha \in \mathcal{LP}$ . The binary relation  $\succeq_{\mathcal{LP}}$  over  $\text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$  is defined as follows:  $\mathcal{O}'_1 \succeq_{\mathcal{LP}} \mathcal{O}'_2$  iff

1.  $\mathcal{O}'_1$   $\mathcal{LP}$ -entails  $\mathcal{O}'_2$ , and
2.  $\mathcal{O}'_2 \cap \mathcal{O} \subseteq \mathcal{O}'_1 \cap \mathcal{O}$ .

It is well-known that a preorder induces an equivalence relation as given next.

**Definition 6.** The equivalence relation  $\equiv_{\mathcal{LP}}$  induced by  $\succeq_{\mathcal{LP}}$  is defined as follows:  $\mathcal{O}'_1 \equiv_{\mathcal{LP}} \mathcal{O}'_2$  iff  $\mathcal{O}'_1 \succeq_{\mathcal{LP}} \mathcal{O}'_2$  and  $\mathcal{O}'_2 \succeq_{\mathcal{LP}} \mathcal{O}'_1$ . Given  $\mathcal{O}' \in \text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$ , we denote with  $[\mathcal{O}']$  the equivalence class to which  $\mathcal{O}'$  belongs.

Ontologies in the same equivalence class are indistinguishable from the point of view of our framework: they coincide in their entailments over the preservation language, and they contain the same axioms from  $\mathcal{O}$ .

We can now establish a partial order  $\succeq_{\mathcal{LP}}$  over the quotient set, and define *optimal* evolutions as the ontologies belonging to an  $\succeq_{\mathcal{LP}}$ -maximal equivalence class.

**Definition 7.** The relations  $\succeq_{\mathcal{LP}}$  and  $\succ_{\mathcal{LP}}$  over the quotient set  $\text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C}) \setminus \equiv_{\mathcal{LP}}$  are as follows:

- $[\mathcal{O}'] \succeq_{\mathcal{LP}} [\mathcal{O}']$  iff  $\mathcal{O}' \succeq_{\mathcal{LP}} \mathcal{O}''$ ; and
- $[\mathcal{O}'] \succ_{\mathcal{LP}} [\mathcal{O}']$  iff  $[\mathcal{O}'] \succeq_{\mathcal{LP}} [\mathcal{O}']$  and  $[\mathcal{O}'] \neq [\mathcal{O}']$ .

We say that  $\mathcal{O}' \in \text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$  is  $\mathcal{LP}$ -optimal if  $[\mathcal{O}']_{\mathcal{LP}}$  is a  $\succeq_{\mathcal{LP}}$ -maximal element in  $\text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C}) \setminus \equiv_{\mathcal{LP}}$ .

In our example, the evolution  $\mathcal{O}'_{\text{ex}} = \{\beta_1, \beta_2\}$  is not  $\mathcal{EL}$ -optimal: first, it does not include “harmless” axioms from  $\mathcal{O}_{\text{ex}}$  such as  $\delta_2$ - $\delta_5$ ; second,  $\mathcal{O}'_{\text{ex}}$  does not entail consequences of  $\mathcal{O}_{\text{ex}}$  such as  $\text{JointDisorder} \sqsubseteq \text{SkeletDisorder}$  or  $\text{Cortisone} \sqsubseteq \text{Steroid} \sqcap \exists \text{treats.SkeletDisorder}$ , which do not cause  $\gamma_1$  to be entailed.

Definition 7 thus motivates the following tasks.

- (T5) Check if  $\mathcal{O}'$  is  $\mathcal{LP}$ -optimal in  $\text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$ ;  
(T6) Check if an  $\mathcal{LP}$ -optimal  $\mathcal{O}' \in \text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$  exists.

Note that there can be multiple  $\mathcal{LP}$ -optimal evolutions of  $\mathcal{O}$ . In particular, each ontology in an  $\succeq_{\mathcal{LP}}$ -maximal class is  $\mathcal{LP}$ -optimal; furthermore, ontologies in the same  $\succeq_{\mathcal{LP}}$ -maximal class are *indistinguishable*, whereas ontologies in different  $\succeq_{\mathcal{LP}}$ -maximal classes are *incomparable*. Hence, the following additional task is also of interest.

- (T7) Compute [some/all]  $\mathcal{LP}$ -optimal evolutions.

In an application, it may be desirable to single out a “preferred” optimal evolution. This could be achieved, by imposing a preference relation over axioms in  $\mathcal{O}$  (e.g., using trust values), or by taking into account users’ feedback. Such mechanisms, however, are application dependent and hence external to our framework.

## Framework Instantiations & Belief Revision

We next argue that our evolution semantics is theoretically well founded, and does not lead to unexpected results. To this end, we discuss several instantiations of our framework, and establish a connection with the Belief Revision postulates.

R1	$\mathcal{O} * (\mathcal{C}^+, \emptyset) \in \mathcal{LO}$	C1	$\mathcal{O} \div (\emptyset, \mathcal{C}^-) \in \mathcal{LO}$
R2	$\mathcal{O} * (\mathcal{C}^+, \emptyset) \models \mathcal{C}^+$	C2	$\mathcal{O} \models \mathcal{O} \div (\emptyset, \mathcal{C}^-)$
R3	$\mathcal{O} \cup \mathcal{C}^+ \models \mathcal{O} * (\mathcal{C}^+, \emptyset)$	C3	If $\mathcal{O} \propto \mathcal{C}^-$ , then $\mathcal{O} \div (\emptyset, \mathcal{C}^-) \equiv \mathcal{O}$
R4	If $\mathcal{O} \cup \mathcal{C}^+$ is satisfiable, then $\mathcal{O} * (\mathcal{C}^+, \emptyset) \models \mathcal{O} \cup \mathcal{C}^+$	C4	If $\emptyset \not\models \mathcal{C}^-$ , then $\mathcal{O} \div (\emptyset, \mathcal{C}^-) \propto \mathcal{C}^-$
R5	If $\mathcal{C}^+$ is satisfiable, then $\mathcal{O} * (\mathcal{C}^+, \emptyset)$ satisfiable	C5	If $\mathcal{O} \models \mathcal{C}^-$ , then $(\mathcal{O} \div (\emptyset, \mathcal{C}^-)) \cup \mathcal{C}^- \models \mathcal{O}$
R6	If $\mathcal{C}_1^+ \equiv \mathcal{C}_2^+$ , then $\mathcal{O} * (\mathcal{C}_1^+, \emptyset) \equiv \mathcal{O} * (\mathcal{C}_2^+, \emptyset)$	C6	If $\mathcal{C}_1^- \equiv \mathcal{C}_2^-$ , then $\mathcal{O} \div (\emptyset, \mathcal{C}_1^-) \equiv \mathcal{O} \div (\emptyset, \mathcal{C}_2^-)$

Figure 1: Basic revision and contraction postulates in Belief Revision

**No evolution.** Intuitively, constraints  $\mathcal{C}$  act as “triggers” to the evolution of  $\mathcal{O}$ . In particular, if  $\mathcal{O}$  already conforms to  $\mathcal{C}$  and  $\mathcal{LO}' = \mathcal{LO}$ , then  $\mathcal{O}$  should not evolve. The following proposition ensures that our framework behaves as expected in such situation, and (the equivalence class of)  $\mathcal{O}$  is singled out as optimal.

**Proposition 8.** *Let  $\mathcal{LO}' = \mathcal{LO} = \mathcal{LC}$ , let  $\mathcal{O} \in \mathcal{LO}$ , let  $\mathcal{C}$  be an  $\mathcal{LC}$ -constraint, and  $\mathcal{O} \propto \mathcal{C}$ . Then, for each  $\mathcal{O}' \in \text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$  and each DL  $\mathcal{LP}$ ,  $\mathcal{O}'$  is  $\mathcal{LP}$ -optimal iff  $\mathcal{O}' \in [\mathcal{O}]$ .*

**Revision.** *Revision* is the process of accommodating new information while preserving satisfiability whenever possible (Alchourrón, Gärdenfors, and Makinson 1985; Peppas 2007). The new information is perceived as reliable, and it prevails over all conflicting knowledge in the ontology. More precisely, the process of revision is formalised by means of a *revision function* “\*”, which maps each theory  $\mathcal{K}$  in a language  $\mathcal{L}$  and each  $\mathcal{L}$ -formula  $\varphi$  to a new  $\mathcal{L}$ -theory  $\mathcal{K} * \varphi$ ; in this setting, the principle of minimal change is formalised as a set of *postulates* that each revision function ought to satisfy (Peppas 2007; Alchourrón, Gärdenfors, and Makinson 1985).

