(Non-)Succinctness of Uniform Interpolants of General Terminologies in the Description Logic $\mathcal{EL}$[☆]

Nadeschda Nikitina[a,*], Sebastian Rudolph[b]

[a] *University of Oxford, Department of Computer Science*
Wolfson Building, Parks Road, Oxford, OX1 3QD, UK
Tel.: +44 (1865) 610747, Fax: +44 (1865) 283531
[b] *Technische Universität Dresden, Fakultät Informatik*
Nöthnitzer Straße 46, Dresden, 01062, Germany
Tel.: +49 (351) 463 38516, Fax: +49 (351) 463 32827

Abstract

$\mathcal{EL}$ is a popular description logic, used as a core formalism in large existing knowledge bases. Uniform interpolants of knowledge bases are of high interest, e.g. in scenarios where a knowledge base is supposed to be partially reused. However, to the best of our knowledge no procedure has yet been proposed that computes uniform $\mathcal{EL}$ interpolants of general $\mathcal{EL}$ terminologies. Up to now, also the bound on the size of uniform $\mathcal{EL}$ interpolants has remained unknown. In this article, we propose an approach to computing a finite uniform interpolant for a general $\mathcal{EL}$ terminology if it exists. To this end, we develop a quadratic representation of $\mathcal{EL}$ TBoxes as regular tree grammars. Further, we show that, if a finite uniform $\mathcal{EL}$ interpolant exists, then there exists one that is at most triple exponential in the size of the original TBox, and that, in the worst case, no smaller interpolants exist, thereby establishing tight worst-case bounds on their size. Beyond showing these bounds, the notions and results established in this paper also provide useful insights for designing efficient ontology reformulation algorithms, for instance, within the context of module extraction.

Keywords: Ontologies, Knowledge Representation, Automated Reasoning, Description Logics, Uniform Interpolation, Forgetting, $\mathcal{EL}$

[☆] This is a revised and extended version of previous work [1].

[*] Corresponding Author

Email addresses: nadejda.nikitina@cs.ox.ac.uk (Nadeschda Nikitina), sebastian.rudolph@tu-dresden.de (Sebastian Rudolph)

1. Introduction

With the wide-spread adoption of ontological modeling by means of the W3C-specified OWL Web Ontology Language [2], description logics (DLs, [3, 4]) have developed into one of the most popular family of formalisms employed for knowledge representation and reasoning [5, 6, 7, 8]. For application scenarios where scalability of reasoning is of utmost importance, specific tractable sublanguages (the so-called *profiles* [9]) of OWL have been put into place, among them OWL 2 EL which in turn is based on DLs of the $\mathcal{EL}$ family [10, 11].

In view of the practical deployment of OWL and its profiles [12, 13, 14], non-standard reasoning services for supporting modeling activities gain in importance. An example of such reasoning services supporting knowledge engineers in different tasks is that of *uniform interpolation*: given a theory using a certain vocabulary, and a subset of "relevant terms" of that vocabulary, find a theory (referred to as a *uniform interpolant*, short: UI) that uses only the relevant terms and gives rise to the same consequences (expressible via relevant terms) as the original theory. Intuitively, this provides a view on the ontology where all irrelevant (asserted as well as implied) statements have been filtered out.

Uniform interpolation has many applications within ontology engineering. For instance, it can help ontology engineers understand existing ontological specifications by visualizing implicit dependencies between relevant concepts and roles, as used, for instance, for interactive ontology revision [15]. In particular for understanding and developing complex knowledge bases, e.g., those consisting of *general concept inclusions* (GCIs), appropriate tool support of this kind would be beneficial. Another application of uniform interpolation is ontology reuse: given an ontology that is to be reused in a different scenario, most likely not all aspects of this ontology are relevant to the new usage requirements. In combination with module extraction, uniform interpolation can be used to reduce the amount of irrelevant information within an ontology employed in a new context.

For DL-Lite, the problem of uniform interpolation has been investigated [16, 17] and a tight exponential bound on the size of uniform interpolants has been shown. Lutz and Wolter [18] propose an approach to uniform interpolation in expressive description logics such as $\mathcal{ALC}$ featuring general terminologies showing a tight triple-exponential bound on the size of uniform interpolants. Koopman and Schmidt [19] and Ludwig and Konev [20] propose practical approaches to computing uniform interpolants in expressive description logics. For the lightweight description logic $\mathcal{EL}$, the problem of uniform interpolation has, however, not been solved. To the best of our knowledge, the only existing approach [21] to uniform interpolation in $\mathcal{EL}$ is restricted to terminologies containing each concept symbol at most once on the left-hand side of concept inclusions and additionally satisfying particular acyclicity conditions which are sufficient, but not necessary for the existence of a uniform interpolant. Recently, Lutz, Seylan and Wolter [22] proposed an EXPTIME procedure for deciding, whether a finite uniform $\mathcal{EL}$ interpolant exists for a particular general terminology and a particular set of relevant terms. However, the authors do not address the actual computation of such a uniform interpolant. Up to now, also the bounds on the size of uniform $\mathcal{EL}$ interpolants have remained unknown.

In this paper, we propose a worst-case-optimal approach to computing a finite uniform $\mathcal{EL}$ interpolant for a general terminology. Our approach is based on proof theory and regular tree languages. We develop a grammar representation of $\mathcal{EL}$ TBoxes. These grammars are quadratic in the size of the initial TBox and capture all of its logical consequences except for a certain kind of *weak consequences* – consequences that can be trivially derived from other logical consequences but are not equivalent to those. We show via a proof-theoretic analysis that the tree languages generated by the proposed grammars indeed capture all non-weak consequences of the initial terminology expressed using the set of relevant terms.

Further, we show that certain finite subsets of the languages generated by these grammars can be transformed into a uniform $\mathcal{EL}$ interpolant of at most triple exponential size, if such a finite uniform $\mathcal{EL}$ interpolant exists for the given terminology and a set of terms. We also show that, in the worst-case, no shorter interpolants exist, thereby establishing tight bounds on the size of uniform interpolants in $\mathcal{EL}$.

It should be noted that the notions and results presented in this article go beyond the mere purpose of showing the triple exponential blowup and have practical applications. In fact, the proposed grammars have served as a basis for a module extraction tool in follow-up work by Nikitina and Glimm [23]. Within this tool, the insights gained in the present article are taken into account to derive a blowup-avoiding algorithm for a kind of partial uniform interpolation that conditionally eliminates concept symbols one by one after a careful analysis.

The article is structured as follows: In Section 2, we recall the necessary preliminaries on $\mathcal{EL}$. In Section 3, we introduce a calculus for deriving general subsumptions in $\mathcal{EL}$ terminologies, which is used as a major tool in the proofs of this work. Section 4 formally introduces the notion of inseparability and defines the task of uniform interpolation. Section 5 demonstrates that the smallest uniform interpolants in $\mathcal{EL}$ can be triple exponential in the size of the original knowledge base. In Section 6.1, we describe a normalisation of terminologies that enables a representation of non-weak logical consequences as languages of regular tree grammars. In Section 6.2, we recall the necessary preliminaries on regular tree languages/grammars and introduce regular tree grammars representing subsumees and subsumers of concept symbols, which are the basis for computing uniform $\mathcal{EL}$ interpolants as shown in Section 6.3. In the same section, we also show the upper bound on the size of uniform interpolants. After giving an overview of related work in Section 7, we summarize the contributions in Section 8 and discuss some ideas for future work. This is a revised and extended version of our previous paper [1] and contains technical enhancements, a more detailed argumentation, examples and the full proofs.

2. Preliminaries

In this section, we formally introduce the description logic $\mathcal{EL}$, and recall some of its well-known properties. Let N_C and N_R be countably infinite and mutually disjoint sets called *concept symbols* and *role symbols*, respectively. $\mathcal{EL}$ *concepts* C are defined by

$$C ::= A \mid C \sqcap C \mid \exists r.C$$

where A and r range over $N_C \cup \{\top\}$ and N_R, respectively. In the following, C, D, E, F and G can denote arbitrary concepts, while A, B can only denote concept symbols (i.e., concepts from N_C) or $\top$. We use the term *simple concept* to refer to a simpler form of $\mathcal{EL}$ concepts defined by $C_s ::= A \mid \exists r.A$, where A and r range over $N_C \cup \{\top\}$ and N_R, respectively.

A *terminology* or *TBox* consists of *concept inclusion* axioms $C \sqsubseteq D$ and *concept equivalence* axioms $C \equiv D$, the latter used as a shorthand for the mutual inclusion $C \sqsubseteq D$ and $D \sqsubseteq C$.[1] The *signature* of an $\mathcal{EL}$ concept C, an axiom α or a TBox $\mathcal{T}$, denoted by $\text{sig}(C)$, $\text{sig}(\alpha)$ or $\text{sig}(\mathcal{T})$, respectively, is the set of concept and role symbols occurring in it. To distinguish between the set of concept symbols and the set of role symbols, we use $\text{sig}_C(\cdot)$ and $\text{sig}_R(\cdot)$, respectively. Further, we use $\text{sub}(\mathcal{T})$ to denote the set of all subconcepts in $\mathcal{T}$.

For a concept C, let the *role depth* of C (denoted by $d(C)$) be the maximal nesting depth of existential restrictions within C. For instance, $d(\exists r.(\exists s.A \sqcap B) \sqcap$

[1] While knowledge bases in general can also include a specification of individuals with the corresponding concept and role assertions (ABox), in this paper we do not consider ABoxes, but concentrate on TBoxes.

$\exists s.B) = 2$. For a TBox $\mathcal{T}$, the role depth is given by the maximal role depth of its subconcepts.

Next, we recall the semantics of the DL constructs introduced above, which is defined by the means of interpretations. An *interpretation* $\mathcal{I}$ is given by a set $\Delta^{\mathcal{I}}$, called the *domain*, and an *interpretation function* $\cdot^{\mathcal{I}}$ assigning to each concept $A \in N_C$ a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ and to each role $r \in N_R$ a subset $r^{\mathcal{I}}$ of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The interpretation of $\top$ is fixed to $\Delta^{\mathcal{I}}$. The interpretation of arbitrary $\mathcal{EL}$ concepts is defined inductively via $(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$ and $(\exists r.C)^{\mathcal{I}} = \{x \mid (x, y) \in r^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}} \text{ for some } y\}$. An interpretation $\mathcal{I}$ *satisfies* an axiom $C \sqsubseteq D$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$, if it satisfies all axioms in $\mathcal{T}$. We say that $\mathcal{T}$ entails an axiom α (in symbols, $\mathcal{T} \models \alpha$), if α is satisfied by all models of $\mathcal{T}$. The *deductive closure* of a TBox $\mathcal{T}$ is the set of all axioms entailed by $\mathcal{T}$. For $\mathcal{EL}$ concepts C, D such that $\mathcal{T} \models C \sqsubseteq D$, we call C a *subsumee* of D and D a *subsumer* of C.

2.1. Model-Theoretic Properties of $\mathcal{EL}$ Concepts

In the following, we provide some results concerning model-theoretic properties of $\mathcal{EL}$ concepts, which are essentially common knowledge. Nevertheless, to make the paper self-contained, we include the proofs in the appendix. We first define pointed interpretations as well as homomorphisms between them. Moreover we define the notion of a characteristic interpretation of an $\mathcal{EL}$ concept. Intuitively, a concept's characteristic interpretation describes a partial model with one distinguished element which represents necessary and sufficient conditions for a domain element to be an instance of this concept.

Definition 1. *A* pointed interpretation *is a pair $(\mathcal{I}, x)$ with $x \in \Delta^{\mathcal{I}}$. Given two pointed interpretations $(\mathcal{I}_1, x_1)$ and $(\mathcal{I}_2, x_2)$, a* homomorphism *from $(\mathcal{I}_1, x_1)$ to $(\mathcal{I}_2, x_2)$ is a mapping $\varphi : \Delta^{\mathcal{I}_1} \to \Delta^{\mathcal{I}_2}$ such that*

- $\varphi(x_1) = x_2$,
- $x \in A^{\mathcal{I}_1}$ *implies* $\varphi(x) \in A^{\mathcal{I}_2}$ *for all* $A \in N_C$,
- $(x, y) \in r^{\mathcal{I}_1}$ *implies* $(\varphi(x), \varphi(y)) \in r^{\mathcal{I}_2}$ *for all* $r \in N_R$.

Given an $\mathcal{EL}$ concept C, we define its characteristic pointed interpretation *$(\mathcal{I}_C, x_C)$ inductively over the structure of C as follows:*

- *For $\top$ we let $\Delta^{\mathcal{I}_\top} = \{x_\top\}$ with*

- $B^{\mathcal{I}_\top} = \emptyset$ *for all* $B \in N_C$, *and*
- $r^{\mathcal{I}_\top} = \emptyset$ *for all* $r \in N_R$.

- *For* $A \in N_C$ *we let* $\Delta^{\mathcal{I}_A} = \{x_A\}$ *with*
  - $A^{\mathcal{I}_A} = \{x_A\}$,
  - $B^{\mathcal{I}_A} = \emptyset$ *for all* $B \in N_C \setminus \{A\}$, *and*
  - $r^{\mathcal{I}_A} = \emptyset$ *for all* $r \in N_R$.

- *For* $C = C_1 \sqcap C_2$, *we define* $\Delta^{\mathcal{I}_C} = \{x_C\} \cup \bigcup_{\iota \in \{1,2\}} (\Delta^{\mathcal{I}_{C_\iota}} \setminus \{x_{C_\iota}\}) \times \{\iota\}$ *with*
  - $A^{\mathcal{I}_C} = \{x_C \mid x_{C_1} \in A^{\mathcal{I}_{C_1}}$ *or* $x_{C_2} \in A^{\mathcal{I}_{C_2}}\} \cup \bigcup_{\iota \in \{1,2\}} (A^{\mathcal{I}_{C_\iota}} \setminus \{x_{C_\iota}\}) \times \{\iota\}$ *for all* $A \in N_C$, *and*
  - $r^{\mathcal{I}_C} = \bigcup_{\iota \in \{1,2\}} \{(x_C, (y, \iota)) \mid (x_{C_\iota}, y) \in r^{\mathcal{I}_{C_\iota}}\} \cup \bigcup_{\iota \in \{1,2\}} \{((y, \iota), (y', \iota)) \mid (y, y') \in r^{\mathcal{I}_{C_\iota}}, y \neq x_{C_\iota}\}$ *for all* $r \in N_R$.

- *For* $C = \exists r.C'$, *we define* $\Delta^{\mathcal{I}_C} = \{x_C\} \cup \Delta^{\mathcal{I}_{C'}}$ *with*
  - $A^{\mathcal{I}_C} = A^{\mathcal{I}_{C'}}$ *for all* $A \in N_C$, *and*
  - $(r')^{\mathcal{I}_C} = \{(x_C, x_{C'}) \mid r' = r\} \cup (r')^{\mathcal{I}_{C'}}$ *for all* $r' \in N_R$.

The subsequent lemma shows that characteristic interpretations indeed characterize $\mathcal{EL}$ concept membership via the existence of appropriate homomorphisms.

Lemma 1 (structurality of validity of $\mathcal{EL}$ concepts)**.** *For any $\mathcal{EL}$ concept C and any interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ and $x \in \Delta^\mathcal{I}$ it holds that $x \in C^\mathcal{I}$ if and only if there is a homomorphism from $(\mathcal{I}_C, x_C)$ to $(\mathcal{I}, x)$.*

The next lemma shows that $\mathcal{EL}$ concept subsumption in the absence of terminological background knowledge can as well be characterized via homomorphisms between characteristic interpretations.

Lemma 2 (Structurality of $\mathcal{EL}$ concept subsumption)**.** *Let C and C' be two $\mathcal{EL}$ concepts. Then $\emptyset \models C \sqsubseteq C'$ if and only if there is a homomorphism from $(\mathcal{I}'_C, x'_C)$ to $(\mathcal{I}_C, x_C)$.*

The proofs of both lemmas can be found in Appendix A.

3. A Gentzen-Style Proof System for $\mathcal{EL}$

The aim of this section is to provide a proof-theoretic calculus that is sound and complete for general subsumption in $\mathcal{EL}$. We will use this calculus in the subsequent sections to prove particular properties of TBoxes of a certain form in the context of consequence-preserving rewriting. The Gentzen-style calculus for $\mathcal{EL}$ is shown in Fig. 1 and is a variation of the calculus given by Hofmann [24].

$$\frac{}{C \sqsubseteq C}(\textsc{Ax}) \quad \frac{}{C \sqsubseteq \top}(\textsc{AxTop})$$

$$\frac{D \sqsubseteq E}{C \sqcap D \sqsubseteq E}(\textsc{AndL})$$

$$\frac{C \sqsubseteq E \quad C \sqsubseteq D}{C \sqsubseteq D \sqcap E}(\textsc{AndR})$$

$$\frac{C \sqsubseteq D}{\exists r.C \sqsubseteq \exists r.D}(\textsc{Ex})$$

$$\frac{C \sqsubseteq E \quad E \sqsubseteq D}{C \sqsubseteq D}(\textsc{Cut})$$

Figure 1: Gentzen-style proof system for general $\mathcal{EL}$ terminologies with C, D, E arbitrary concepts.

The calculus operates on sequents. A *sequent* is of the form $C \sqsubseteq D$, where C, D are $\mathcal{EL}$ concepts. The rules depicted in Fig. 1 can be used to derive new sequents from sequents that have already been derived. For instance, if we have derived the sequent $C \sqsubseteq D$, we can derive the sequent $\exists r.C \sqsubseteq \exists r.D$ using rule (Ex). A *derivation* (or *proof*) of a sequent $C \sqsubseteq D$ is a finite tree with whose nodes are labeled with sequents. The tree root is labeled with the sequent $C \sqsubseteq D$. Within the tree, a parent node is always labeled by the conclusion of a proof rule from Fig. 1 whose antecedent(s) are the labels of the child nodes. The leaves of a derivation are either labeled by axioms from $\mathcal{T}$ or conclusions of (Ax) or (AxTop). We use the notation $\mathcal{T} \vdash C \sqsubseteq D$ to indicate that there is a derivation of $C \sqsubseteq D$. In our calculus, we assume commutativity of conjunction for convenience.[2] Fig. 2 shows an example derivation of the sequent $\exists r.C_1 \sqsubseteq C_2$ in our

[2]Alternatively, commutativity of conjunction can be realised by adding a rule $\frac{}{C \sqcap D \sqsubseteq D \sqcap C}$.

$$\dfrac{\exists r.C_1 \sqsubseteq C_1 \sqcap C_2 \qquad \dfrac{\dfrac{}{C_2 \sqsubseteq C_2}\,(\textsc{Ax})}{C_1 \sqcap C_2 \sqsubseteq C_2}\,(\textsc{AndL})}{\exists r.C_1 \sqsubseteq C_2}\,(\textsc{Cut})$$

Figure 2: Example derivation of $\exists r.C_1 \sqsubseteq C_2$ from $\mathcal{T}_e$.

calculus w.r.t. the $\mathcal{EL}$ TBox $\mathcal{T}_e = \{\exists r.C_1 \sqsubseteq C_1 \sqcap C_2\}$.

We show that the above calculus is sound and complete for subsumptions between arbitrary $\mathcal{EL}$ concepts.

Lemma 3 (Soundness and Completeness)**.** *Let $\mathcal{T}$ be an arbitrary $\mathcal{EL}$ TBox, C, D $\mathcal{EL}$ concepts. Then $\mathcal{T} \models C \sqsubseteq D$ iff $\mathcal{T} \vdash C \sqsubseteq D$.*

Proof. While the soundness of the proof system (if-direction) can be easily verified for each rule separately, the proof of completeness is more sophisticated. Analogously to other proof-theoretic approaches [11, 25], we show the only-if-direction of the lemma by constructing a model $\mathcal{I}$ for $\mathcal{T}$ wherein *only* the GCIs derivable from $\mathcal{T}$ are valid. The construction of the model is rather standard (a similar construction is, e.g., given by Lutz and Wolter [26]). The model is defined as follows:

- $\Delta^{\mathcal{I}}$ is the set of elements δ_C where C is an $\mathcal{EL}$ concept;
- $A^{\mathcal{I}} := \{\delta_C \in \Delta^{\mathcal{I}} \mid \mathcal{T} \vdash C \sqsubseteq A\}$, where $A \in N_C$;
- $r^{\mathcal{I}} := \{(\delta_C, \delta_D) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid \mathcal{T} \vdash C \sqsubseteq \exists r.D\}$ where $r \in N_R$.

We will show that the following claim holds for $\mathcal{I}$:
For all $\delta_E \in \Delta^{\mathcal{I}}$ and $\mathcal{EL}$ concepts F, it holds that $\delta_E \in F^{\mathcal{I}}$ iff $\mathcal{T} \vdash E \sqsubseteq F$. *(*)*

This claim can be exploited in two ways: First, we use it to show that $\mathcal{I}$ is indeed a model of $\mathcal{T}$. Let $C \sqsubseteq D \in \mathcal{T}$ and consider an arbitrary concept G with $\delta_G \in C^{\mathcal{I}}$. Via (*) we obtain $\mathcal{T} \vdash G \sqsubseteq C$. Further, $\mathcal{T} \vdash C \sqsubseteq D$ is due to $C \sqsubseteq D \in \mathcal{T}$. Thus we can derive $\mathcal{T} \vdash G \sqsubseteq D$ via (Cut) and consequently, applying (*) again, we obtain $\delta_G \in D^{\mathcal{I}}$. Thereby, we have proved that $\mathcal{I} \models \mathcal{T}$.

Second, we use (*) to show that $\mathcal{I}$ is a counter-model for all GCIs not derivable from $\mathcal{T}$ as follows: Assume $\mathcal{T} \not\vdash C \sqsubseteq D$. From $\mathcal{T} \vdash C \sqsubseteq C$ and (*) we derive $\delta_C \in C^{\mathcal{I}}$. From $\mathcal{T} \not\vdash C \sqsubseteq D$ and (*) we obtain $\delta_C \notin D^{\mathcal{I}}$. Hence we get $C^{\mathcal{I}} \not\subseteq D^{\mathcal{I}}$ and therefore $\mathcal{I} \not\models C \sqsubseteq D$.

It remains to prove (*). This is done by induction over the structure of the concept F. There are two base cases:

- for $F = \top$, the claim trivially follows from (AXTOP),
- for a concept symbol F, it is a direct consequence of the definition of our model $\mathcal{I}$.

We now consider the cases where F is a complex concept

- for $F = C_1 \sqcap \ldots \sqcap C_n$, we note that $\delta_E \in F^{\mathcal{I}}$ exactly if $\delta_E \in C_i^{\mathcal{I}}$ for all $i \in \{1 \ldots n\}$. By induction hypothesis, this means $\mathcal{T} \vdash E \sqsubseteq C_i$ for all $i \in \{1 \ldots n\}$. Finally, observe that $\{E \sqsubseteq C_i \mid 1 \leq i \leq n\}$ and $E \sqsubseteq C_1 \sqcap \ldots \sqcap C_n$ can be mutually derived from each other:
  - $\{E \sqsubseteq C_i \mid 1 \leq i \leq n\} \vdash E \sqsubseteq C_1 \sqcap \ldots \sqcap C_n$ is a straightforward consequence of (ANDR);
  - To derive $E \sqsubseteq C_1 \sqcap \ldots \sqcap C_n \vdash \{E \sqsubseteq C_i \mid 1 \leq i \leq n\}$, we first derive $C_1 \sqcap \ldots \sqcap C_n \sqsubseteq C_i$ from $C_i \sqsubseteq C_i$ (obtained using (AX)) by applying (ANDL) multiple times. Since $\mathcal{T} \vdash E \sqsubseteq C_1 \sqcap \ldots \sqcap C_n$, we can apply (CUT) (with $E \sqsubseteq C_1 \sqcap \ldots \sqcap C_n$ as the left antecedent and $C_1 \sqcap \ldots \sqcap C_n \sqsubseteq C_i$ as the right antecedent) to derive $E \sqsubseteq C_i$.
- for $F = \exists r.G$, we prove the two directions separately. First assuming $\delta_E \in F^{\mathcal{I}}$ we must find $(\delta_E, \delta_H) \in r^{\mathcal{I}}$ for some H with $\delta_H \in G^{\mathcal{I}}$. This implies both $\mathcal{T} \vdash E \sqsubseteq \exists r.H$ (by the definition of the model) and $\mathcal{T} \vdash H \sqsubseteq G$ (via the induction hypothesis). From the latter, we can deduce $\mathcal{T} \vdash \exists r.H \sqsubseteq \exists r.G$ by (EX) and consequently $\mathcal{T} \vdash E \sqsubseteq \exists r.G$. For the other direction, note that by definition, $\mathcal{T} \vdash E \sqsubseteq \exists r.G$ implies $(\delta_E, \delta_G) \in r^{\mathcal{I}}$. On the other hand, we get $\mathcal{T} \vdash G \sqsubseteq G$ by (AX) and therefore $\delta_G \in G^{\mathcal{I}}$ by the induction hypothesis which yields us $\delta_E \in F^{\mathcal{I}}$. □

Alternatively, the completeness of the calculus could be shown by a reduction to the calculus of Hofmann [24].

4. Uniform Interpolation

Uniform interpolation has many potential applications in ontology engineering due to its ability to reduce the amount of irrelevant information within a terminology while preserving all relevant consequences given the set of relevant signature elements. The task of computing terminologies with such properties is not trivial. For instance, it is not sufficient to simply eliminate axioms containing only irrelevant entities, since it can change the meaning of the relevant entities and cause a loss of relevant information. Example 1 demonstrates the effect of such an elimination.

Example 1. *Consider the terminology $\mathcal{T}$ given by*

$$
\begin{aligned}
A_{i+1} &\sqsubseteq A_i && 1 \leq i \leq 3 \\
A_4 &\sqsubseteq \exists r.A_4
\end{aligned}
$$

If we are only interested in entities A_1, A_4, r, then we might consider to eliminate all axioms except for those that contain at least one relevant entity, obtaining $\mathcal{T}' = \mathcal{T} \setminus \{A_3 \sqsubseteq A_2\}$. However, in this way we would lose the information about the connection between the relevant entities, for instance $A_4 \sqsubseteq A_1, A_4 \sqsubseteq \exists r.A_1, A_4 \sqsubseteq \exists r.\exists r.A_1, \ldots$. Indeed, $\mathcal{T}'$ does not entail any of these statements.

In typical ontology reuse scenarios, it is required to preserve the meaning of the relevant entities while computing a terminology that contains as little irrelevant information as possible. We say that the meaning of relevant entities is preserved, if every logical statement that follows from the original terminology and contains only relevant entities also follows from the resulting terminology. The logical foundation for such a preservation of relevant consequences can be defined using the notion of *inseparability*. Two terminologies, $\mathcal{T}_1$ and $\mathcal{T}_2$, are inseparable w.r.t. a signature Σ if they have the same Σ-consequences, i.e., consequences whose signatures are subsets of Σ. Depending on the particular application requirements, the expressivity of those Σ-consequences can vary from subsumption axioms and concept assertions to conjunctive queries. In the following, we consider *concept-inseparability* of general $\mathcal{EL}$ terminologies as given, for instance, in [17, 21, 18]:

Definition 2. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two general $\mathcal{EL}$ terminologies and Σ a signature. $\mathcal{T}_1$ and $\mathcal{T}_2$ are* concept-inseparable *w.r.t. Σ, in symbols $\mathcal{T}_1 \equiv_{\Sigma}^{\mathcal{EL}} \mathcal{T}_2$, if for all $\mathcal{EL}$ concepts C, D with $sig(C) \cup sig(D) \subseteq \Sigma$ it holds that $\mathcal{T}_1 \models C \sqsubseteq D$ iff $\mathcal{T}_2 \models C \sqsubseteq D$.*

Due to its usefulness for different ontology engineering tasks, concept-inseparability has been investigated by different authors in the last decade. For instance, in the context of ontology reuse, the notion of inseparability can be used to derive a terminology that is inseparable from the initial terminology and is using only terms from Σ. This is an established non-standard reasoning task called forgetting or uniform interpolation.