Revision can be captured in our framework by using  $\mathcal{C}^+$  to represent the new information and  $\mathcal{C}^-$  to be the empty set. Furthermore, since in Belief Revision no distinction is usually made between the languages of  $\mathcal{K}$ ,  $\varphi$  and  $\mathcal{K} * \varphi$ , we will also assume that that languages  $\mathcal{LO}$ ,  $\mathcal{LO}'$  and  $\mathcal{LC}$  coincide. In this setting, revision functions can be defined as given next.

**Definition 9.** *Let  $\mathcal{LO} = \mathcal{LO}' = \mathcal{LC}$  and  $\mathcal{LP}$  be DLs. A revision function  $*$  is a binary function that maps each  $\mathcal{LO}$ -ontology  $\mathcal{O}$  and each  $\mathcal{LC}$ -constraint  $\mathcal{C}$  of the form  $\mathcal{C} = (\mathcal{C}^+, \emptyset)$  to an ontology  $\mathcal{O} * \mathcal{C}$  such that*

- (i)  $\mathcal{O} * \mathcal{C}$  is  $\mathcal{LP}$ -optimal in  $\text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$ ; and
- (ii) If  $\mathcal{C}_1^+ \equiv \mathcal{C}_2^+$ , then  $\mathcal{O} * (\mathcal{C}_1^+, \emptyset) \equiv \mathcal{O} * (\mathcal{C}_2^+, \emptyset)$ .

Our next step is to show that revision functions as in Definition 9 are consistent with the basic postulates of Belief Revision. These postulates can be formulated in the context of our framework as given in Figure 1. Postulate R1 says that the result of the revision is also an ontology in the relevant language; postulate R2 says that the new information always holds after revision; postulate R3 and R4 together state that, whenever the new information does not contradict  $\mathcal{O}$ , there is no reason to remove any information from  $\mathcal{O}$ ; postulate R5 says that satisfiability should be preserved whenever possible (unsatisfiability is only acceptable if the new

information itself is unsatisfiable); finally, postulate R6 states that the syntax of the new information is irrelevant to the revision process.

**Theorem 10.** *Let  $\mathcal{LO} = \mathcal{LO}' = \mathcal{LC}$  and  $\mathcal{LP}$  be DLs, and let  $*$  be a revision function as in Definition 9. Then, function  $*$  satisfies postulates R1 to R6 in Figure 1.*

**Contraction.** *Contraction* is the process of retracting information that is no longer considered to hold. Like revision, contraction is defined using a function “ $\div$ ” mapping each theory  $\mathcal{K}$  and formula  $\alpha$  to a theory  $\mathcal{K} \div \alpha$  while satisfying a given set of postulates (Alchourrón, Gärdenfors, and Makinson 1985; Peppas 2007). Similarly to revision, contraction can be captured in our framework using a constraint  $\mathcal{C}$  by setting  $\mathcal{C}^+ = \emptyset$  and  $\mathcal{C}^-$  to represent the information to be retracted. Contraction functions can be defined in our framework as follows.

**Definition 11.** *Let  $\mathcal{LO} = \mathcal{LO}' = \mathcal{LC}$  and  $\mathcal{LP}$  be DLs. A contraction function  $\div$  is a binary function that maps each  $\mathcal{LO}$ -ontology  $\mathcal{O}$  and each  $\mathcal{LC}$ -constraint  $\mathcal{C}$  of the form  $\mathcal{C} = (\emptyset, \mathcal{C}^-)$  to an ontology  $\mathcal{O} \div \mathcal{C}$  such that*

- (i)  $\mathcal{O} \div \mathcal{C}$  is  $\mathcal{LP}$ -optimal in  $\text{Evol}_{\mathcal{LO}'}(\mathcal{O}, \mathcal{C})$ ; and
- (ii) If  $\mathcal{C}_1^- \equiv \mathcal{C}_2^-$ , then  $\mathcal{O} \div (\emptyset, \mathcal{C}_1^-) \equiv \mathcal{O} \div (\emptyset, \mathcal{C}_2^-)$ .

Contraction postulates can also be adapted to our framework (Figure 1). Postulates C1 and C2 are self-explanatory; C3 says that if  $\mathcal{O}$  already conforms to  $\mathcal{C}$  (i.e., it does not entail any axiom in  $\mathcal{C}^-$ ), then there is no reason to change  $\mathcal{O}$ ; C4 says that tautologies are the only sentences that cannot be retracted; C5, called the *recovery postulate*, states that we can get back the initial theory by first retracting some information and then adding it back; finally, C6 is the analogue to R6.

**Theorem 12.** *Let  $\mathcal{LO} = \mathcal{LO}' = \mathcal{LC}$ , and  $\mathcal{LP}$  be DLs, and let  $\div$  be a contraction function as in Definition 11. Then,  $\div$  satisfies postulates C1-C4, and C6 in Figure 1.*

Our contraction functions satisfy all postulates except for the recovery postulate. Consider our running example and let  $\mathcal{O}_{\text{ex}}^1 = \{\beta_1, \delta_1\}$  and  $\mathcal{C}^- = \{\gamma_1\}$ . The ontology  $\mathcal{O}_{\text{ex}}^2 = \{\beta_1, \text{JointDisorder} \sqsubseteq \text{SkeletDisorder}\}$  is  $\mathcal{L}^{\text{min}}$ -optimal, so let  $\mathcal{O}_{\text{ex}}^1 \div \mathcal{C} = \mathcal{O}_{\text{ex}}^2$ . Since  $\mathcal{O}_{\text{ex}}^2 \cup \mathcal{C}^- \not\models \delta_1$ , we have  $\mathcal{O}_{\text{ex}}^2 \cup \mathcal{C}^- \not\models \mathcal{O}_{\text{ex}}^1$ , which falsifies C5. Failure to satisfy C5 is hence intuitive (and expected).

**Syntactic Repair.** We finally show that “syntactic approaches to contraction” in ontologies – often called *ontology repair* techniques (Kalyanpur et al. 2005; 2006; Schlobach et al. 2007) – can be easily captured in our framework using the empty preservation language.

---

**Algorithm 1:** Evolution for finite  $\mathcal{LP}$ 

---

**INPUT** :  $\mathcal{O}$  : satisfiable,  
 $\mathcal{C} = (\mathcal{C}^+, \mathcal{C}^-)$ :  $\mathcal{LO}'$ -conformant,  
 $\mathcal{LP}$ : finite and computable

**OUTPUT**:  $\mathcal{LO}'$ -ontology  $\mathcal{O}'$

- 1 If  $\mathcal{C}^+$  is unsatisfiable, Return  $\mathcal{O}' := \mathcal{O} \cup \mathcal{C}^+$ ;
  - 2  $\mathcal{O}_m := \max.$  subset of  $\mathcal{O}$  such that  $\mathcal{O}_m \cup \mathcal{C}^+$  is satisfiable and  $(\mathcal{O}_m \cup \mathcal{C}^+) \propto \mathcal{C}$ ;
  - 3  $\mathcal{S}^1 := \{\alpha \mid \alpha \in \text{all}_{\mathcal{LP}}(\text{sig}(\mathcal{O})), \text{ and } \mathcal{O} \models \alpha\}$ ;
  - 4  $\mathcal{S}_m := \max.$  subset of  $\mathcal{S}^1$  such that  $\mathcal{O}_m \cup \mathcal{C}^+ \cup \mathcal{S}_m$  is satisfiable and  $(\mathcal{O}_m \cup \mathcal{C}^+ \cup \mathcal{S}_m) \propto \mathcal{C}$ ;
  - 5 Return  $\mathcal{O}' := \mathcal{O}_m \cup \mathcal{C}^+ \cup \mathcal{S}_m$ .
- 

**Definition 13.** Let  $\mathcal{O}$  and  $U$  be ontologies s.t.  $\mathcal{O} \models U$ . A syntactic repair of  $\mathcal{O}$  for  $U$  is an ontology  $\mathcal{O}' \subseteq \mathcal{O}$  s.t.