Definition 3. *Given a signature* Σ *and a terminology* $\mathcal{T}$*, the task of* uniform interpolation *is to determine a terminology* $\mathcal{T}'$ *with sig*$(\mathcal{T}') \subseteq \Sigma$ *such that* $\mathcal{T} \equiv_{\Sigma}^{\mathcal{EL}} \mathcal{T}'$. $\mathcal{T}'$ *is also called a* uniform Σ-interpolant *of* $\mathcal{T}$.

For the TBox $\mathcal{T}$ in Example 1, one possible uniform Σ-interpolant for $\Sigma = \{A_1, A_4, r\}$ would be $\mathcal{T}_\Sigma = \{A_4 \sqsubseteq A_1, A_4 \sqsubseteq \exists r.A_4\}$. We see that, by introducing a shortcut axiom $A_4 \sqsubseteq A_1$, we preserve all relevant logical consequences (those expressed using Σ) while eliminating all other logical consequences, e.g., $A_{i+1} \sqsubseteq A_i$ for $0 \leq i \leq 3$.

In practice, uniform interpolants are required to be finite, i.e., expressible by a finite set of finite axioms using only the language constructs of a particular DL. It is well-known (e.g., see [21]) that, in the presence of cyclic concept inclusions, a finite uniform $\mathcal{EL}$ Σ-interpolant might not exist for a particular terminology $\mathcal{T}$ and a particular Σ.

Example 2. *Consider the terminology* $\mathcal{T} = \{A' \sqsubseteq A, A \sqsubseteq A'', A \sqsubseteq \exists r.A, \exists s.A \sqsubseteq A\}$ *and let* $\Sigma = \{s, r, A', A''\}$. *As consequences, we obtain infinite sequences* $A' \sqsubseteq \exists r.\exists r.\exists r....A''$ *and* $\exists s.\exists s.\exists s....A' \sqsubseteq A''$ *which contain nested existential quantifiers of unbounded depth. Those sequences cannot be finitely axiomatized, using only signature elements from* Σ.

Lutz, Seylan and Wolter [22] give an EXPTIME procedure for deciding if a finite uniform $\mathcal{EL}$ interpolant exists. In the following, we extend the results and show that, if a finite uniform $\mathcal{EL}$ interpolant exists for the given terminology and signature, then there exists a uniform $\mathcal{EL}$ interpolant of at most triple exponential size. Further, we show that, in the worst case, no shorter interpolants exist, thereby establishing tight bounds on the size of uniform interpolants in $\mathcal{EL}$.

5. Lower Bound

In this section we will establish the lower bound for the size of uniform interpolants of $\mathcal{EL}$ terminologies, in case they exist. It is interesting that, while deciding the existence of uniform interpolants in $\mathcal{EL}$ [22] is one exponential less complex than the same decision problem for the more complex logic $\mathcal{ALC}$ [18], the size of uniform interpolants remains triple-exponential. An intuitive reason for this rather unexpected result can be seen in the unavailability of disjunction, which would allow for a more succinct representation of the interpolants. In fact, the exponential blowup due to the non-availability of disjunction has been noted before [21]. We show the triple-exponential lower bound by means of a sequence of terminologies (obtained by a slight modification of the corresponding example given in [27] originally demonstrating a double exponential lower bound in the context of conservative extensions).

We start with an intuitive explanation of what the terminology is supposed to express. Assume, given some $n \in \mathbb{N}$ we want to label domain elements with natural numbers $0 \ldots 2^n - 1$ according to the following scheme: domain elements belonging to the concepts A_1 or A_2 are labeled with 0. Further, whenever we find a domain element δ that is linked via an r-role to an ℓ-labeled domain element δ_1 and linked via an s-role to an ℓ-labeled domain element δ_2, then δ will be labeled with $\ell + 1$ (provided $\ell < 2^n - 1$). Finally, we stipulate that every domain element labeled with $2^n - 1$ will belong to the concept B. In order to encode this labeling scheme in a knowledge base whose size is polynomial in n, we encode the number-labels in a binary way as a conjunction of n concepts. Thereby, the concept symbols $X_i, \overline{X_i}$ represent the i^{th} bit of ℓ's binary representation being clear or set.

Definition 4. *The $\mathcal{EL}$ TBox $\mathcal{T}_n$ for a natural number n is given by*

$$A_1 \sqsubseteq \overline{X_0} \sqcap \ldots \sqcap \overline{X_{n-1}} \quad (1)$$

$$A_2 \sqsubseteq \overline{X_0} \sqcap \ldots \sqcap \overline{X_{n-1}} \quad (2)$$

$$\bigsqcap_{\sigma \in \{r,s\}} \exists \sigma.(\overline{X_i} \sqcap X_0 \sqcap \ldots \sqcap X_{i-1}) \sqsubseteq X_i \qquad i < n \quad (3)$$

$$\bigsqcap_{\sigma \in \{r,s\}} \exists \sigma.(X_i \sqcap X_0 \sqcap \ldots \sqcap X_{i-1}) \sqsubseteq \overline{X_i} \qquad i < n \quad (4)$$

$$\bigsqcap_{\sigma \in \{r,s\}} \exists \sigma.(\overline{X_i} \sqcap \overline{X_j}) \sqsubseteq \overline{X_i} \;\; j < i < n \quad (5)$$

$$\bigsqcap_{\sigma \in \{r,s\}} \exists \sigma.(X_i \sqcap \overline{X_j}) \sqsubseteq X_i \;\; j < i < n \quad (6)$$

$$X_0 \sqcap \ldots \sqcap X_{n-1} \sqsubseteq B \quad (7)$$

In the above TBox, Axiom (3) ensures that a clear bit will be set in the successor number, if all lower bits are already set. The subsequent Axiom (4) ensures that a set bit will be clear in the successor number, if all lower bits are also set. Axioms (5) and (6) ensure that in all other cases, bits are not toggled. For instance, Axiom (5) states that, if any of the bits lower than i is clear, then bit i will remain clear also in the successor number.

If we now consider sets $\mathcal{C}_i$ of concept descriptions inductively defined by $\mathcal{C}_0 = \{A_1, A_2\}$, $\mathcal{C}_{i+1} = \{\exists r.C_1 \sqcap \exists s.C_2 \mid C_1, C_2 \in \mathcal{C}_i\}$, then we find that $|\mathcal{C}_{i+1}| = |\mathcal{C}_i|^2$ and consequently $|\mathcal{C}_i| = 2^{(2^i)}$. Thus, the set $\mathcal{C}_{2^n-1}$ contains triply exponentially many different concepts, each of which is doubly exponential in the size of $\mathcal{T}_n$ (intuitively, we obtain concepts having the shape of binary trees of exponential depth, thus having doubly exponentially many leaves, each of which can be A_1 or A_2, which gives rise to a triply exponential number of such trees). Then we will show that for each concept $C \in \mathcal{C}_{2^n-1}$ it holds that $\mathcal{T}_n \models C \sqsubseteq B$ and that there cannot be a smaller uniform interpolant with respect to the signature $\Sigma = \{A_1, A_2, B, r, s\}$ than the one containing all these GCIs.

Based on the above definition, we now prove the following result.

Theorem 1. *There exists a sequence of $\mathcal{EL}$ TBoxes and a fixed signature Σ such that for each TBox $(\mathcal{T}_n)$ within this sequence the following hold:*

- *the size of $\mathcal{T}_n$ is polynomial in n and*
- *the size of the smallest uniform interpolant of $\mathcal{T}_n$ with respect to Σ is at least $2^{(2^{(2^n-1)})}$.*

Proof. Obviously, the size of $\mathcal{T}_n$ is polynomial in n. As discussed above, the set $\mathcal{C}_{2^n-1}$ contains triply exponentially many different concepts, each of which is doubly exponential in the size of $\mathcal{T}_n$. By definition, for any k, every concept from $\mathcal{C}_k$ contains only signature elements from A_1, A_2, r, s.

It is rather straightforward to check that $\mathcal{T}_n \models C \sqsubseteq B$ holds for each concept $C \in \mathcal{C}_{2^n-1}$: by induction on k, we can show that for any $C \in \mathcal{C}_k$ with $k < 2^n$ it holds that $\mathcal{T}_n \models C \sqsubseteq Y_0^k \sqcap \ldots \sqcap Y_{n-1}^k$ with

$$Y_i^k = \begin{cases} X_i \text{ if } \lfloor \frac{k}{2^i} \rfloor \mathsf{mod}\ 2 = 1 \\ \overline{X_i} \text{ if } \lfloor \frac{k}{2^i} \rfloor \mathsf{mod}\ 2 = 0 \end{cases},$$

i.e., Y_i^k indicates the ith bit of the number k in binary encoding. Then, $C \sqsubseteq B$ follows via the last axiom of $\mathcal{T}_n$.

Toward the claimed triple-exponential lower bound, we now show that every uniform interpolant of $\mathcal{T}_n$ for $\Sigma = \{A_1, A_2, B, r, s\}$ must contain for each $C \in \mathcal{C}_{2^n-1}$ a GCI of the form $C \sqsubseteq B'$ with $B' = B$ or $B' = B \sqcap F$ for some F (where we consider structural variants – i.e., concepts whose characteristic interpretations are isomorphic – as syntactically equal). Toward a contradiction, we assume that this is not the case, i.e., there is a uniform interpolant $\mathcal{T}'$ and a $C \in \mathcal{C}_{2^n-1}$ where $C \sqsubseteq B' \notin \mathcal{T}'$ for any B' containing B as a (top-level) conjunct.

Yet, $C \sqsubseteq B$ must be a consequence of $\mathcal{T}'$, since it is a consequence of of $\mathcal{T}_n$ containing only signature elements from Σ and $\mathcal{T}'$ is a uniform interpolant of $\mathcal{T}_n$ w.r.t. Σ by assumption. Therefore, there must be a derivation of it. Looking at the derivation calculus from the last section, the last derivation step must be (ANDL) or (CUT). We can exclude (ANDL) since neither $\exists r.C' \sqsubseteq B$ nor $\exists s.C' \sqsubseteq B$ is the consequence of $\mathcal{T}'$ for any $C' \in \mathcal{C}_{2^n-2}$ (which can be easily shown by providing appropriate witness models of $\mathcal{T}'$). Consequently, the last derivation step must be an application of (CUT), i.e., there must be a concept $E \neq C$ such that $\mathcal{T}' \models C \sqsubseteq E$ and $\mathcal{T}' \models E \sqsubseteq B$. Without loss of generality, we assume that we consider a derivation tree where the subtree deriving $C \sqsubseteq E$ has minimal depth.

We now distinguish two cases: either E contains B as a conjunct or not.

- First we assume $E = E' \sqcap B$, i.e. the (CUT) rule was used to derive $C \sqsubseteq B$ from $C \sqsubseteq E' \sqcap B$ and $E' \sqcap B \sqsubseteq B$. The former cannot be contained in $\mathcal{T}'$ by assumption, hence it must have been derived itself. We can exclude (ANDR) due to the minimality of the proof. Again, it cannot have been derived via (ANDL) for the same reasons as given above, which again leaves (CUT) as the only possible derivation rule for obtaining $C \sqsubseteq E' \sqcap B$. Thus, there must be some concept G with $\mathcal{T}' \models C \sqsubseteq G$ and $\mathcal{T}' \models G \sqsubseteq E' \sqcap B$. Once more, we distinguish two cases: either G contains B as a conjunct or not.

  – If G contains B as a conjunct, i.e., $G = G' \sqcap B$, the derivation of $C \sqsubseteq E$ was not depth-minimal since there is a better proof where $C \sqsubseteq B$ is derived from $C \sqsubseteq G' \sqcap B$ and $G' \sqcap B \sqsubseteq B$ via (CUT). Hence we have a contradiction.

  – If G does not contain B as a conjunct, the original derivation of $C \sqsubseteq E$ was not depth-minimal since we can construct a better one that derives $C \sqsubseteq B$ directly from $C \sqsubseteq G$ and $G \sqsubseteq B$ (the latter being derived from $G \sqsubseteq E' \sqcap B$ via (ANDR)).

- Now assume E does not contain B as a conjunct.

  We construct a specific interpretation $(\Delta^\mathcal{I}, \cdot^\mathcal{I})$ as follows (ϵ denotes the empty word):

  – $\Delta^\mathcal{I} = \{w \mid w \in \{r, s\}^*,\ \text{length}(w) < 2^n\}$

– We define an auxiliary function χ associating a concept to each domain element: we let $\chi(\epsilon) = C$ (with ϵ being the empty word) and, for every $wr, ws \in \Delta^{\mathcal{I}}$ with $\chi(w) = \exists r.C_1 \sqcap \exists s.C_2$, we let $\chi(wr) = C_1$ and $\chi(ws) = C_2$.
– the concepts and roles are interpreted as follows:
  * $A_\iota^{\mathcal{I}} = \{w \mid \chi(w) = A_\iota\}$ for $\iota \in \{1, 2\}$
  * $B^{\mathcal{I}} = \{\epsilon\}$
  * $X_i^{\mathcal{I}} = \{w \mid \lfloor \frac{\text{length}(w)}{2^i} \rfloor \mathsf{mod}\ 2 = 0\}$ for $i < n$
  * $\overline{X_i}^{\mathcal{I}} = \{w \mid \lfloor \frac{\text{length}(w)}{2^i} \rfloor \mathsf{mod}\ 2 = 1\}$ for $i < n$
  * $r^{\mathcal{I}} = \{\langle w, wr\rangle \mid wr \in \Delta^{\mathcal{I}}\}$
  * $s^{\mathcal{I}} = \{\langle w, ws\rangle \mid ws \in \Delta^{\mathcal{I}}\}$

It is straightforward to check that $\mathcal{I}$ is a model of $\mathcal{T}_n$. Furthermore using descending induction on the length of w, we can show that $w \in (\chi(w))^{\mathcal{I}}$ for every $w \in \Delta^{\mathcal{I}}$; in particular, $\epsilon \in C^{\mathcal{I}}$. Consequently, due to our assumption, $\epsilon \in E^{\mathcal{I}}$ must hold. Now we observe that the restriction of $\mathcal{I}$ to the signature elements A_1, A_2, r, s is isomorphic to $\mathcal{I}_C$ (with x_C corresponding to ϵ). On the other hand, as $\epsilon \in E^{\mathcal{I}}$ we find by Lemma 1 a homomorphism from $(\mathcal{I}_E, x_E)$ to $(\mathcal{I}, \epsilon)$ and hence to $(\mathcal{I}_C, x_C)$, thus, by Lemma 2, E is a proper "structural superconcept" of C, i.e., $\emptyset \models C \sqsubseteq E$ and $\emptyset \not\models E \sqsubseteq C$ must hold.

We now obtain $\widetilde{E}$ by enriching E as follows: starting from $k = 0$ and iteratively incrementing k up to $2^n - 1$, every subconcept G of E satisfying $\emptyset \models G \sqsubseteq C'$ for some $C' \in \mathcal{C}_k$ is substituted by $G \sqcap Y_0^k \sqcap \ldots \sqcap Y_{n-1}^k$ where, as before,

$$Y_i^k = \begin{cases} X_i \text{ if } \lfloor \frac{k}{2^i} \rfloor \mathsf{mod}\ 2 = 1 \\ \overline{X_i} \text{ if } \lfloor \frac{k}{2^i} \rfloor \mathsf{mod}\ 2 = 0 \end{cases},$$

i.e., Y_i^k indicates the ith bit of the number k in binary encoding.

Then, $\widetilde{E}$'s characteristic pointed interpretation $(\mathcal{I}_{\widetilde{E}}, x_{\widetilde{E}})$ satisfies the following conditions: $\mathcal{I}_{\widetilde{E}}$ is a model of $\mathcal{T}_n$ (following from structural induction on subconcepts of $\widetilde{E}$) and its root individual $x_{\widetilde{E}}$ is in the extension of $\widetilde{E}$. Still, we find $x_{\widetilde{E}} \notin C^{\mathcal{I}_{\widetilde{E}}}$ for the following reason: C does only contain signature elements from $\{A_1, A_2, B, r, s\}$, and the restriction of $(\mathcal{I}_{\widetilde{E}}, x_{\widetilde{E}})$ to these signature elements is isomorphic to $(\mathcal{I}_E, x_E)$, therefore $x_{\widetilde{E}} \in C_{\mathcal{I}_{\widetilde{E}}}$ iff $x_E \in C^{\mathcal{I}_E}$. The latter is however not the case as this would imply by

Lemma 1 that there is a homomorphism from $(\mathcal{I}_C, x_C)$ to $(\mathcal{I}_E, x_E)$ and consequently, via Lemma 2 $\emptyset \models E \sqsubseteq C$, contradicting our finding above.

Yet, the root individual $x_{\widetilde{E}}$ cannot satisfy any other concept C'' from $\mathcal{C}_{2^n-1} \setminus \{C\}$ either, since this, via $\emptyset \models E \sqsubseteq C''$, would imply $\emptyset \models C \sqsubseteq C''$ which is not the case (by induction on k one can show that there cannot be a homomorphism between the characteristic pointed interpretations of any two distinct concepts from any $\mathcal{C}_k$). In particular, we note that $x_{\widetilde{E}} \notin B^{\mathcal{I}_{\widetilde{E}}}$. Thus, we have found a model of $\mathcal{T}_n$ witnessing $\mathcal{T}_n \not\models E \sqsubseteq B$, contradicting our assumption that $\mathcal{T}' \models E \sqsubseteq B$.

□

Hence we have found a class $\mathcal{T}_n$ of TBoxes giving rise to uniform $\mathcal{EL}$ interpolants of triple-exponential size in terms of the original TBox.

6. Upper Bound

Now we discuss the upper bound on the size of uniform $\mathcal{EL}$ interpolants as well as their computation. Since, for a TBox $\mathcal{T}$ and a signature Σ, there are in general infinitely many Σ-consequences, in the following, we aim at identifying a subset of such consequences, the deductive closure of which contains the whole set. Interestingly, there exists a bound on the role depth of Σ-consequences such that, for the set $\mathcal{T}_{\Sigma,N}$ of all Σ-consequences of $\mathcal{T}$ with the maximal role depth N the following holds: either $\mathcal{T}_{\Sigma,N}$ is a uniform $\mathcal{EL}$ interpolant of $\mathcal{T}$ with respect to Σ or such a finite uniform $\mathcal{EL}$ interpolant of $\mathcal{T}$ does not exist. This is an easy consequence of results obtained by Lutz, Seylan and Wolter [22] while investigating the problem of existence of uniform $\mathcal{EL}$ interpolants (a proof can be found in Appendix B).

Lemma 4 (Reformulation of Lemma 55 from [22])**.** *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, Σ a signature. The following statements are equivalent:*

1. *There exists a uniform $\mathcal{EL}$ Σ-interpolant of $\mathcal{T}$.*
2. *There exists a uniform $\mathcal{EL}$ Σ-interpolant $\mathcal{T}'$ of $\mathcal{T}$ such that $d(\mathcal{T}') \leq 2^{4\cdot|sub(\mathcal{T})|} + 1$.*

However, an upper bound on the role depth is only sufficient for showing a non-elementary upper bound on the size of uniform interpolants for the following reasons. There are 2^n many different conjunctions of n different conjuncts, and, accordingly, for each role, 2^m many different existential restrictions of depth $i+1$

if m is the number of existential restrictions of depth i. Moreover, for any role depth i, we can find a TBox such that i is the corresponding maximal role depth. Subsequently, the upper bound on the role depth does not suffice to obtain an upper bound for the number i of exponents bounding the size of the uniform interpolant.

In order to obtain a tight upper bound, we need to further narrow down the subset of Σ-consequences required to obtain a uniform interpolant. To this end, we show the following:

- If we "flatten" terminologies, i.e., we reduce the maximal role depth of $\mathcal{T}$ to 1 by recursively introducing fresh concept symbols for all subconcepts occurring in $\mathcal{T}$, it is sufficient to consider the Σ-consequences stating subsumees and subsumers of all concept symbols referenced by the flattened terminology $\mathcal{T}'$ in order to preserve all Σ-consequences.
- Lemma 4 can be transferred to flattened TBoxes such that it is sufficient to consider subsumees and subsumers of role depth $2^{4\cdot|\mathsf{sub}(\mathcal{T}')|} + 1$ in order to preserve all Σ-consequences of $\mathcal{T}$.
- There is a particular type of subsumees and subsumers that do not add any consequences to the deductive closure, which we call *weak* subsumees and subsumers. These are subsumees obtained by adding arbitrary conjuncts to arbitrary subconcepts of other subsumees and, accordingly, subsumers obtained from other subsumers by omitting conjuncts from arbitrary subconcepts. When included into the uniform interpolant, weak subsumees and subsumers have a negative impact on its size. Given the exponential bound on the role depth, each concept has non-elementary many weak subsumees. Since weak subsumers and subsumees do not add any new Σ-consequences, we can safely exclude them.

We show that, in case a finite uniform $\mathcal{EL}$ interpolant of $\mathcal{T}$ with respect to Σ exists, there are at most triple-exponentially many such non-weak subsumers and subsumees of role depth up to $2^{4\cdot|\mathsf{sub}(\mathcal{T})|}+1$. Moreover, we show that each of them is of at most double-exponential size.

6.1. Flattening

Recall that we want to compute the uniform interpolant of a TBox $\mathcal{T}$ by rewriting the latter, ensuring that the part of the deductive closure of $\mathcal{T}$ consisting of Σ-consequences is preserved throughout the rewriting process. Since rewriting operates on the syntactic structure of $\mathcal{T}$, it is desirable that the syntactic structure has a close relation to the deductive closure of $\mathcal{T}$ such that we can easily manipulate the deductive closure via changes of the syntactic structure. As in other syntax-based approaches [11, 25, 21], we decompose complex axioms into syntactically simple ones. We refer to this process as *flattening*: assigning a temporary concept symbol to each complex subconcept occurring in $\mathcal{T}$, so that the terminology can be represented without nested expressions, namely using only axioms of the form $A \sqsubseteq B$, $A \equiv B_1 \sqcap \ldots \sqcap B_n$, and $A \equiv \exists r.B$, where A and $B_{(i)}$ are concept symbols or $\top$ and r is a role. For this purpose, we introduce a minimal required set of fresh concept symbols N_D with exactly one equivalence axiom $A' \equiv C'$ for each $A' \in N_D$, where C' is the subconcept of $\mathcal{T}$ replaced by A'.

In what follows, we assume terminologies to be flattened and all concepts symbols from N_D to be in $\mathrm{sig}_C(\mathcal{T}) \setminus \Sigma$. W.l.o.g., we also assume that $\mathcal{EL}$ concepts do not contain any equivalent concepts in conjunctions and that whenever several concept symbols are equivalent in $\mathcal{T}$, all their occurrences have been replaced by a single representative of the corresponding equivalence class. Concept symbols from Σ are preferred to be selected as representatives. Note that this is a preprocessing step that can be performed in polynomial time as $\mathcal{EL}$ allows for polytime reasoning. The following lemma postulates the close semantic relation between a TBox and its flattening.

Lemma 5 (Model-conservativity)**.** *Any $\mathcal{EL}$ TBox $\mathcal{T}$ can be rewritten into a flattened TBox $\mathcal{T}'$ so that each model of $\mathcal{T}'$ is a model of $\mathcal{T}$ and each model of $\mathcal{T}$ can be extended into a model of $\mathcal{T}'$.*

In the next subsection, we represent the corresponding subsumees and subsumers explicitly stated within a classified, flattened TBox $\mathcal{T}$ as a pair of regular tree grammars on ranked trees (with concept symbols interpreted as non-terminals and $\exists r, \sqcap$ as functions). We show that all non-weak subsumees and subsumers entailed by $\mathcal{T}$ can be generated by these grammars. To this end, we now analyse the derivation of subsumptions in flattened TBoxes by means of the deduction calculus introduced in Section 3.

First, we consider the derivation of subsumees. We use the auxiliary function $\mathtt{Pre} : \mathrm{sig}_C(\mathcal{T}) \to 2^{2^{\mathrm{sig}_C(\mathcal{T})}}$ which allows us for any concept symbol A to refer to its subsumees of the form $B_1 \sqcap \ldots \sqcap B_n$, where $B_{(i)}$ are concept symbols. For each such conjunction, the set of its conjuncts is an element of $\mathtt{Pre}$.

Definition 5. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and $A \in sig_C(\mathcal{T})$. $\mathtt{Pre}(A)$ is the smallest set with the following properties:*

- $\{A\} \in \mathtt{Pre}(A)$.
- *For each $K \in \mathtt{Pre}(A)$ and each $B \in K$, if there is $\mathcal{T} \models B' \sqsubseteq B$, then also $(K \setminus \{B\}) \cup \{B'\} \in \mathtt{Pre}(A)$.*
- *For each $K \in \mathtt{Pre}(A)$ and each $B \in K$, if there is $B \equiv B_1 \sqcap ... \sqcap B_n \in \mathcal{T}$, then also $(K \setminus \{B\}) \cup \{B_1, ..., B_n\} \in \mathtt{Pre}(A)$.*

We can show the following closure property of Pre.

Lemma 6. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and $A \in sig_C(\mathcal{T})$. For each $K \in \mathtt{Pre}(A)$, each $B \in K$ and each $M \in \mathtt{Pre}(B)$, we have $(K \setminus \{B\}) \cup M \in \mathtt{Pre}(A)$.*

The above lemma can be shown by an easy induction over the derivation of M from B.

In essence, the lemma below implies that, in case of flattened terminologies explicitly containing all elements of Pre, we can derive all subsumees of a concept by (1) applying the rule (Ex) to construct existential restrictions from two concepts in a subsumption relation and/or (2) replacing concepts occurring within subsumees by their subsumees.

Lemma 7. *Let $\mathcal{T}$ be a flattened $\mathcal{EL}$ TBox and C, D two $\mathcal{EL}$ concepts with $sig(C) \cup sig(D) \subseteq sig(\mathcal{T})$ such that $\mathcal{T} \models C \sqsubseteq D$. Let*

$$C = \bigsqcap_{1 \leq j \leq n} A_j \sqcap \bigsqcap_{1 \leq k \leq m} \exists r_k.E_k$$

where A_j are concept symbols, r_k are role symbols and E_k are arbitrary $\mathcal{EL}$ concepts. Then, for all conjuncts D_i of D, the following is true: If D_i is a concept symbol, then there is a set $M \in \mathtt{Pre}(D_i)$ of concept symbols from $sig_C(\mathcal{T})$ such that, for each $B \in M$, either:

(a1) There is an A_j in C such that $A_j = B$.

(a2) There are r_k, E_k and $B' \in sig_C(\mathcal{T})$ such that $\mathcal{T} \models E_k \sqsubseteq B'$ and $B \equiv \exists r_k.B' \in \mathcal{T}$.

If $D_i = \exists r'.D'$ for a role r' and an $\mathcal{EL}$ concept D', then either:

(a3) There are r_k, E_k such that $r_k = r'$ and $\mathcal{T} \models E_k \sqsubseteq D'$.

(a4) *There is* $B \in sig_C(\mathcal{T})$ *such that* $\mathcal{T} \models B \sqsubseteq \exists r'.D'$ *and* $\mathcal{T} \models C \sqsubseteq B$ *and either (a1) or (a2) holds for* $C \sqsubseteq B$.

Proof. We apply induction on the length of the proof. We start with the last applied rule and show for each possibility that the lemma holds. Rules AXTOP, AX and the case $C \bowtie D \in \mathcal{T}$ are the basis of induction, since each proof begins with one of them.