- (i)  $\mathcal{O}' \not\models \alpha$  for each  $\alpha \in U$ ; and
- (ii) for all  $\beta \in \mathcal{O} \setminus \mathcal{O}'$ , there is  $\alpha \in U$  s.t.  $\mathcal{O}' \cup \{\beta\} \models \alpha$ .

Syntactic repairs are in fact evolutions:

**Proposition 14.** Let  $\mathcal{LO} = \mathcal{LO}' = \mathcal{LC}$ , and  $\mathcal{LP} = \emptyset$ . Let  $\mathcal{O}, U \in \mathcal{LO}$ , and  $\mathcal{O}' \in \mathcal{LO}'$  be a syntactic repair of  $\mathcal{O}$  for  $U$ . Then,  $\mathcal{O}' \in \text{Evol}_{\mathcal{LO}'}(\mathcal{O}, (\emptyset, U))$  is  $\mathcal{LP}$ -optimal.

Note that there can be exponentially many different syntactic repairs of  $\mathcal{O}$  for  $U$ ; thus, the non-determinism inherent to the choice of an optimal evolution already manifests itself in syntactic approaches.

## Computing Optimal Evolutions

Having established our framework, the focus in the remainder of this paper will be on the computation of  $\mathcal{LP}$ -optimal evolutions of an ontology  $\mathcal{O}$ .

Existence of an optimal evolution critically depends on the properties of preservation language  $\mathcal{LP}$ . In particular, we will make a clear distinction between *finite* and *infinite* languages, as given next.

**Definition 15.** A DL  $\mathcal{L}$  over a signature  $\Sigma$  is *finite* if for every finite  $\Sigma' \subseteq \Sigma$  there are only finitely many non-equivalent  $\mathcal{L}$ -sentences over  $\Sigma'$ . Otherwise,  $\mathcal{L}$  is *infinite*. A finite DL  $\mathcal{L}$  is *computable* if an algorithm  $\text{all}_{\mathcal{L}}$  exists that given a finite signature  $\Sigma'$  computes a set  $\text{all}_{\mathcal{L}}(\Sigma')$  of non-equivalent  $\mathcal{L}$ -sentences over  $\Sigma'$  such that any other  $\mathcal{L}$ -sentence over  $\Sigma'$  is equivalent to some  $\alpha \in \text{all}_{\mathcal{L}}(\Sigma')$ .

## Finite Preservation Languages

Many practically relevant languages are finite and computable as in of Definition 15; these include, for example, propositional logic, the language  $\mathcal{L}^{\min}$ , or the description logic *DL-Lite* (Calvanese et al. 2007).

If  $\mathcal{LP}$  is finite, an optimal  $\mathcal{LP}$ -evolution is guaranteed to exist provided that the constraints are conformant; indeed, the closure  $\text{Cl}_{\mathcal{LP}}(\mathcal{O})$  contains only finitely many non-equivalent axioms, and thus the (non-empty) quotient set  $\text{Evol}(\mathcal{O}, \mathcal{C}) \equiv_{\mathcal{LP}}$  contains finitely many equivalence classes. Furthermore, if  $\mathcal{LP}$  is computable, then Algorithm 1 computes one such  $\mathcal{LP}$ -optimal evolution.

**Theorem 16.** Let  $\mathcal{LC} \cup \mathcal{LP} \cup \mathcal{LO} \subseteq \mathcal{LO}'$  and let entailment in  $\mathcal{LO}'$  be a decidable problem; let  $\mathcal{LP}$  be finite and computable, and let  $\mathcal{C}$  be an  $\mathcal{LO}'$ -conformant  $\mathcal{LC}$ -constraint. Then, Algorithm 1 computes an  $\mathcal{LP}$ -optimal  $\mathcal{LO}'$ -evolution of a satisfiable ontology  $\mathcal{O}$  under  $\mathcal{C}$ .

Note that Algorithm 1 generalises the algorithm in (Calvanese et al. 2010) for computing so-called *Bold Evolution Semantics* for *DL-Lite* ontologies.

## Infinite Languages: Inexpressibility

Many DLs, however, are infinite in the sense of Definition 15. We next study the case where  $\mathcal{LP}$  is infinite and present inexpressibility results for  $\mathcal{FL}_0$  and  $\mathcal{EL}$ .

More precisely, we consider the case where  $\mathcal{LP}$  is either  $\mathcal{FL}_0$  or  $\mathcal{EL}$ , and where  $\mathcal{LO}$  and  $\mathcal{LO}'$  coincide with  $\mathcal{LP}$ ; for each choice of  $\mathcal{LP}$ , we provide an  $\mathcal{LO}$ -TBox  $\mathcal{T}$  and conformant constraints  $\mathcal{C}$  for which *no*  $\mathcal{LP}$ -optimal evolution of  $\mathcal{T}$  under  $\mathcal{C}$  exists. We focus on the simplest case of contraction, where  $\mathcal{C}^-$  consists of a single axiom of the form  $A \sqsubseteq B$  with  $A$  and  $B$  atomic concepts.

**Inexpressibility for  $\mathcal{FL}_0$ .** Suppose that we want to retract axiom  $\gamma_1$  in our running example from the singleton TBox  $\mathcal{T}_{\text{ex}} = \{\gamma_1\}$  while maximally preserving all  $\mathcal{FL}_0$ -consequences of  $\mathcal{T}$  w.r.t.  $\Sigma = \text{sig}(\mathcal{T}) \cup \{\text{located}\}$ . Clearly, the  $\mathcal{FL}_0$ -closure of  $\mathcal{T}$  w.r.t.  $\Sigma$  is an infinite set containing all axioms  $\text{Arthropathy} \sqcap X \sqsubseteq \text{JointFinding}$  with  $X$  an  $\mathcal{FL}_0$ -concept over  $\Sigma$ ; in particular, the following axioms  $\alpha_k$  are in the closure for each  $k \geq 1$ :

$\text{Arthropathy} \sqcap \forall \text{located}^k. \text{JointFinding} \sqsubseteq \text{JointFinding}$ .

Unfortunately, inexpressibility can already be shown in this simple setting. Intuitively, each of these axioms  $\alpha_k$  can be “recovered” without introducing the undesired consequence  $\gamma_1$ ; furthermore, no finite subset of these axioms entails the remaining ones. As a result, one would need to recover an infinite set of axioms in the closure in order to ensure maximality.

These intuitions can be made precise as given in the following lemma.

**Lemma 17.** Let  $\Sigma = \{A, B, R\}$  with  $A, B$  concepts and  $R$  a role, let  $\mathcal{T} = \{A \sqsubseteq B\}$ , and let  $\Lambda$  be the following (infinite) set of axioms:

$$\Lambda = \{A \sqcap \forall R^n. Z \sqsubseteq B \mid n \in \mathbb{N}, Z \in \{A, B\}\}.$$

Then, the following conditions hold:

- (i)  $\Lambda \subseteq \text{Cl}_{\mathcal{FL}_0}(\mathcal{T})$ ;
- (ii)  $\Lambda \not\models A \sqsubseteq B$ ;
- (iii) if a finite  $\Gamma \subseteq \text{Cl}_{\mathcal{FL}_0}(\mathcal{T})$  satisfies  $\Gamma \not\models A \sqsubseteq B$ , then there is a finite  $\Gamma' \subseteq \Lambda$  s.t.  $\Gamma' \equiv \Gamma$ ; and
- (iv) each  $\alpha \in \Lambda$  satisfies  $\Lambda \setminus \{\alpha\} \not\models \alpha$ .

Lemma 17 immediately leads to an inexpressibility result for  $\mathcal{FL}_0$ : it suffices to consider  $\mathcal{T} = \{A \sqsubseteq B\}$  and  $\mathcal{C} = (\emptyset, \mathcal{T})$  and use Lemma 17 to show that no  $\mathcal{FL}_0$ -optimal evolution of  $\mathcal{T}$  under  $\mathcal{C}$  exists.