($C \bowtie D \in \mathcal{T}$) In the case that $C \sqsubseteq D \in \mathcal{T}$ or $C \equiv D \in \mathcal{T}$, the lemma holds due to the flattening. Axioms within $\mathcal{T}$ can have the following form:

- $C \in \text{sig}_C(\mathcal{T}), D = D_1 \sqcap ... \sqcap D_m$ with $m \geq 1$ and $D_1, ..., D_m \in \text{sig}_C(\mathcal{T})$. In this case, we have $\{C\} \in \texttt{Pre}(D_i)$ for each D_i with $1 \leq i \leq m$. Therefore, condition (a1) holds for each D_i.
- $C \in \text{sig}_C(\mathcal{T}), D = \exists r'.D'$ with $D' \in \text{sig}_C(\mathcal{T})$. This case corresponds to the condition (a4).

(AXTOP) Since the conjunction is empty in case $D = \top$, the lemma holds.

(AX) Since $C = D$, for each D_i there is a conjunct C_i of C with $C_i = D_i$. If D_i is a concept symbol, condition (a1) holds. Otherwise, (a3).

(EX) If EX was the last applied rule, then $D_i = \exists r_k.D'$ and $\mathcal{T} \vdash D_k \sqsubseteq D'$. Therefore, (a3) holds.

(ANDL) Assume that $C' \sqcap C'' = C$ such that $C' \sqsubseteq D$ is the antecedent. By induction hypothesis, the lemma holds for $C' \sqsubseteq D$. Since all conjuncts of C' are also conjuncts of C, the lemma holds also for $C \sqsubseteq D$.

(ANDR) Assume that $D = D_1 \sqcap D_2$, therefore, $C \sqsubseteq D_1$ and $C \sqsubseteq D_2$ is the antecedent. By induction hypothesis, the lemma holds for both $C \sqsubseteq D_1$ and $C \sqsubseteq D_2$. Since all conjuncts of D are from either D_1 or D_2, the lemma also holds for $C \sqsubseteq D$.

(CUT) By induction hypothesis, the lemma holds for both elements of the antecedent, $C \sqsubseteq C_1$ and $C_1 \sqsubseteq D$. Let $C_1 = \bigsqcap_{1 \leq p \leq r} A_p \sqcap \bigsqcap_{1 \leq s \leq t} \exists r'_s.E'_s$.

1. Assume that D_i is a concept symbol. Then, there is $M_1 \in \texttt{Pre}(D_i)$ such that (a1) or (a2) holds for each $B_u \in M_1$. We now consider each $C \sqsubseteq B_u$ and distinguish three cases, in one of which (a2) holds. In the remaining two cases, we can obtain M_{new} by replacing B_u within M_1

by the elements of some $M'_u \in \texttt{Pre}(B_u)$ such that (a1) or (a2) holds for each $B' \in M_{\text{new}}$ and $C \sqsubseteq B'$:

a. Assume that (a1) holds and that there is a conjunct A_p of C_1 with $A_p = B_1$. Then, by induction hypothesis, for $C \sqsubseteq A_p$, there is $M'_u \in \texttt{Pre}(A_p)$ such that (a1) or (a2) holds for each $B' \in M'_u$. We can replace B_u within M_1 by the elements of M'_u.

b. Assume that (a2) holds and that, for B_u, there are r'_s, E'_s and there exists $B' \in \text{sig}_C(\mathcal{T})$ such that $\mathcal{T} \models E'_s \sqsubseteq B'$ and $B \equiv \exists r'_s.B' \in \mathcal{T}$. Then, for $C \sqsubseteq \exists r'_s.E'_s$ either (a3) or (a4) holds.

   – Assume that (a3) holds. Then there are r_k, E_k such that $r_k = r'_s$ and $\mathcal{T} \models E_k \sqsubseteq E'_s$. Then (a2) holds for $C \sqsubseteq B_u$, since $\mathcal{T} \models E_k \sqsubseteq B'$ and $B \equiv \exists r_k.B' \in \mathcal{T}$.

   – Assume that (a4) holds. Then there is $B'' \in \text{sig}_C(\mathcal{T})$ such that $\mathcal{T} \models B'' \sqsubseteq \exists r'_s.E'_s$, $\mathcal{T} \models C \sqsubseteq B''$ and there is a set $M'_u \in \texttt{Pre}(B'')$ such that for each element B' of M'_u at least one of the conditions (a1)-(a2) holds with respect to $C \sqsubseteq B'$.

Let M_{A1} be the set of all $B_u \in M_1$ for which (a1) holds and let M_{A4} be the set of all $B_u \in M_1$ for which (a2) holds and (a4) holds for $C \sqsubseteq \exists r'_s.E'_s$. Now we replace each B_u within M_1 by the elements of the corresponding set $M'_u \in \texttt{Pre}(B_u)$ that we have specified above and obtain $M_{\text{new}} = M_1 \setminus (M_{A1} \cup M_{A4}) \cup \bigcup\{M'_u \mid B_u \in M_{A1} \cup M_{A4}\}$. Clearly, $M_{\text{new}} \in \texttt{Pre}(D_i)$ and (a1) or (a2) holds for each $B' \in M_{\text{new}}$ with respect to $C \sqsubseteq B'$, i.e., the lemma holds for $C \sqsubseteq D_i$.

2. Assume that $D_i = \exists r'.D'$. Then, (a3) or (a4) hold.

a. Assume that (a3) holds. Then there are r'_s, E'_s such that $r' = r'_s$ and $\mathcal{T} \models E'_s \sqsubseteq D'$. Then, for $C \sqsubseteq \exists r'_s.E'_s$ one of (a3), (a4) holds:

   – Assume that (a3) holds. Then there are r_k, E_k such that $r_k = r'_s$ and $\mathcal{T} \models E_k \sqsubseteq E'_s$. Then (a3) holds for $C \sqsubseteq D_i$, since $\mathcal{T} \models E_k \sqsubseteq D'$ and $r_k = r'$.

   – Assume that (a4) holds. Then there is a concept symbol B'' such that $\mathcal{T} \models B'' \sqsubseteq \exists r'_s.E'_s$, $\mathcal{T} \models C \sqsubseteq B''$ and there is a set $M'' \in \texttt{Pre}(B'')$ of concept symbols such that at least one of the conditions (a1)-(a2) holds for each element B' of M'' and $C \sqsubseteq B'$. Since $\mathcal{T} \models B'' \sqsubseteq D_i$, (a4) holds for $\mathcal{T} \models C \sqsubseteq D_i$.

b. Assume that (a4) holds. Then there is a concept symbol B such that $\mathcal{T} \models B \sqsubseteq \exists r'.D'$, $\mathcal{T} \models C_1 \sqsubseteq B$ and there is a set $M_1 \in$

$\mathtt{Pre}(B)$ such that at least one of the conditions (a1)-(a2) holds for each element B_u of M_1 and for $C_1 \sqsubseteq B_u$. The argumentation is the same as for 1 (D_i is a concept symbol). We consider each $C \sqsubseteq B_u$ and distinguish three cases, in one of which (a2) holds. In the remaining two cases, we can obtain M_{new} by replacing B_u within M_1 by the elements of some $M'_u \in \mathtt{Pre}(B_u)$ such that (a1) or (a2) holds for each $B' \in M_{\text{new}}$ and $C \sqsubseteq B'$. Therefore, there is $M_1 \in \mathtt{Pre}(B)$ such that either (a1) or (a2) holds for each $B_u \in M_1$. Then, (a4) holds for $C \sqsubseteq D_i$. □

The above lemma is focused on the derivation of subsumees. For the computation of uniform interpolants, we additionally need to show that, in flattened terminologies, every subsumption relation with a concept symbol and its subsumer being an existential restriction is derived from an equivalence axiom of the form $B_1 \equiv \exists r.B_2$ in $\mathcal{T}$.

Lemma 8. *Let $\mathcal{T}$ be a flattened $\mathcal{EL}$ TBox, $A \in sig_C(\mathcal{T})$ and $r \in sig_R(\mathcal{T})$. Let C be an $\mathcal{EL}$ concept such that $\mathcal{T} \models A \sqsubseteq \exists r.C$. Then, there are B_1, B_2 with $B_1 \equiv \exists r.B_2 \in \mathcal{T}$ such that $\mathcal{T} \models A \sqsubseteq B_1$, $\mathcal{T} \models B_2 \sqsubseteq C$.*

Proof. Lemma 16 [27] states that for a general $\mathcal{EL}$ TBox $\mathcal{T}$ with $\mathcal{T} \models C_1 \sqsubseteq \exists r.C_2$, where C_1, C_2 are $\mathcal{EL}$-concepts one of the following holds:

- there is a conjunct $\exists r.C'$ of C_1 such that $\mathcal{T} \models C' \sqsubseteq C_2$;
- there is a subconcept $\exists r.C'$ of $\mathcal{T}$ such that $\mathcal{T} \models C_1 \sqsubseteq \exists r.C'$ and $\mathcal{T} \models C' \sqsubseteq C_2$;

The first condition does not hold in this lemma, since A is a concept symbol. Moreover, since in our case $\mathcal{T}$ is flattened, for each subconcept $\exists r.C'$ of $\mathcal{T}$ containing an existential restriction holds: there is an concept symbol B_2 such that $B_2 = C'$ and there is an axiom of the form $B_1 \equiv \exists r.B_2 \in \mathcal{T}$ with B_1. Additionally, from the above Lemma 16 it follows that $\mathcal{T} \models A \sqsubseteq \exists r.B_2$ and $\mathcal{T} \models B_2 \sqsubseteq C$. Since $\mathcal{T} \models B_1 \equiv \exists r.B_2$, it also follows that $\mathcal{T} \models A \sqsubseteq B_1$. □

6.2. Grammar Representation of Subsumees and Subsumers

In this section, we show how the sets of Σ-subsumees and Σ-subsumers of each concept symbol in a flattened $\mathcal{EL}$ TBox $\mathcal{T}$ can be described as languages generated by regular tree grammars on ranked ordered trees for a particular signature Σ. First, we briefly recall the basics of tree languages and regular tree grammars.

6.2.1. Regular Tree Grammars

A *ranked alphabet* is a pair ($\mathcal{F}$, Arity) where $\mathcal{F}$ is a finite set and Arity is a mapping from $\mathcal{F}$ into $\mathbb{N}$. We use superscripts to denote the arity of alphabet symbols (if it is not 0), e.g., $f^2(g^1(a), a)$. The set of ground terms over the alphabet $\mathcal{F}$ (which are also simply referred to as *trees*) is denoted by $T(\mathcal{F})$. Let $\mathcal{X}$ be a set of variables. Then, $T(\mathcal{F}, \mathcal{X})$ denotes the set of terms over the alphabet $\mathcal{F}$ and the set of variables $\mathcal{X}$. A term $C \in T(\mathcal{F}, \mathcal{X})$ containing each variable from $\mathcal{X}$ at most once is called a *context*.

Example 3. *Let $\mathcal{F} = \{f^2, g^1, a\}$ and X, Y two variables. Terms $f^2(g^1(a), X)$, $f^2(g^1(Y), X)$ and $f^2(Y, X)$ are contexts obtained by replacing terminal symbols within the term $f^2(g^1(a), a)$ with a variable. The term $f^2(g^1(X), X)$ is not a context, since it contains the variable X more than once.*

A *regular tree grammar* $G = (S, \mathcal{N}, \mathcal{F}, R)$ is composed of a *start symbol* S, a set $\mathcal{N}$ of *non-terminal symbols* (of arity 0) with $S \in \mathcal{N}$, a ranked alphabet $\mathcal{F}$ of *terminal symbols* with a fixed arity such that $\mathcal{F} \cap \mathcal{N} = \emptyset$, and a set R of derivation rules, each of which is of the form $N \to \beta$ where N is a non-terminal from $\mathcal{N}$ and β is a term from $T(\mathcal{F} \cup \mathcal{N})$. Let $\mathcal{X}$ be a set of variables disjoint from the ranked alphabet $\mathcal{F} \cup \mathcal{N}$. Given a regular tree grammar $G = (S, \mathcal{N}, \mathcal{F}, R)$, the derivation relation $\to_G$ associated to G is a relation on terms from $T(\mathcal{F} \cup \mathcal{N})$ such that $s \to_G t$ if and only if there is a rule $N \to \alpha \in R$ and there is a context C such that $s = C[N/X]$ and $t = C[\alpha/X]$, where X is a variable from $\mathcal{X}$. The subset of $T(\mathcal{F} \cup \mathcal{N})$ which can be generated by successive derivations starting with the start symbol is denoted by $L_u(G) = \{s \in T(\mathcal{F} \cup \mathcal{N}) \mid S \to_G^+ s\}$ where $\to_G^+$ is the transitive closure of $\to_G$. We omit the subscript G when the grammar G is clear from the context. The language generated by G denoted by $L(G) = T(\mathcal{F}) \cap L_u(G)$.

By definition of the derivation relation, it does not necessarily hold that $f^2(B, A) \in L(G) \Rightarrow f^2(A, B) \in L(G)$ or $f^2(A, f^2(B, C)) \in L(G) \Rightarrow f^2(f^2(A, B)), C) \in L(G)$. In contrast, for DL constructs with conjunction, the order of conjuncts is not significant. For instance, for concepts C and D, it always holds that $C \sqcap D \equiv D \sqcap C$. It also holds that $D \sqcap (E \sqcap C)) \equiv ((D \sqcap E) \sqcap C)$. Since we are interested in languages consisting of DL concepts, it will be convenient to consider the *commutative associative closure* $L_u^*(G)$ and $L^*(G)$ of languages $L_u(G)$ and $L(G)$, respectively. This closure can be defined as follows.

Definition 6. *Let $G = (S, \mathcal{N}, \mathcal{F}, R)$ be a regular tree grammar and let $\mathcal{F}'$ be a subset of $\mathcal{F}$. Then the commutative associative closure $L^*(G)$ of $L(G)$ wrt. $\mathcal{F}'$ is* *the smallest set for which the following hold:*

- $L(G) \subseteq L^*(G)$*;*
- *Let* $\mathcal{X} = \{X_1, \ldots, X_n\}$ *be a set of variables and let* C *and* D *be contexts from* $T(\mathcal{F}', \mathcal{X})$*, in which every variable from* $\mathcal{X}$ *occurs exactly once. Moreover let* E *be a context over* $\mathcal{F}$ *containing a variable* X *and let* λ *be a function mapping variables from* $\mathcal{X}$ *to ground terms over* $\mathcal{F}$*. Let* $C' = C[\lambda(X_1)/X_1, \ldots, \lambda(X_n)/X_n]$ *and let* $D' = D[\lambda(X_1)/X_1, \ldots, \lambda(X_n)/X_n]$ *(where, for a variable* X *and two terms* t_1 *and* t_2*, we use* $t_1[t_2/X]$ *to denote the simultaneous replacement of all occurrences of* X *in* t_1 *by* t_2*). Then* $L^*(G)$ *contains* $E[C'/X]$ *iff it contains* $E[D'/X]$*.*

The following example demonstrates the application of derivation rules.

Example 4. *Let* $G = (A, \{A, B\}, \{f^2, g^1, a, b\}, R)$ *with* R *given by the following derivation rules:*

- $A \to f^2(B, A)$
- $A \to a$
- $B \to g^1(A)$
- $B \to b$

Then, $f^2(g^1(a), a) \in L(G)$*, since* $A \to f^2(B, A) \to f^2(B, a) \to f^2(g^1(A), a) \to f^2(g^1(a), a)$*. While* $f^2(a, g^1(a))$ *is not in* $L(G)$*, it is contained in* $L^*(G)$ *wrt.* $\{f^2\}$*.*

For further details on regular tree grammars, we refer the reader, for instance, to [28].

6.2.2. Subsumees and Subsumers as Grammars

Now, we define regular tree grammars that capture, for a signature Σ and a flattened $\mathcal{EL}$ TBox $\mathcal{T}$, the sets of Σ-subsumees and Σ-subsumers of each concept symbol. In our definition of grammars, we uniquely represent each concept symbol $A \in \mathrm{sig}_C(\mathcal{T})$ by a non-terminal $\mathfrak{n}_A$ (and denote the set of all non-terminals by $\mathcal{N}^{\mathcal{T}} = \{\mathfrak{n}_B \mid B \in \mathrm{sig}_C(\mathcal{T}) \cup \{\top\}\}$). In what follows, we use the ranked alphabet $\mathcal{F} = (\mathrm{sig}_C(\mathcal{T}) \cap \Sigma) \cup \{\top\} \cup \{\exists r^1 \mid r \in \mathrm{sig}_R(\mathcal{T}) \cap \Sigma\} \cup \{\sqcap^i \mid 2 \leq i\}$, where $\top$ and concept symbols in $\mathrm{sig}_C(\mathcal{T}) \cap \Sigma$ are constants, $\exists r^1$ for $r \in \mathrm{sig}_R(\mathcal{T}) \cap \Sigma$

are unary functions and $\sqcap^i$ are functions of the arity greater than 2. Due to flattening, $|\mathrm{sig}_C(\mathcal{T})|$ is the highest arity of conjunctions that can occur in our TBox and, as we will see, also within the languages generated by our grammars. However, higher arities are required for terms within the commutative associate closure of these languages wrt. $\{\sqcap^i \mid 2 \leq i\}$. In the following, it will be convenient to simply write $\sqcap$ and $\exists r$ if the arity of the corresponding function is clear from the context. Clearly, every $\mathcal{EL}$ concept C with $\mathrm{sig}(C) \subseteq \Sigma$ and at most $|\mathrm{sig}_C(\mathcal{T})|$ conjuncts in each subconcept has a unique representation by means of the above functions. We denote such a term representation of C using $\mathcal{F}$ by t_C. For a term t, we denote its concept representation by C_t. Additionally, we use a substitution function $\sigma_{\mathcal{T},\mathcal{F}} : \{C \mid \mathrm{sig}(C) \subseteq \mathrm{sig}(\mathcal{T})\} \to T(\mathcal{F}, \mathcal{N}^{\mathcal{T}})$ with $\sigma_{\mathcal{T},\mathcal{F}}(C) = t_C\{\mathfrak{n}_\top/\top, \mathfrak{n}_{B_1}/B_1, ..., \mathfrak{n}_{B_n}/B_n\}$, where $B_1, ..., B_n$ are all concept symbols occurring in C. If the TBox and the set of non-terminals are clear from the context, we will denote such a representation of a concept C simply by $\sigma(C)$.

As mentioned above, weak subsumees and subsumers are not required in order to obtain a uniform $\mathcal{EL}$ interpolant. Including weak subsumees into our definition of the grammars would lead to a significant redundancy within the generated language, which would become infinite even for most simple TBoxes containing roles. Weak subsumers would lead to an exponential blow-up in the size of the corresponding grammar. Thus, we avoid generating weak subsumees and subsumers by the corresponding grammars.

Definition 7. *Let $\mathcal{T}$ be a flattened $\mathcal{EL}$ TBox, Σ a signature. Further, for each $B \in sig_C(\mathcal{T}) \cup \{\top\}$, let $R^{\sqsupseteq}$ be given by*

(GL1) $\mathfrak{n}_B \to B$ *if* $B \in \Sigma \cup \{\top\}$,

(GL2) $\mathfrak{n}_B \to \mathfrak{n}_{B'}$ *for all* $B' \in sig_C(\mathcal{T}) \cup \{\top\}$ *with* $\mathcal{T} \models B' \sqsubseteq B$, $B \neq B'$

(GL3) $\mathfrak{n}_B \to \sqcap(\mathfrak{n}_{B_1}, ..., \mathfrak{n}_{B_n})$ *for all* $B \equiv B_1 \sqcap \ldots \sqcap B_n \in \mathcal{T}$,

(GL4) $\mathfrak{n}_B \to \exists r(\mathfrak{n}_{B'})$ *for all* $B \equiv \exists r.B' \in \mathcal{T}$ *with* $r \in sig_R(\mathcal{T}) \cap \Sigma$.

Let $R^{\sqsubseteq}$ be given for all $B \in sig_C(\mathcal{T}) \cup \{\top\}$ by

(GR1) $\mathfrak{n}_B \to B$ *if* $B \in \Sigma \cup \{\top\}$,

(GR2) $\mathfrak{n}_B \to \mathfrak{n}_{B'}$ *if* $B \neq B'$ *and either* $B' = \top$ *or* B' *is the only concept symbol such that* $\mathcal{T} \models B \sqsubseteq B'$,

(GR3) $\mathfrak{n}_B \to \sqcap(\mathfrak{n}_{B_1}, ..., \mathfrak{n}_{B_n})$ *if* $\{B_1, \ldots, B_n\} = \{B' \in sig_C(\mathcal{T}) \mid \mathcal{T} \models B \sqsubseteq B'\}$ *and* $n \geq 2$,

(GR4) $\mathfrak{n}_B \to \exists r(\mathfrak{n}_{B'})$ *for all* $B \equiv \exists r.B' \in \mathcal{T}$ *with* $r \in sig_R(\mathcal{T}) \cap \Sigma$.

For every $A \in sig_C(\mathcal{T})$*, the regular tree grammar* $G^{\sqsupseteq}(\mathcal{T}, \Sigma, A)$ *is given by* $(\mathfrak{n}_A, \mathcal{N}^{\mathcal{T}}, \mathcal{F}, R^{\sqsupseteq})$*. Likewise, the regular tree grammar* $G^{\sqsubseteq}(\mathcal{T}, \Sigma, A)$ *is given by* $(\mathfrak{n}_A, \mathcal{N}^{\mathcal{T}}, \mathcal{F}, R^{\sqsubseteq})$*.*

We denote the set of tree grammars $\{G^{\sqsupseteq}(\mathcal{T}, \Sigma, A) \mid A \in \mathrm{sig}_C(\mathcal{T})\}$ by $\mathbb{G}^{\sqsupseteq}(\mathcal{T}, \Sigma)$ and the set $\{G^{\sqsubseteq}(\mathcal{T}, \Sigma, A) \mid A \in \mathrm{sig}_C(\mathcal{T})\}$ by $\mathbb{G}^{\sqsubseteq}(\mathcal{T}, \Sigma)$. In what follows, $L^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$ and $L^*(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$ refer to the commutative associate closure of $L(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$ and $L(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$ wrt. $\{\sqcap^i \mid 2 \leq i\}$. For the construction of grammars the following result holds.

Theorem 2. *Let* $\mathcal{T}$ *be a flattened* $\mathcal{EL}$ *TBox and let* Σ *be a signature.* $\mathbb{G}^{\sqsupseteq}(\mathcal{T}, \Sigma)$ *and* $\mathbb{G}^{\sqsubseteq}(\mathcal{T}, \Sigma)$ *can be computed from* $\mathcal{T}$ *in polynomial time and are at most quadratic in the size of* $\mathcal{T}$*.*

Proof. Flattening and classification can be done all together in polynomial time [11] and yield an at most quadratic result. From this result, the grammars are constructed in linear time. □

The following example demonstrates the grammar construction.

Example 5. *Let* $\mathcal{T} = \{A_1 \sqsubseteq \exists r.A_2, \exists r.B_1 \sqcap B_3 \sqsubseteq B_2, A_2 \sqsubseteq B_1\}$*. In order to flatten the given TBox, we introduce fresh concept names for* $\exists r.A_2, \exists r.B_1$ *and* $B'_1 \sqcap B_3$ *to obtain* $\mathcal{T}'$*:*

$$\begin{aligned} A_1 &\sqsubseteq A'_2 & A_2 &\sqsubseteq B_1 \\ B'_2 &\sqsubseteq B_2 & B'_1 \sqcap B_3 &\equiv B'_2 \\ \exists r.B_1 &\equiv B'_1 & \exists r.A_2 &\equiv A'_2 \end{aligned}$$

Let $\Sigma = sig(\mathcal{T}) \setminus \{B_1\}$*. Then, we introduce terminals for each concept symbol from* Σ *and the* $\top$ *concept according to (GL1) and (GR1):*

$$\mathfrak{n}_{A_1} \to A_1 \qquad \mathfrak{n}_{A_2} \to A_2 \qquad \mathfrak{n}_{B_2} \to B_2 \qquad \mathfrak{n}_{\top} \to \top \tag{8}$$

If we only use subsumees explicitly given in $\mathcal{T}'$*, we obtain the following set of* *transitions $R^{\sqsupseteq}$ for generating subsumees of concept symbols:*

$$\mathfrak{n}_{A_2'} \rightarrow \mathfrak{n}_{A_1} \qquad \mathfrak{n}_{B_1} \rightarrow \mathfrak{n}_{A_2} \tag{9}$$

$$\mathfrak{n}_{B_2} \rightarrow \mathfrak{n}_{B_2'} \qquad \mathfrak{n}_{B_2'} \rightarrow \sqcap(\mathfrak{n}_{B_1'}, \mathfrak{n}_{B_3}) \tag{10}$$

$$\mathfrak{n}_{B_1'} \rightarrow \exists r(\mathfrak{n}_{B_1}) \qquad \mathfrak{n}_{A_2'} \rightarrow \exists r(\mathfrak{n}_{A_2}) \tag{11}$$

We see that the subsumee $\exists r.A_2 \sqcap B_3$ of B_2 is not generated by the above set of transitions. If we take inferred inclusions into consideration, we obtain additionally

$$\mathfrak{n}_{B_1'} \rightarrow \mathfrak{n}_{A_2'} \qquad \mathfrak{n}_{B_1'} \rightarrow \mathfrak{n}_{B_2'} \qquad \mathfrak{n}_{B_3} \rightarrow \mathfrak{n}_{B_2'} \qquad \mathfrak{n}_{B_1'} \rightarrow \mathfrak{n}_{A_1} \tag{12}$$

Accordingly, $R^{\sqsubseteq}$ is given by Rules (8),(11) and, additionally

$$\mathfrak{n}_{A_1} \rightarrow \mathfrak{n}_{\top} \qquad \mathfrak{n}_{A_2} \rightarrow \mathfrak{n}_{\top} \qquad \mathfrak{n}_{B_1} \rightarrow \mathfrak{n}_{\top} \qquad \mathfrak{n}_{B_2} \rightarrow \mathfrak{n}_{\top} \tag{13}$$

$$\mathfrak{n}_{B_3} \rightarrow \mathfrak{n}_{\top} \qquad \mathfrak{n}_{A_2'} \rightarrow \mathfrak{n}_{\top} \qquad \mathfrak{n}_{B_1'} \rightarrow \mathfrak{n}_{\top} \qquad \mathfrak{n}_{B_2'} \rightarrow \mathfrak{n}_{\top} \tag{14}$$

$$\mathfrak{n}_{A_1} \rightarrow \mathfrak{n}_{A_2'} \qquad \mathfrak{n}_{A_2} \rightarrow \mathfrak{n}_{B_1} \qquad \mathfrak{n}_{A_2'} \rightarrow \mathfrak{n}_{B_1'} \qquad \mathfrak{n}_{B_2'} \rightarrow \sqcap(\mathfrak{n}_{B_1'}, \mathfrak{n}_{B_3}, \mathfrak{n}_{B_2}) \tag{15}$$

In the above example, we can generate all non-weak subsumees using the complete grammar construction, i.e., after including the results of classification in addition to transitions representing explicitly given subsumptions. For instance, the subsumee $\exists r.A_2 \sqcap B_3$ of B_2 can be generated using the first additional rule in (12) as follows: $\mathfrak{n}_{B_2} \rightarrow \mathfrak{n}_{B_2'} \rightarrow \sqcap(\mathfrak{n}_{B_1'}, \mathfrak{n}_{B_3}) \rightarrow \sqcap(\mathfrak{n}_{A_2'}, \mathfrak{n}_{B_3}) \rightarrow \sqcap(\exists r(\mathfrak{n}_{A_1}), \mathfrak{n}_{B_3}) \rightarrow \sqcap(\exists r(A_1), B_3)$.

We now consider various properties of the above grammars that are of interest for the computation of uniform interpolants. The following theorem states that the grammars derive only terms representing Σ-subsumees and Σ-subsumers of the corresponding concept symbol.