**Theorem 18.** Let  $\mathcal{LO} = \mathcal{LO}' = \mathcal{LP} = \mathcal{FL}_0$  and let  $\mathcal{LC} = \mathcal{L}^{\min}$ . There exists an  $\mathcal{LO}$ -TBox  $\mathcal{T}$ , and an  $\mathcal{LC}$ -constraint of the form  $\mathcal{C} = (\emptyset, \{\alpha\})$  with  $\emptyset \not\models \alpha$  such that no  $\mathcal{LP}$ -optimal  $\mathcal{LO}'$ -evolution of  $\mathcal{T}$  under  $\mathcal{C}$  exists.

**Inexpressibility for  $\mathcal{EL}$ .** The logic  $\mathcal{EL}$  is not only more useful for ontology modeling than  $\mathcal{FL}_0$ , but fortunately it also behaves better in terms of expressibility of optimal evolutions. For example, consider the retraction of the axiom  $\gamma_1$  in  $\mathcal{T}_{\text{ex}}$ , which illustrated our inexpressibility result for  $\mathcal{FL}_0$ . The  $\mathcal{EL}$  analogue to the  $\mathcal{FL}_0$ -axioms  $\alpha_k$  in our example are the following axioms  $\alpha'_k$ :

**Arthropaty  $\sqcap \exists \text{located}^k . \text{JointFinding} \sqsubseteq \text{JointFinding}$ .**

As in the case of  $\mathcal{FL}_0$ , any subset of these axioms can be included in an evolution of  $\mathcal{T}_{\text{ex}}$  without regaining  $\gamma_1$ ; in contrast to the previous case, however, it suffices to recover the following axiom, which entails all the others:

**Arthropaty  $\sqcap \exists \text{located} . \top \sqsubseteq \text{Joint}$ .**

Inexpressibility results for  $\mathcal{EL}$  originate from non-trivial interactions between cyclic axioms of the form  $A \sqsubseteq \exists R.A$  and of the form  $\exists R.B \sqsubseteq B$ . The former axiomatises existence of  $R$ -connected instances of  $A$  and entails all axioms of the form  $A \sqsubseteq \exists R^n.A$ ; the latter axiomatises recursion and entails  $\exists R^n.B \sqsubseteq B$  for each  $n \in \mathbb{N}$ . The “harmful” interaction between these kinds of axioms is formally described by the following lemma.

**Lemma 19.** *Let  $\mathcal{T}$  be the following  $\mathcal{EL}$ -TBox:*

$$\mathcal{T} = \{Z \sqsubseteq \exists R.A, A \sqsubseteq \exists R.A, \exists R.B \sqsubseteq B, A \sqsubseteq B\}.$$

Furthermore, for each  $k \in \mathbb{N}$ , let

$$\begin{aligned} \alpha_k &= Z \sqsubseteq \exists R^k.(A \sqcap B); \\ \beta_k &= Z \sqsubseteq \exists R^k.B; \\ \Lambda_k &= \{\alpha_i \mid 1 \leq i \leq k\}. \end{aligned}$$

Finally, let  $\Lambda = \bigcup_{k=1}^{\infty} \Lambda_k$  and let  $\mathcal{T}' = \mathcal{T} \setminus \{A \sqsubseteq B\}$ . Then, the following conditions hold:

- (i)  $\Lambda \subseteq \text{Cl}_{\mathcal{EL}}(\mathcal{T})$ ;
- (ii)  $\mathcal{T}' \cup \Lambda \not\models A \sqsubseteq B$ ;
- (iii) If a finite  $\Gamma \subseteq \text{Cl}_{\mathcal{EL}}(\mathcal{T})$  satisfies  $\mathcal{T}' \subseteq \Gamma$  and  $\Gamma \not\models A \sqsubseteq B$ , then  $\mathcal{T}' \cup \Lambda_k \models \Gamma$  for some  $k \in \mathbb{N}$ ; and
- (iv)  $\mathcal{T}' \cup \Lambda_k \not\models \beta_{k+1}$  for each  $k \in \mathbb{N}$ .

Lemma 19 implies that, regardless of how large a subset  $\Gamma \subseteq \text{Cl}_{\mathcal{EL}}(\mathcal{T})$  we pick as the result of contracting  $\mathcal{T}$  with  $A \sqsubseteq B$  while preserving  $\mathcal{T}'$ , we can always find  $k \geq 1$  such that  $\Gamma \not\models \beta_{k+1}$  and adding  $\beta_{k+1}$  to  $\Gamma$  will not make us recover the undesired entailment  $A \sqsubseteq B$ . The following inexpressibility result immediately follows.

**Theorem 20.** *Let  $\mathcal{LO} = \mathcal{LO}' = \mathcal{LP} = \mathcal{EL}$  and let  $\mathcal{LC} = \mathcal{L}^{\text{min}}$ . There exists an  $\mathcal{LO}$ -TBox  $\mathcal{T}$ , and an  $\mathcal{LC}$ -constraint  $\mathcal{C} = (\mathcal{C}^+, \{\alpha\})$  with  $\mathcal{C}^+ \subseteq \mathcal{T}$  and  $\mathcal{C}^+ \not\models \alpha$  such that no  $\mathcal{LP}$ -optimal  $\mathcal{LO}'$ -evolution of  $\mathcal{T}$  under  $\mathcal{C}$  exists.*

### Contraction in $\mathcal{EL}$

Lemma 19 suggests that inexpressibility can be overcome by constraining the structure of  $\mathcal{T}$ ; more precisely, one could devise sufficient conditions for precluding in  $\text{Cl}_{\mathcal{EL}}(\mathcal{T})$  either axioms of the form  $A \sqsubseteq \exists R.A$  or axioms of the form  $\exists R.B \sqsubseteq B$ . The former can be achieved

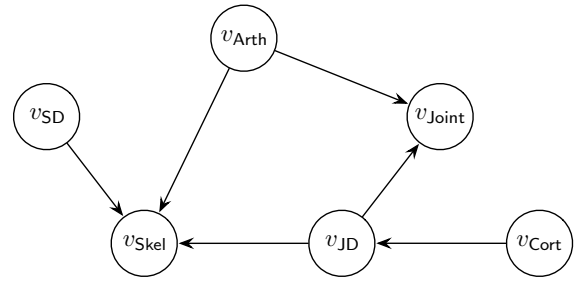


Figure 2: Graph for  $\mathcal{O}_{\text{ex}} = \{\beta_1, \beta_2, \gamma_1, \delta_1, \dots, \delta_5\}$ . Abbreviations: “SD”, “Arth”, “Skel”, “JD” and “Cort” stand for SkeletDisorder, Arthropaty, Skeleton, JointDisorder, and Cortisone, respectively.

with a suitable *acyclicity condition*; the latter involves *precluding recursion*.

In this section, we study  $\mathcal{EL}$ -contraction under each of these alternatives. For convenience, we restrict ourselves to  $\mathcal{EL}$ -TBoxes in normal form. We consider w.l.o.g. the case where  $\mathcal{C}^- = \{\alpha\}$  with  $\alpha \in \mathcal{L}^{\text{min}}$ ; furthermore, instead of assuming  $\mathcal{C}^+ = \emptyset$ , we consider a slightly more general setting where  $\mathcal{C}^+$  may contain a “protected” subset of axioms in  $\mathcal{T}$  that must survive the contraction.

In our technical results, we restrict ourselves to a preservation language  $\mathcal{LP}$  that is a fragment of  $\mathcal{EL}$ , and which we call  $\mathcal{EL}^c$ .

**Definition 21.** *The DL  $\mathcal{EL}^c$  consists of all  $\mathcal{EL}$ -TBoxes containing only axioms of the form  $Z \sqsubseteq \exists w.Z'$ , or of the form  $\exists w.Z' \sqsubseteq Z$ , where  $w$  is a word of roles and  $Z, Z'$  are either atomic concepts or  $\top$ .<sup>2</sup>*

Essentially,  $\mathcal{EL}^c$  disallows conjunction, but allows for arbitrarily deep nesting of existential concepts on both the left and right hand side of axioms. Thus,  $\mathcal{EL}^c$  is an infinite language, in the sense of Definition 15.

Although extending this preservation language with conjunction might increase the size of the computed optimal evolutions by an exponential factor, we believe that  $\mathcal{EL}^c$  is a sufficiently large fragment of  $\mathcal{EL}$  to illustrate the key issues involving existence of optimal contractions. (Note that the inexpressibility result that follows from Lemma 19 only relies on the preservation of the  $\mathcal{EL}^c$  entailments  $\beta_k = Z \sqsubseteq \exists R^k.B$ ). We conjecture that the contraction algorithms presented in this section can be extended to the case where  $\mathcal{LP} = \mathcal{EL}$ , and leave the details for future work.

### Contraction in acyclic $\mathcal{EL}$

We next study contraction for  $\mathcal{EL}$  TBoxes  $\mathcal{T}$  under a suitable acyclicity condition.

**Acyclicity** Our notion of acyclicity is formulated in terms of the canonical model  $\mathcal{I}_{\mathcal{T}}$  of  $\mathcal{T}$ , and can be checked in polynomial time w.r.t. the size of  $\mathcal{T}$ .

<sup>2</sup>Note that  $w$  could be the empty word, in which case we have a subsumption between atomic concepts.

**Definition 22.** A normalised  $\mathcal{EL}$ -TBox  $\mathcal{T}$  with canonical model  $\mathcal{I}_{\mathcal{T}} = (\Delta^{\mathcal{I}_{\mathcal{T}}}, \mathcal{I}_{\mathcal{T}})$  is acyclic if the graph  $(V, E)$  consisting of nodes  $V = \{v_A \mid v_A \in \Delta^{\mathcal{I}_{\mathcal{T}}}\}$  and (directed) edges  $E = \{(v_A, v_B) \mid (v_A, v_B) \in R^{\mathcal{I}_{\mathcal{T}}}\}$  for some role  $R$  is acyclic. With  $\delta(\mathcal{T})$  we denote the length of the longest path in this graph. By  $\mathcal{EL}^a$  we denote the DL consisting of all acyclic  $\mathcal{EL}$  TBoxes in normal form.<sup>3</sup>

Our acyclicity condition generalises the usual acyclicity condition in  $\mathcal{EL}$ -terminologies (Baader et al. 2003; Konev, Walther, and Wolter 2008); that is, the normalisation of each acyclic  $\mathcal{EL}$ -terminology is acyclic as in Definition 22. In particular, the positive results presented in this section could be applicable to reference bio-medical ontologies such as SNOMED and (the  $\mathcal{EL}$  versions of) NCI, which are acyclic terminologies.

For instance, note that the (normalisation of) example ontology  $\mathcal{O}_{\text{ex}} = \{\beta_1, \beta_2, \gamma_1, \delta_1, \dots, \delta_5\}$  about skeletal disorders is also acyclic. The interesting fragment of the graph corresponding to the (normalisation of)  $\mathcal{O}_{\text{ex}}$  as given in Definition 22 is depicted in Figure 2.<sup>4</sup>

Acyclicity of  $\mathcal{T}$  immediately ensures that the closure  $\text{Cl}_{\mathcal{EL}}(\mathcal{T})$  cannot contain axioms of the form  $A \sqsubseteq \exists w.A$  with  $A$  atomic and  $w$  a word of atomic roles. More generally, acyclicity of  $\mathcal{T}$  establishes a bound on the quantifier depth of concepts that can occur on the right-hand-side of an axiom derived from  $\mathcal{T}$ .

**Lemma 23.** Let  $\mathcal{T} \in \mathcal{EL}^a$ , let  $A$  be an atomic concept or  $\top$ , and let  $D$  an arbitrary  $\mathcal{EL}$ -concept. If  $\mathcal{T} \models A \sqsubseteq D$ , then  $\text{depth}(D) \leq \delta(\mathcal{T})$ .

In contrast to concepts on the right-hand side of derived axioms, the quantifier depth of concepts on the left-hand side is not limited by acyclicity, which makes  $\mathcal{EL}^a$  an infinite language (e.g., recursive axioms such as  $\exists R.B \sqsubseteq B$ , which entails  $\exists R^n.B \sqsubseteq B$  for each  $n \in \mathbb{N}$ , are allowed in  $\mathcal{T}$ ). Adding an axiom of the form  $C \sqsubseteq D$  to an acyclic  $\mathcal{T}$  where  $\text{depth}(C) > \delta(\mathcal{T})$ , however, will not introduce new subsumption relations between atomic concepts in  $\mathcal{T}$ .

**Lemma 24.** Let  $\mathcal{T} \in \mathcal{EL}^a$ , let  $C$  and  $D$  be  $\mathcal{EL}$ -concepts, and assume that  $\text{depth}(C) > \delta(\mathcal{T})$ . If  $\mathcal{T} \not\models A \sqsubseteq B$  with  $A$  and  $B$  atomic concepts, then  $\mathcal{T} \cup \{C \sqsubseteq D\} \not\models A \sqsubseteq B$ .

**The contraction algorithm** Algorithm AContr (see Algorithm 2) computes an  $\mathcal{EL}^c$ -optimal contraction  $\mathcal{T}'$  of an  $\mathcal{EL}^a$ -TBox  $\mathcal{T}$ . The algorithm works as follows.

In Step 1, a TBox  $\mathcal{T}_m \subseteq \mathcal{T}$  conforming to  $\mathcal{C}$  is (non-deterministically) selected; this ensures that the output  $\mathcal{T}'$  preserves a maximal syntactic subset of  $\mathcal{T}$ . (Note that existence of such  $\mathcal{T}_m$  is ensured by the preconditions of the algorithm). Steps 2 and 3 compute the subset  $\mathcal{S}^1 \cup \mathcal{S}^2$  of  $\mathcal{EL}^c$ -axioms in  $\text{Cl}_{\mathcal{EL}^c}(\mathcal{T})$  of quantifier depth at most  $\delta(\mathcal{T})$ ; clearly,  $\mathcal{S}^1 \cup \mathcal{S}^2$  is finite and exponential

<sup>3</sup>The  $a$  in  $\mathcal{EL}^a$  stands for “acyclic”.

<sup>4</sup>For simplicity, we did not depict in Figure 2 the nodes corresponding to  $\top$ , Disorder, Steroid, ArthropatyTest and JointFinding, as well as the nodes corresponding to the fresh concepts introduced by normalisation.

---

### Algorithm 2: AContr

---

**INPUT** :  $\mathcal{T} \in \mathcal{EL}^a$ ,  $\mathcal{C} = (\mathcal{C}^+, \mathcal{C}^-)$ , such that  $\mathcal{C}^+ \subseteq \mathcal{T}$ ,  $\mathcal{C}^- = \{\alpha\}$  for  $\alpha \in \mathcal{L}^{\text{min}}$ , and  $\mathcal{T} \not\models \mathcal{C}^-$ ,  $\mathcal{C}^+ \propto \mathcal{C}^-$

**OUTPUT**:  $\mathcal{EL}$  TBox  $\mathcal{T}'$

- 1  $\mathcal{T}_m := \text{max. subset of } \mathcal{T} \text{ s.t. } \mathcal{T}_m \propto \mathcal{C}^- \ \& \ \mathcal{C}^+ \subseteq \mathcal{T}_m$ ;
  - 2  $\mathcal{S}^1 := \{\alpha = Z_1 \sqsubseteq \exists w.Z_2 \mid \mathcal{T} \models \alpha \text{ and } |w| \leq \delta(\mathcal{T})\}$ ;
  - 3  $\mathcal{S}^2 := \{\alpha = \exists w.Z_1 \sqsubseteq Z_2 \mid \mathcal{T} \models \alpha \text{ and } |w| \leq \delta(\mathcal{T})\}$ ;
  - 4  $\mathcal{S}_m := \text{max. subset of } \mathcal{S}^1 \cup \mathcal{S}^2 \text{ s.t. } \mathcal{T}_m \cup \mathcal{S}_m \propto \mathcal{C}^-$ ;
  - 5  $\mathcal{S}^3 := \{\alpha = \exists w.Z_1 \sqsubseteq Z_2 \mid \mathcal{T} \models \alpha \text{ and } |w| \in [\delta(\mathcal{T}) + 1, 2 \times \delta(\mathcal{T}) + 1]\}$ ;
  - 6 Return  $\mathcal{T}' = \mathcal{T}_m \cup \mathcal{S}_m \cup \mathcal{S}^3$ .
- 

in size. In Step 4, the algorithm computes a maximal  $\mathcal{S}_m \subseteq \mathcal{S}^1 \cup \mathcal{S}^2$  that can be added to  $\mathcal{T}_m$  without regaining  $\alpha$ . At this point, Algorithm AContr needs to consider the axioms in  $\text{Cl}_{\mathcal{EL}^c}(\mathcal{T})$  with concepts of quantifier depth greater than  $\delta(\mathcal{T})$ . By Lemma 23, no such axiom of the form  $A_1 \sqsubseteq \exists w.A_2$  exists; however,  $\mathcal{T}$  might entail  $\mathcal{EL}^c$ -axioms of the form  $\exists w.A_1 \sqsubseteq A_2$  with  $|w| > \delta(\mathcal{T})$ , and by Lemma 24 all such axioms must also be entailed by each optimal evolution (since they cannot make us recover  $\alpha$ ). Even if there can be infinitely many such axioms, Algorithm AContr only computes in Step 5 those of quantifier depth at most  $2 \times \delta(\mathcal{T}) + 1$ , which we prove sufficient. The intuition behind this bound is given by the following example.