Theorem 3. *Let $\mathcal{T}$ be a flattened $\mathcal{EL}$ TBox, Σ a signature and $A \in sig_C(\mathcal{T})$.*

1. *For each $t \in L(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$ it holds that $sig(C_t) \subseteq \Sigma$ and $\mathcal{T} \models C_t \sqsubseteq A$.*
2. *For each $t \in L(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$ it holds that $sig(C_t) \subseteq \Sigma$ and $\mathcal{T} \models A \sqsubseteq C_t$.*

Proof. It is easy to check that $\text{sig}(C_t) \subseteq \Sigma$ in 1 and 2 by examining Definition 7: the grammars derive only terms containing concept symbols and roles from Σ, since $\mathfrak{n}_B \rightarrow B$ only if $B \in \Sigma \cup \{\top\}$ and $\mathfrak{n}_B \rightarrow \exists r(t')$ only if $r \in \Sigma$. Therefore, for any $A \in \text{sig}_C(\mathcal{T})$ and any $t \in L(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A)) \cup L(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$ it holds that

$\text{sig}(C_t) \subseteq \Sigma$. To show that grammars only generate subsumees and subsumers, we investigate the above two cases separately:

1. We use an easy induction on the maximal nesting depth of functions in t using the rules given in Definition 7:

   - Assume that C_t is a concept symbol B or $\top$. The term B can only be derived from $\mathfrak{n}_A$ by n transitions (GL2), and, once $\mathfrak{n}_B$ is derived, the rule (GL1). Let $B_1, ..., B_n$ be such that $\mathfrak{n}_A \to \mathfrak{n}_{B_1} \to ... \to \mathfrak{n}_{B_n} \to \mathfrak{n}_B$. Then, by Definition 7, $\mathcal{T} \models A \sqsupseteq B_1$, $\mathcal{T} \models B_i \sqsupseteq B_{i+1}$, for $i < n$, and $\mathcal{T} \models B_n \sqsupseteq B$. Thus, $\mathcal{T} \models A \sqsupseteq B$.

   - Assume that $t = \exists r(t')$ for some term t'. Then, the derivation of t from $\mathfrak{n}_A$ starts with n transitions (GL2) such that $\mathfrak{n}_{B'}$ for some $B' \in \text{sig}_C(\mathcal{T}) \cup \{\top\}$ is derived, and a subsequent application of (GL4) such that $\mathfrak{n}_B$ for some $B \in \text{sig}_C(\mathcal{T}) \cup \{\top\}$ is derived. As argued above about the applications of transitions (GL2), $\mathcal{T} \models A \sqsupseteq B'$ holds. Moreover, by (GL4) it holds that $B' \equiv \exists r.B \in \mathcal{T}$, and, therefore, $\mathcal{T} \models A \sqsupseteq \exists r.B$. Let $C' = C_{t'}$. Then, by induction hypothesis, $\mathcal{T} \models B \sqsupseteq C'$. Therefore, $\mathcal{T} \models A \sqsupseteq \exists r.C'$, while $\exists r.C' = C_t$.

   - Assume that $t = \sqcap(t_1, ..., t_n)$ for a set of terms $t_1, ..., t_n$. Then, the derivation of t from $\mathfrak{n}_A$ starts with m transitions (GL2) such that $\mathfrak{n}_{B'}$ for some $B' \in \text{sig}_C(\mathcal{T}) \cup \{\top\}$ is derived, and a subsequent application of (GL3) such that we derive $\sqcap(\mathfrak{n}_{B_1}, ..., \mathfrak{n}_{B_n})$, where $t_i \in L(G^{\sqsupseteq}(\mathcal{T}, \Sigma, \mathfrak{n}_{B_i}))$ for $1 \leq i \leq n$. As argued above about the applications of transitions (GL2), $\mathcal{T} \models A \sqsupseteq B'$. Let $C_i = C_{t_i}$. By induction hypothesis, $\mathcal{T} \models B_i \sqsupseteq C_i$. By Definition 7, $B' \equiv B_1 \sqcap ... \sqcap B_n \in \mathcal{T}$. Therefore, $\mathcal{T} \models B' \sqsupseteq C_1 \sqcap ... \sqcap C_n$ and $\mathcal{T} \models A \sqsupseteq C_1 \sqcap ... \sqcap C_n$ with $C_1 \sqcap ... \sqcap C_n = C_t$.

2. The proof of soundness of $\mathbb{G}^{\sqsubseteq}(\mathcal{T}, \Sigma)$ can be done in the same manner, i.e., by induction on the maximal nesting depth of functions in t:

   - Assume that C_t is a concept symbol B or $\top$. The term B can only be derived from $\mathfrak{n}_A$ by n transitions (GR2), and, once $\mathfrak{n}_B$ is derived, the rule (GR1). Let $B_1, ..., B_n$ be such that $\mathfrak{n}_A \to \mathfrak{n}_{B_1} \to ... \to \mathfrak{n}_{B_n} \to \mathfrak{n}_B$. Then, by Definition 7, for each pair B_i, B_{i+1} it holds that $\mathcal{T} \models B_i \sqsubseteq B_{i+1}$, for B_n, B it holds that $\mathcal{T} \models B_n \sqsubseteq B$ and for A, B_1 it holds that $\mathcal{T} \models A \sqsubseteq B_1$. It follows that also $\mathcal{T} \models A \sqsubseteq B$ with $C_t = B$.

- Assume that $t = \exists r(t')$ for some term t'. Then, the derivation of t from $\mathfrak{n}_A$ starts with n transitions (GR2) such that $\mathfrak{n}_{B'}$ for some $B' \in \mathsf{sig}_C(\mathcal{T}) \cup \{\top\}$ is derived, and a subsequent application of a transition (GR4) such that $\exists r.\mathfrak{n}_B$ for some $B \in \mathsf{sig}_C(\mathcal{T}) \cup \{\top\}$ is derived. As argued above about the applications of transitions (GR2), $\mathcal{T} \models A \sqsubseteq B'$ holds. Moreover, By Definition 7, it holds that $\mathcal{T} \models B' \equiv \exists r.B$, and, therefore, $\mathcal{T} \models A \sqsubseteq \exists r.B$. Let $C' = C_{t'}$. By induction hypothesis, $\mathcal{T} \models B \sqsubseteq C'$. Therefore, $\mathcal{T} \models A \sqsubseteq \exists r.C'$ with $C_t = \exists r.C'$.
- Assume that $t = \sqcap(t_1, ..., t_n)$ for a set of terms $t_1, ..., t_n$. Then, the derivation of t from $\mathfrak{n}_A$ starts with m transitions (GR2) such that $\mathfrak{n}_{B'}$ for some $B' \in \mathsf{sig}_C(\mathcal{T}) \cup \{\top\}$ is derived, and a subsequent application of (GR3) such that we derive $\sqcap(\mathfrak{n}_{B_1}, ..., \mathfrak{n}_{B_n})$, where $t_i \in L(G^{\sqsupseteq}(\mathcal{T}, \Sigma, \mathfrak{n}_{B_i}))$ for $1 \leq i \leq n$ and $n \geq 2$. As argued above about the applications of transitions (GR2), $\mathcal{T} \models A \sqsubseteq B'$. Let $C_i = C_{t_i}$. By induction hypothesis, $\mathcal{T} \models B_i \sqsubseteq C_i$. By Definition 7, $\mathcal{T} \models B' \sqsubseteq B_1 \sqcap ... \sqcap B_n$. Therefore, $\mathcal{T} \models B' \sqsubseteq C_1 \sqcap ... \sqcap C_n$ and $\mathcal{T} \models A \sqsubseteq C_1 \sqcap ... \sqcap C_n$ with $C_1 \sqcap ... \sqcap C_n = C_t$.

To be able to show completeness of the grammars, we first show that the commutative associative closure of the generated $G^{\sqsupseteq}$ language contains all elements of `Pre`.

Lemma 9. *Let $\mathcal{T}$ be flattened $\mathcal{EL}$ TBox, Σ a signature, A a concept symbol and $K \in \mathtt{Pre}(A)$. Then, $\sigma(\bigsqcap_{B \in K} B) \in L_u^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$.*

Proof. The lemma can be shown by an easy induction on the depth of derivation of K from A. We distinguish three cases for the last derivation step.

- If $K = \{A\}$, then the lemma is a direct consequence of (GL1).
- Assume that K has been obtained from $K' \in \mathtt{Pre}(A)$ by replacing some B by some B' such that $\mathcal{T} \models B' \sqsubseteq B$. By induction hypothesis, $\sigma(\bigsqcap_{B'' \in K'} B'') \in L_u^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$. By (GL2), we have $\mathfrak{n}_B \rightarrow \mathfrak{n}_{B'} \in R^{\sqsupseteq}$. Thus, also $\sigma(\bigsqcap_{A' \in K} A') \in L_u^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$.
- Assume that K has been obtained from $K' \in \mathtt{Pre}(A)$ by replacing some B by some $B_1, \ldots, B_n$ such that $B \equiv B_1 \sqcap ... \sqcap B_n \in \mathcal{T}$. By induction hypothesis, $\sigma(\bigsqcap_{B'' \in K'} B'') \in L_u^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$. By (GL3), we have $\mathfrak{n}_B \rightarrow \sqcap(\mathfrak{n}_{B_1}, ..., \mathfrak{n}_{B_n}) \in R^{\sqsupseteq}$. Thus, also $\sigma(\bigsqcap_{A' \in K} A') \in L_u^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$. □

As discussed above, grammars do not guarantee to capture weak subsumees and subsumers. Therefore, we obtain the following result for the completeness of the grammars.

Theorem 4. *Let $\mathcal{T}$ be a flattened $\mathcal{EL}$ TBox, Σ a signature and A a concept symbol.*

1. *For each C with sig$(C) \subseteq \Sigma$ such that $\mathcal{T} \models C \sqsubseteq A$ there is a concept C' with $t_{C'} \in L^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$ such that C can be obtained from C' by adding arbitrary conjuncts to arbitrary subconcepts.*
2. *For each C with sig$(C) \subseteq \Sigma$ such that $\mathcal{T} \models A \sqsubseteq C$ there is a concept C' with $t_{C'} \in L^*(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$ such that C can be obtained from C' by removing $\top$ conjuncts from arbitrary subconcepts.*

Proof. The theorem is proved by induction on the role depth of C using the properties of the flattening, for instance, stated in Lemma 7, in addition to Definition 7 and Lemma 9. Let

$$C = \bigsqcap_{1 \leq j \leq n} A_j \sqcap \bigsqcap_{1 \leq k \leq m} \exists r_k.E_k$$

where A_j are concept symbols, r_k are role symbols and E_k are arbitrary $\mathcal{EL}$ concepts. W.l.o.g., we can assume that all A_j are pairwise different. We prove the first claim as follows:

- Assume role depth is 0. Then, C is a conjunction of concept symbols, i.e., $C = \bigsqcap_{1 \leq j \leq n} A_j$. By Lemma 7, there is a set $M' \in \texttt{Pre}(A)$ of concept symbols such that, for each $B \in M'$, there is an A_j with $A_j = B$. By Lemma 9, $\sigma(\bigsqcap_{B \in M'} B) \in L_u^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$. Since each $B \in M'$ is in Σ, by (GL1), $\mathfrak{n}_B \to B \in R^{\sqsupseteq}$. It follows that $t_C \in L^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$.

- Assume that the role depth is greater than 0. As in the case above, there is a set $M' \in \texttt{Pre}(A)$ of concept symbols such that, for each $B \in M'$, [A1] or [A2] holds. Let M_1' be the subset of M' where [A1] holds, i.e., $M_1' = M' \cap \{A_1, ...A_n\}$, and let $M_2' = M' \setminus M_1'$. In accordance with this separation of M' into M_1' and M_2', we can also identify the two corresponding subconjunctions of C: Let $C_1' = \bigsqcap_{B \in M_1'} B$, and $C_2' = \bigsqcap_{1 \leq f \leq p} \exists r_f'.E_f'$ such that for each f there is a corresponding $B_f \in M_2'$.

  For each f, there exists a concept symbol B_f' with $\mathcal{T} \models E_f' \sqsubseteq B_f'$ and $B_f \equiv \exists r.B_f' \in \mathcal{T}$. By induction hypothesis, for each f there exists a concept E_f'' such that $t_{E_f''} \in L^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, B_f'))$ and E_f' can be obtained from E_f'' by adding arbitrary conjuncts to arbitrary subconcepts. By (GL4),

$\mathfrak{n}_{B_f} \rightarrow \exists r'_f(\mathfrak{n}_{B'_f}) \in R^{\sqsupseteq}$. Therefore, $\exists r'_f(t_{E''_f}) \in L^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, B_f))$ and $\exists r'_f.E'_f$ can be obtained from $\exists r'_f.E''_f$ by adding arbitrary conjuncts to arbitrary subconcepts.

Since each $B \in M'_1$ is in Σ, we have $\mathfrak{n}_B \rightarrow B \in R^{\sqsupseteq}$ by (GL1). By Lemma 9, $\sigma(\bigsqcap_{B \in M'} B) \in L^*_u(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$. Thus, we obtain a concept $C'' = \bigsqcap_{B \in M'_1} B \sqcap \bigsqcap_{B_f \in M'_2} \exists r'_f.E''_f$ with $t_{C''} \in L^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$ such that C can be obtained from it by adding arbitrary conjuncts to arbitrary subconcepts.

We proceed with showing that for each such general C with $\mathrm{sig}(C) \subseteq \Sigma$ such that $\mathcal{T} \models A \sqsubseteq C$ there is a concept C' such that $t_{C'} \in L^*(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$ and C can be obtained from C' by removing $\top$ conjuncts from arbitrary subconcepts. For each A_j, we know that $\mathcal{T} \models A \sqsubseteq A_j$ and $A_j \in \Sigma \cup \{\top\}$. By (GR1) $\mathfrak{n}_{A_j} \rightarrow A_j \in R^{\sqsubseteq}$ for all A_j. Assume a role depth 0.

- Assume that $n = 1$, i.e., $C = A_1$, and assume that A_1 is the only concept symbol such that $\mathcal{T} \models A \sqsubseteq A_1$. By (GR2) $\mathfrak{n}_A \rightarrow \mathfrak{n}_{A_1} \in R^{\sqsubseteq}$. Thus, $t_C \in L^*(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$.