**Example 25.** Consider the following TBox  $\mathcal{T}$ :

$$\mathcal{T} = \{A \sqsubseteq \exists R.C, C \sqsubseteq \exists R.B, \exists R.B \sqsubseteq B\}$$

Clearly,  $\mathcal{T}$  is acyclic with  $\delta(\mathcal{T}) = 2$ . Let us apply Algorithm AContr to  $\mathcal{T}$ ,  $\mathcal{C}^+ = \{A \sqsubseteq \exists R.C, C \sqsubseteq \exists R.B\}$  and  $\mathcal{C}^- = \{A \sqsubseteq B\}$ .<sup>5</sup> We have  $\mathcal{T}_m = \mathcal{C}^+$  and

$$\begin{aligned} \mathcal{S}^1 &= \mathcal{T}_m \cup \{C \sqsubseteq B, A \sqsubseteq \exists R.B\}; \\ \mathcal{S}^2 &= \{\exists R.Z \sqsubseteq B, \exists R^2.Z \sqsubseteq B \mid Z \in \{A, B, C\}\}; \\ \mathcal{S}_m &= \mathcal{S}^1 \cup \{\exists R.A \sqsubseteq B, \exists R^2.A \sqsubseteq B, \exists R^2.C \sqsubseteq B\}; \\ \mathcal{S}^3 &= \{\exists R^k.Z \sqsubseteq B \mid Z \in \{A, B, C\}, 3 \leq k \leq 5\}. \end{aligned}$$

The algorithm then returns  $\mathcal{T}' = \mathcal{T}_m \cup \mathcal{S}_m \cup \mathcal{S}^3$ . To see that  $\mathcal{T}'$  is optimal, consider, for example, axiom  $\beta = \exists R^{10}.B \sqsubseteq B$ , which follows from  $\mathcal{T}$ . Note that we can decompose  $k = 10$  as a sum of numbers from 3 to 5 as follows:  $k = 3 \times 2 + 4$ ; since  $\exists R^3.B \sqsubseteq B$  and  $\exists R^4.B \sqsubseteq B$  are in  $\mathcal{S}^3$ , we have  $\mathcal{T}' \models \beta$ .

The following lemma makes the intuitions in Example 25 precise. In particular, it shows that if  $\beta = \exists w.A \sqsubseteq B$  with  $|w| > 2 \times \delta(\mathcal{T}) + 1$  follows from  $\mathcal{T}$ , we can “decompose”  $w$  and derive  $\beta$  from axioms in  $\mathcal{S}^3$ .

**Lemma 26.** Let  $\mathcal{T} \in \mathcal{EL}^a$  and let  $\alpha = \exists w.Z_n \sqsubseteq Z_0$  be an  $\mathcal{EL}^c$ -axiom with  $|w| > (2 \times \delta(\mathcal{T}) + 1)$  s.t.  $\mathcal{T} \models \alpha$ .

<sup>5</sup>For simplicity, we do not include in sets  $\mathcal{S}^i$  axioms that are already entailed by  $\mathcal{T}_m$



---

**Algorithm 3: NRContr**

---

**INPUT** :  $\mathcal{T} \in \mathcal{EL}^{nr}$ ,  $\mathcal{C} = (\mathcal{C}^+, \mathcal{C}^-)$ , such that  
 $\mathcal{C}^+ \subseteq \mathcal{T}$ ,  $\mathcal{C}^- = \{\alpha\}$  for  $\alpha \in \mathcal{L}^{\min}$ , and  
 $\mathcal{T} \not\sqsubseteq \mathcal{C}^-$ ,  $\mathcal{C}^+ \sqsubseteq \mathcal{C}^-$

**OUTPUT**:  $\mathcal{EL}^{nr}$  TBox  $\mathcal{T}'$

- 1  $\mathcal{T}_m := \max.$  subset of  $\mathcal{T}$  s.t.  $\mathcal{T}_m \sqsubseteq \mathcal{C}^-$  &  $\mathcal{C}^+ \subseteq \mathcal{T}_m$ ;
- 2  $\mathcal{S}_m := \max.$  subset of  $\text{BCl}(\mathcal{T})$  s.t.  $\mathcal{T}_m \cup \mathcal{S}_m \sqsubseteq \mathcal{C}^-$ ;
- 3 Return  $\mathcal{T}' := \mathcal{T}_m \cup \mathcal{S}_m$ .

---

Then, there exists an integer  $\ell \geq 2$ , subwords  $u_1, \dots, u_\ell$  of  $w$ , and concepts  $\{Y_0, \dots, Y_\ell\} \subseteq \text{sig}(\mathcal{T}) \cup \{\top\}$  where  $Y_\ell = Z_n$  and  $Y_0 = Z_0$  such that

- (i)  $w = u_1 \circ \dots \circ u_\ell$  with  $\delta(\mathcal{T}) < |u_j| \leq (2 \times \delta(\mathcal{T}) + 1)$ ;
- (ii)  $\mathcal{T} \models \exists u_j.Y_j \sqsubseteq Y_{j-1}$  for each  $j \in [1, \ell]$ .

We can now show the correctness of our algorithm.

**Theorem 27.** *Algorithm AContr computes an  $\mathcal{EL}^c$ -optimal evolution  $\mathcal{T}'$  of  $\mathcal{T} \in \mathcal{EL}^a$  under  $\mathcal{C}$ . Furthermore the size of  $\mathcal{T}'$  is exponential in the size of  $\mathcal{T}$ .*

### Contraction in non-recursive $\mathcal{EL}$

We next study contraction for non-recursive  $\mathcal{EL}$ -TBoxes. The simplest non-recursive fragment of  $\mathcal{EL}$ , which we call  $\mathcal{EL}^{nr}$ , is defined as follows.

**Definition 28.** *The DL  $\mathcal{EL}^{nr}$  consists of all normalised  $\mathcal{EL}$ -TBoxes where neither (i)  $\top$  nor (ii) concepts of the form  $\exists R.C$  occur on the left hand side of axioms.<sup>6</sup>*

Note that  $\mathcal{EL}^{nr}$  can express cyclic axioms of the form  $A \sqsubseteq \exists R.A$  and hence  $\mathcal{EL}^{nr} \not\subseteq \mathcal{EL}^a$ . Since  $\mathcal{EL}^a$  allows for axioms with  $\exists R.C$  on any side, we also have  $\mathcal{EL}^a \not\subseteq \mathcal{EL}^{nr}$ . Furthermore, the normalisation of an  $\mathcal{EL}$ -TBox  $\mathcal{T}$  satisfying properties (i) and (ii) in Definition 28 leads to a TBox in  $\mathcal{EL}^{nr}$ . Note also that  $\mathcal{EL}^{nr}$  is an infinite language, e.g., the ontology  $\{A \sqsubseteq \exists R.A\}$  entails infinitely many axioms of the form  $A \sqsubseteq \exists R^n.A$  for  $n \in \mathbb{N}$ .

**The contraction algorithm** We describe algorithm NRContr (see Algorithm 3), which computes an  $\mathcal{EL}^c$ -optimal contraction  $\mathcal{T}'$  of a TBox  $\mathcal{T} \in \mathcal{EL}^{nr}$ .

Step 1 is identical to Step 1 in Algorithm AContr. In Step 2, Algorithm NRContr computes a maximal subset of axioms in the basic closure of  $\mathcal{T}$  that can be added to  $\mathcal{T}_m$  without recovering the undesired entailment  $\alpha$ . The basic closure of any  $\mathcal{EL}$ -TBox contains only  $\mathcal{EL}^{nr}$ -axioms, so the output  $\mathcal{T}'$  is an  $\mathcal{EL}^{nr}$ -TBox; furthermore, the basic closure is of size polynomial in the size of  $\mathcal{T}$  (and hence so is the output  $\mathcal{T}'$ ). We next illustrate the intuition behind this algorithm with an example.

**Example 29.** *Consider the following  $\mathcal{EL}^{nr}$ -TBox:*

$$\mathcal{T} = \{Z \sqsubseteq \exists R.A, A \sqsubseteq \exists R.A, A \sqsubseteq B\}.$$

We apply NRContr to  $\mathcal{T}$ ,  $\mathcal{C}^+ = \{Z \sqsubseteq \exists R.A, A \sqsubseteq \exists R.A\}$ , and  $\mathcal{C}^- = \{A \sqsubseteq B\}$ . We then have that  $\mathcal{T}_m = \mathcal{C}^+$  and

<sup>6</sup>The *nr* in  $\mathcal{EL}^{nr}$  stands for “non-recursive”.

$ \mathcal{T} \setminus \mathcal{T}_m $	$ \mathcal{S}_m $	$ \mathcal{S}_m \cap \mathcal{L}^{\min} $	Time (s)	# of
	(max/avg/min)	(max/avg/min)	(avg)	tests
1	150/24/0	52/5/0	135	52
2	282/76/7	96/24/0	217	51
3	616/206/9	195/70/0	176	51
4	822/447/92	257/138/26	169	39
5	826/530/281	281/162/75	165	42

Table 1: Summary of experimental results

$\mathcal{S}_m = \{Z \sqsubseteq \exists R.B, A \sqsubseteq \exists R.B\}$ ; hence, the algorithm returns  $\mathcal{T}' = \mathcal{T}_m \cup \mathcal{S}_m$ .<sup>7</sup>

To see that  $\mathcal{T}'$  is  $\mathcal{EL}^c$ -optimal, we make two observations. First, a TBox  $\mathcal{T} \in \mathcal{EL}^{nr}$  cannot entail  $\mathcal{EL}^c$ -axioms of the form  $\exists w.C \sqsubseteq C$ , unless  $C = \top$ , which are then tautological; note that  $\mathcal{EL}^{nr}$  does not allow for  $\top$  on the l.h.s. of axioms, and hence  $\mathcal{T} \models \top \sqsubseteq C$  implies  $C = \top$ . Second, although  $\mathcal{T}$  entails infinitely many axioms  $\alpha_k = A \sqsubseteq \exists R^k.B$  for  $k \in \mathbb{N}$ , these are entailed by axioms  $A \sqsubseteq \exists R.A$  and  $A \sqsubseteq \exists R.B$ , which are in  $\mathcal{T}'$ .

These intuitions can be made precise, and we can then show correctness of our contraction algorithm.<sup>8</sup>

**Theorem 30.** *Algorithm NRContr computes an  $\mathcal{EL}^c$ -optimal evolution  $\mathcal{T}'$  of  $\mathcal{T} \in \mathcal{EL}^{nr}$  under  $\mathcal{C}$ . Furthermore the size of  $\mathcal{T}'$  is polynomial in the size of  $\mathcal{T}$ .*

## Experiments

We have implemented an optimised version of Algorithm 1 for the particular case of contraction for a TBox  $\mathcal{T}$  and an  $\mathcal{L}^{\min}$ -axiom  $\alpha$  (i.e., where  $\mathcal{C}^+ \subseteq \mathcal{T}$  and  $\mathcal{C}^- = \{\alpha\}$ ). Concerning the preservation language, we have chosen the (finite) language that extends  $\mathcal{L}^{\min}$  with all axioms of the form  $D \sqsubseteq E$ , with  $D$  and  $E$  (possibly complex) subconcepts occurring in  $\mathcal{T}$ .

In this setting, Step 2 in Algorithm 1 amounts to computing a syntactic repair  $\mathcal{T}_m$  of  $\mathcal{T}$  for  $\alpha$ . Our implementation uses the reasoner HermiT (Motik, Shearer, and Horrocks 2009) and the OWL API facility for computing justifications (Kalyanpur et al. 2007).

We have applied our algorithm to a fragment of SNOMED with 6802 atomic concepts. In each test, we have selected an entailed subsumption relationship  $\alpha$  at random and computed the corresponding contraction; we have recorded the size (number of axioms) in  $\mathcal{T} \setminus \mathcal{T}_m$  (Step 2 in Algorithm 1), the size of  $\mathcal{S}_m$  (Step 4 Algorithm 1), the size of the  $\mathcal{L}^{\min}$ -subset of  $\mathcal{S}_m$  and the time overhead w.r.t. computing a syntactic repair (i.e., the computation time for Steps 3-5 in Algorithm 1).

Table 1 summarises the obtained results. Since there is a clear correlation between the number of axioms deleted by the syntactic repair (i.e.,  $|\mathcal{T} \setminus \mathcal{T}_m|$ ) and the number of recovered entailments (i.e.,  $|\mathcal{S}_m|$ ), we have grouped our tests according to  $|\mathcal{T} \setminus \mathcal{T}_m|$ .

<sup>7</sup>For simplicity, we do not include in  $\mathcal{S}_m$  axioms that follow already from  $\mathcal{T}_m$ .

<sup>8</sup>[www.inf.unibz.it/~zheleznyakov/krfull.pdf](http://www.inf.unibz.it/~zheleznyakov/krfull.pdf)

$\mathcal{LO} = \mathcal{LO}'$	$\mathcal{C}^- \in \mathcal{L}^{\min}$	$\mathcal{LP}$	Evol. size
$\mathcal{FL}_0$	$\{\alpha\}$	$\mathcal{FL}_0$	Inexpressible
$\mathcal{EL}$	$\{\alpha\}$	$\mathcal{EL}$	Inexpressible
$\mathcal{EL}^a$	$\{\alpha\}$	$\mathcal{EL}^c$	Exponential
$\mathcal{EL}^{nr}$	$\{\alpha\}$	$\mathcal{EL}^c$	Polynomial

Table 2: Size of computed optimal evolutions, where  $\alpha$  is an atomic subsumption of the form  $A \sqsubseteq B$

Note that  $\mathcal{S}_m$  contains surprisingly many axioms, especially when considering syntactic repairs that involve the deletion of several axioms from  $\mathcal{T}$ . To estimate the degree of redundancy in  $\mathcal{S}_m$  we have checked for each axiom  $\alpha \in \mathcal{S}_m$  whether  $\mathcal{T}_m \cup (\mathcal{S}_m \setminus \{\alpha\}) \models \alpha$  and found that in more than 95% of cases such entailment does not hold, and hence  $\alpha$  is likely to be non-redundant. Thus, a purely syntactic approach to contraction would result in a substantial and unnecessary loss of information; as shown in the table, such loss of information is already significant if we consider  $\mathcal{L}^{\min}$  as preservation language.

Finally, note that the computation of optimal contractions implies an overhead of 2-4 minutes on average. Moreover, this overhead does not depend on the size of  $\mathcal{S}_m$ . These times are promising, considering that our implementation is an early-stage prototype.

## Conclusion and Future Work

We have presented a logic-based framework for ontology evolution that can capture revision and contraction at a fine-grained deductive level. Our framework is novel and it opens many possibilities for future research. In particular, many challenging problems are left open; these include decidability of checking whether an optimal evolution exists, and complexity of computing optimal evolutions, among others. The relationships between the problem of computing optimal evolutions and other relevant reasoning problems, such as computing the logical difference between DL ontologies (Konev, Walther, and Wolter 2008), also need to be explored.

We have studied contraction for the DLs  $\mathcal{FL}_0$  and  $\mathcal{EL}$  and shown that, in general, optimal contractions cannot be expressed using finitely many axioms. Note that one could potentially overcome these inexpressibility results by allowing the “target” language  $\mathcal{LO}'$  of the evolution  $\mathcal{O}'$  to be a more powerful than the language  $\mathcal{LO}$  in which the original ontology  $\mathcal{O}$  is expressed; however, on the one hand,  $\mathcal{LO}'$  might have much less favourable computational properties than  $\mathcal{LO}$  and, on the other hand, it might not be possible to perform further contractions on the evolved ontology  $\mathcal{O}'$ . Furthermore, we conjecture that the inexpressibility results presented in this paper for  $\mathcal{LO} = \mathcal{EL}$  and  $\mathcal{LO} = \mathcal{FL}_0$  hold even if  $\mathcal{LO}'$  is an expressive DL such as  $\mathcal{SHIQ}$  (Horrocks, Sattler, and Tobies 2000).

We have devised sufficient conditions for existence of finite optimal contractions and proposed suitable contraction algorithms for such cases (see Table 2 for a

summary of our results). We are currently working on relaxing these sufficient conditions and extending them towards  $\mathcal{EL}^{++}$  (the DL underpinning the OWL 2 EL profile), and we are also investigating the applicability of our framework to the contraction of ABox assertions.

To test the feasibility of our approach in practice, we have performed contraction experiments on a fragment of SNOMED using finite preservation languages. Our results suggest that syntactic approaches to contraction (i.e., ontology repair techniques) lead to a significant and unnecessary loss of (non-redundant) information. Although from a practical point of view, users may intend to recover only a part of all this missing information, understanding the impact of changes is important in ontology modeling (Konev, Walther, and Wolter 2008; Jiménez-Ruiz et al. 2011), and knowing which entailments could be “harmlessly” regained can be very valuable. Finally, we are planning to integrate our implementation in the CONTENTCVS ontology versioning system (Jiménez-Ruiz et al. 2011) and to make our contraction algorithms for  $\mathcal{EL}$  practical by making them more “goal oriented” as well as by exploiting ontology modularisation techniques (Cuenca Grau et al. 2008a).

## Acknowledgements

Cuenca Grau and Jimenez-Ruiz are supported by the EPSRC project LOGMAP. Cuenca Grau is supported by a Royal Society University Research Fellowship. Kharlamov is supported by the EPSRC EP/G004021/1, EP/H017690/1, and ERC FP7 grant Webdam (n. 226513). Kharlamov and Zheleznyakov are supported by the EU project ACSI (FP7-ICT-257593).

## References

- Alchourrón, C. E.; Gärdenfors, P.; and Makinson, D. 1985. On the logic of theory change: Partial meet contraction and revision functions. *J. Symb. Log.* 50(2):510–530.
- Baader, F.; Calvanese, D.; McGuinness, D. L.; Nardi, D.; and Patel-Schneider, P. F., eds. 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.
- Baader, F.; Brandt, S.; and Lutz, C. 2005. Pushing the EL envelope. In *Proc. of IJCAI*, 364–369.
- Cadoli, M.; Donini, F. M.; Liberatore, P.; and Schaerf, M. 1999. The size of a revised knowledge base. *Artif. Intell.* 115(1):25–64.
- Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2007. Tractable reasoning and efficient query answering in Description Logics: The DL-Lite family. *Journal of Automated Reasoning* 39(3):385–429.
- Calvanese, D.; Kharlamov, E.; Nutt, W.; and Zheleznyakov, D. 2010. Evolution of DL-Lite knowledge bases. In *International Semantic Web Conference (1)*, 112–128.
- Cuenca Grau, B.; Horrocks, I.; Kazakov, Y.; and Sattler,

- U. 2008a. Modular reuse of ontologies: Theory and practice. *J. Artif. Intell. Res. (JAIR)* 31:273–318.
- Cuenca Grau, B.; Horrocks, I.; Motik, B.; Parsia, B.; Patel-Schneider, P.; and Sattler, U. 2008b. OWL 2: The next step for OWL. *Journal of Web Semantics* 6(4):309–322.
- Flouris, G.; Manakanatas, D.; Kondylakis, H.; Plexousakis, D.; and Antoniou, G. 2008. Ontology change: classification and survey. *Knowledge Eng. Review* 23(2):117–152.
- Fridman Noy, N.; Kunnatur, S.; Klein, M.; and Musen, M. 2004. Tracking changes during ontology evolution. In *Proc. of ISWC*, 259–273.
- Giacomo, G. D.; Lenzerini, M.; Poggi, A.; and Rosati, R. 2009. On instance-level update and erasure in Description Logic ontologies. *Journal Logic and Computation* 19(5):745–770.
- Gonçalves, R. S.; Parsia, B.; and Sattler, U. 2011. Analysing the evolution of the NCI thesaurus. In *Proc. of CBMS*, 1–6.
- Haase, P., and Stojanovic, L. 2005. Consistent evolution of OWL ontologies. In *Proc. of ESWC*, 182–197.
- Hartung, M.; Kirsten, T.; and Rahm, E. 2008. Analyzing the evolution of life science ontologies and mappings. In *Proc. of DILS*, 11–27.
- Horrocks, I.; Patel-Schneider, P. F.; and van Harmelen, F. 2003. From *SHIQ* and RDF to OWL: the making of a web ontology language. *Journal of Web Semantics* 1(1):7–26.
- Horrocks, I.; Sattler, U.; and Tobies, S. 2000. Practical reasoning for very expressive Description Logics. *Logic Journal of the IGPL* 8(3):239–263.
- Jiménez-Ruiz, E.; Grau, B. C.; Horrocks, I.; and Llavori, R. B. 2011. Supporting concurrent ontology development: Framework, algorithms and tool. *Data Knowl. Eng.* 70(1):146–164.
- Kalyanpur, A.; Parsia, B.; Sirin, E.; and Hendler, J. A. 2005. Debugging unsatisfiable classes in OWL ontologies. *Journal of Web Semantics* 3(4):268–293.
- Kalyanpur, A.; Parsia, B.; Sirin, E.; and Grau, B. C. 2006. Repairing unsatisfiable concepts in OWL ontologies. In *Proc. of ESWC*, 170–184.
- Kalyanpur, A.; Parsia, B.; Horridge, M.; and Sirin, E. 2007. Finding all justifications of OWL DL entailments. In *Proc. of ISWC*, 267–280.
- Katsuno, H., and Mendelzon, A. O. 1991. On the difference between updating a knowledge base and revising it. In *Proc. of KR*, 387–394.
- Kharlamov, E., and Zheleznyakov, D. 2011. Capturing instance level ontology evolution for DL-Lite. In *International Semantic Web Conference (1)*, 321–337.
- Konev, B.; Walther, D.; and Wolter, F. 2008. The logical difference problem for Description Logic terminologies. In *Proc. of IJCAR*, 259–274.
- Lenzerini, M., and Savo, D. F. 2011. On the evolution of the instance level of DL-Lite knowledge bases. In *Description Logics*.
- Liu, H.; Lutz, C.; Milicic, M.; and Wolter, F. 2011. Foundations of instance level updates in expressive Description Logics. *Artificial Intelligence* 175(18):2170–2197.
- Motik, B.; Shearer, R.; and Horrocks, I. 2009. Hyper-tableau Reasoning for Description Logics. *Journal of Artificial Intelligence Research* 36:165–228.
- Peñaloza, R., and Sertkaya, B. 2010. On the complexity of axiom pinpointing in the EL family of description logics. In *Proc. of KR*.
- Peppas, P. 2007. Belief revision. In F. van Harmelen, V. L., and Porter, B., eds., *Handbook of Knowledge Representation*. Elsevier.
- Qi, G., and Du, J. 2009. Model-based revision operators for terminologies in Description Logics. In *Proc. of IJCAI*, 891–897.
- Schlobach, S.; Huang, Z.; Cornet, R.; and van Harmelen, F. 2007. Debugging incoherent terminologies. *J. Automated Reasoning* 39(3):317–349.
- Suntisrivaraporn, B.; Qi, G.; Ji, Q.; and Haase, P. 2008. A modularization-based approach to finding all justifications for OWL DL entailments. In *Proc. of ASWC*, 1–15.
- Wang, Z.; Wang, K.; and Topor, R. W. 2010a. A new approach to knowledge base revision in DL-Lite. In *Proc. of AAAI*.
- Wang, Z.; Wang, K.; and Topor, R. W. 2010b. Revising general knowledge bases in Description Logics. In *Proc. of KR*.