- Assume that there are more than one concept symbol A_i such that $\mathcal{T} \models A \sqsubseteq A_i$. By (GR3), $\mathfrak{n}_A \rightarrow \sqcap(\mathfrak{n}_{A_1}, ..., \mathfrak{n}_{A_l}) \in R^{\sqsubseteq}$ for some $l \geq n$. By (GR2), there is $\mathfrak{n}_{A_i} \rightarrow \mathfrak{n}_\top \in R^{\sqsubseteq}$ for all A_i. By applying (GR1) for all A_j and $\mathfrak{n}_{A_i} \rightarrow \mathfrak{n}_\top$, $\mathfrak{n}_\top \rightarrow \top$ for all $i > n$, we obtain a term $t_{C \sqcap C'}$, where C' is a conjunction of $x - n$ concepts $\top$. Thus, the theorem holds for role depth 0.

Assume that the role depth is greater than 0. For each $\exists r_k.E_k$, it follows from Lemma 8 that there are $B_k, B''_k \in \mathrm{sig}_C(\mathcal{T})$ with $B_k \equiv \exists r_k.B''_k \in \mathcal{T}$ such that $\mathcal{T} \models A \sqsubseteq B_k$, $\mathcal{T} \models B''_k \sqsubseteq E_k$. By (GR4), $\mathfrak{n}_{B_k} \rightarrow \exists r_k(\mathfrak{n}_{B''_k}) \in R^{\sqsubseteq}$. By induction hypothesis, there is a concept E'_k such that $t_{E'_k} \in L^*(G^{\sqsubseteq}(\mathcal{T}, \Sigma, B''_k))$ and E_k can be obtained from E'_k by removing $\top$ conjuncts from arbitrary subconcepts.

- Assume that there is the only one concept symbol B' such that $\mathcal{T} \models A \sqsubseteq B'$. Then, $C = \exists r_1.E_1$ and $B_1 = B'$. By (GR2) $\mathfrak{n}_A \rightarrow \mathfrak{n}_{B'} \in R^{\sqsubseteq}$. Thus, $t_{\exists r_1.E'_1} \in L(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$ and $\exists r_1.E_1$ can be obtained from $\exists r_1.E'_1$ by removing $\top$ conjuncts from arbitrary subconcepts.

- Assume that there are more than one concept symbol B' such that $\mathcal{T} \models A \sqsubseteq B'$. By (GR3), $\mathfrak{n}_A \rightarrow \sqcap(\mathfrak{n}_{B'_1}, ..., \mathfrak{n}_{B'_l}) \in R^{\sqsubseteq}$ for some $l \geq n + m$ such that $B'_j = A_j$ for $1 \leq j \leq n$ and $B'_{n+k} = B_k$ for $1 \leq k \leq m$. By (GR2), there is $\mathfrak{n}_{B'_i} \rightarrow \mathfrak{n}_\top \in R^{\sqsubseteq}$ for all B'_i. Now, we derive the term $t_{C'' \sqcap C'}$ from $\mathfrak{n}_A$ by first applying $\mathfrak{n}_A \rightarrow \sqcap(\mathfrak{n}_{B'_1}, ..., \mathfrak{n}_{B'_l})$ and then proceeding as follows:

- from each B'_i with $i > n + m$, we derive $\top$ by applying $\mathfrak{n}_{B'_i} \rightarrow \mathfrak{n}_\top$, $\mathfrak{n}_\top \rightarrow \top$;
- from each $B'_j = A_j$ with $1 \leq j \leq n$, we derive A_j by applying $\mathfrak{n}_{B'_j} \rightarrow A_j$;
- from each $B'_{n+k} = B_k$ with $1 \leq k \leq m$, we derive $t_{\exists r_k.E'_k}$.

We obtain a term $t_{C'' \sqcap C'} \in L^*(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$, where C' is a conjunction of concepts $\top$ and $C'' = \bigsqcap_{1 \leq j \leq n} A_j \sqcap \bigsqcap_{1 \leq k \leq m} \exists r_k.E'_k$. Clearly, C can be obtained from C'' by removing $\top$ conjuncts from arbitrary subconcepts. Thus, C can be obtained from $C'' \sqcap C'$ by removing the $\top$ conjuncts from arbitrary subconcepts. □

6.3. From Grammars to Uniform Interpolants

Now we show that, as a consequence of Lemma 4 and Theorem 4, in case a finite uniform interpolant exists, we can construct it from the subsumees and subsumers of maximal depth $N = 2^{4 \cdot |\text{sub}(\mathcal{T})|} + 1$ generated by the grammars $\mathbb{G}^{\sqsupseteq}(\mathcal{T}, \Sigma)$ and $\mathbb{G}^{\sqsubseteq}(\mathcal{T}, \Sigma)$. To this end, we represented subsumees and subsumers of maximal depth N as a TBox that uses only concept and role symbols from Σ as follows.

Definition 8. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and Σ a signature. Let N be a natural number. For $\bowtie \in \{\sqsupseteq, \sqsubseteq\}$ and $A \in sig_C(\mathcal{T})$, let $L_{\bowtie}(A) = \{C \mid t_C \in L(G^{\bowtie}(\mathcal{T}, \Sigma, A)), d(C) \leq N\}$. Then the* grammar-generated TBox $\mathbb{T}^{\mathbb{G}}$ *for $\mathcal{T}$, Σ and N is defined as follows:*

$$\begin{aligned}\mathbb{T}^{\mathbb{G}}(\mathcal{T}, \Sigma, N) = &\{C \sqsubseteq A \mid A \in \Sigma \cap sig_C(\mathcal{T}), C \in L_{\sqsupseteq}(A)\} \cup \\ &\{A \sqsubseteq D \mid A \in \Sigma \cap sig_C(\mathcal{T}), D \in L_{\sqsubseteq}(A)\} \cup \\ &\{C \sqsubseteq D \mid A \in sig_C(\mathcal{T}) \setminus \Sigma, \\ &\quad C \in L_{\sqsupseteq}(A), D \in L_{\sqsubseteq}(A)\}.\end{aligned}$$

We now show that $\mathbb{T}^{\mathbb{G}}(\mathcal{T}, \Sigma, N)$ is a uniform $\mathcal{EL}$ Σ-interpolant of $\mathcal{T}$ in case there exists one and obtain an upper bound on its size.

Theorem 5. *Let $\mathcal{T}$ be a flattened version of an $\mathcal{EL}$ TBox $\mathcal{T}_{nf}$ and let Σ be a signature with $\Sigma \cap sig(\mathcal{T}) \subseteq sig(\mathcal{T}_{nf})$. Let $N = 2^{4 \cdot |sub(\mathcal{T}_{nf})|} + 1$. The following statements are equivalent:*

1. *There exists a uniform $\mathcal{EL}$ Σ-interpolant of $\mathcal{T}_{nf}$.*
2. $\mathbb{T}^{\mathbb{G}}(\mathcal{T}, \Sigma, N) \equiv_{\Sigma}^{\mathcal{EL}} \mathcal{T}_{nf}$

3. *There exists a uniform $\mathcal{EL}$ Σ-interpolant $\mathcal{T}'$ of $\mathcal{T}_{nf}$ with $|\mathcal{T}'| \in O(2^{2^{2^{|\mathcal{T}_{nf}|}}})$.*

Proof. We prove the implications $1 \Rightarrow 2$ and $2 \Rightarrow 3$. All other implications are either trivial or follow from the others. For convenience, let $\mathcal{T}_\Sigma$ denote the TBox $\mathbb{T}^{\mathbb{G}}(\mathcal{T}, \Sigma, N)$.

We start by showing the implication $1 \Rightarrow 2$. First, note that the statement $\mathcal{T}_\Sigma \equiv_\Sigma^{\mathcal{EL}} \mathcal{T}_{nf}$ follows from Lemma 5 and the fact that $\Sigma \cap \text{sig}(\mathcal{T}) \subseteq \text{sig}(\mathcal{T}_{nf})$. Thus, it is sufficient to prove $\mathcal{T}_\Sigma \equiv_\Sigma^{\mathcal{EL}} \mathcal{T}$. By Definition 2, the statement $\mathcal{T}_\Sigma \equiv_\Sigma^{\mathcal{EL}} \mathcal{T}$ consists of two directions: (1) for all $\mathcal{EL}$ concepts C, D with $\text{sig}(C) \cup \text{sig}(D) \subseteq \Sigma$ it holds that $\mathcal{T}_\Sigma \models C \sqsubseteq D \Rightarrow \mathcal{T} \models C \sqsubseteq D$ and (2) for all $\mathcal{EL}$ concepts C, D with $\text{sig}(C) \cup \text{sig}(D) \subseteq \Sigma$ it holds that $\mathcal{T}_\Sigma \models C \sqsubseteq D \Leftarrow \mathcal{T} \models C \sqsubseteq D$.

(1) The first direction follows from Theorem 3 and Definition 8. Theorem 3 ensures that the subsumees and subsumers used within Definition 8 are all entailed by $\mathcal{T}$. Theorem 3 and Definition 8 imply that $\mathcal{T}_\Sigma$ does not contain any concept and role symbols not from Σ.

(2) For the second direction, assume that there exists a uniform $\mathcal{EL}$ Σ-interpolant of $\mathcal{T}_{nf}$ and, subsequently, $\mathcal{T}$. Then, by Lemma 4, there exists a uniform $\mathcal{EL}$ Σ-interpolant $\mathcal{T}'$ of $\mathcal{T}_{nf}$ and $\mathcal{T}$ with $d(\mathcal{T}') \leq N$. It is sufficient to show that for each $C \sqsubseteq D \in \mathcal{T}'$ it holds that $\mathcal{T}_\Sigma \models C \sqsubseteq D$. Assume that $C \sqsubseteq D \in \mathcal{T}'$. We prove by induction on maximal role depth of C, D that also $\mathcal{T}_\Sigma \models C \sqsubseteq D$. Let $D = \bigsqcap_{1 \leq i \leq l} D_i$ and

$$C = \bigsqcap_{1 \leq j \leq n} A_j \sqcap \bigsqcap_{1 \leq k \leq m} \exists r_k.E_k$$

where A_j are concept symbols, r_k are role symbols and E_k are arbitrary $\mathcal{EL}$ concepts. Clearly, $\mathcal{T} \models C \sqsubseteq D$ iff $\mathcal{T} \models C \sqsubseteq D_i$ for all i with $1 \leq i \leq l$.

- If D_i is a concept symbol, then, it follows from Theorem 4 that there is a concept C' such that $t_{C'} \in L^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$ and C can be obtained from C' by adding arbitrary conjuncts to arbitrary subconcepts. Since $d(C) \leq N$, also $d(C') \leq N$. From Definition 8, it follows that $\mathcal{T}_\Sigma \models C \sqsubseteq D_i$.
- If $D_i = \exists r.D'$ for some r, D', then, by Lemma 7, one of the following is true:

(a3) There are r_k, E_k in C such that $r_k = r$ and $\mathcal{T} \models E_k \sqsubseteq D'$. Since $d(E_k) < N$ and $d(D') < N$, by induction hypothesis, it holds that $\mathcal{T}_\Sigma \models E_k \sqsubseteq D'$. It follows that $\mathcal{T}_\Sigma \models \exists r_k.E_k \sqsubseteq D_i$ and $\mathcal{T}_\Sigma \models C \sqsubseteq D_i$.

(a4) There is a concept symbol $B \in \mathrm{sig}_C(\mathcal{T})$ such that $\mathcal{T} \models B \sqsubseteq \exists r.D'$ and $\mathcal{T} \models C \sqsubseteq B$. Then, on one hand, it follows from Theorem 4 that there is a concept C'_1 such that $t_{C'_1} \in L^*(G^{\sqsupseteq}(\mathcal{T}, \Sigma, B))$ and C can be obtained from C'_1 by and adding arbitrary conjuncts to arbitrary subconcepts. Since $d(C) \leq N$, also $d(C'_1) \leq N$. Therefore, there is a commutative associative variant C''_1 of C'_1 with $C''_1 \in L_{\sqsupseteq}(B)$. On the other hand, it follows from Theorem 4 that there is a concept C'_2 such that $t_{C'_2} \in L^*(G^{\sqsubseteq}(\mathcal{T}, \Sigma, B))$ and $\exists r.D'$ can be obtained from C'_2 by removing $\top$ conjuncts from arbitrary subconcepts. Since $d(\exists r.D') \leq N$, also $d(C'_2) \leq N$. Therefore, there is a commutative associative variant C''_2 of C'_2 with $C''_2 \in L_{\sqsubseteq}(B)$.
By Definition 8, $C''_1 \sqsubseteq C''_2 \in \mathcal{T}_\Sigma$, and, therefore, $\mathcal{T}_\Sigma \models C \sqsubseteq D_i$.

Now we show the implication $2 \Rightarrow 3$. Observe that $\mathbb{G}^{\sqsupseteq}(\mathcal{T}, \Sigma)$ and $\mathbb{G}^{\sqsubseteq}(\mathcal{T}, \Sigma)$ have $n = |\mathrm{sig}_C(\mathcal{T})|$ non-terminals and n is also the maximal arity of $\sqcap$. Now we consider the stepwise generation of terms in $L(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$ and $L(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$. Initially, terms are given by transitions. Assume that m is the maximal number of transitions in $\mathbb{G}^{\sqsupseteq}(\mathcal{T}, \Sigma)$ and $\mathbb{G}^{\sqsubseteq}(\mathcal{T}, \Sigma)$. Note that m is polynomial in n. Each of these outgoing transitions has at most n occurring non-terminals. For a term t of role depth x, we can obtain a term of the role depth $x + 1$ by first applying transition rules of type GL1-GL3 (GR1-GR3 in case of subsumer terms) to replace non-terminals $\mathfrak{n}$ by terms t' and then applying transitions of type GL4 (GR4). In case of subsumees, we can assume that it is sufficient to consider terms t' with a maximal function depth m (maximal number of transitions), since a repeated application of the same transition of type GL3 generates a weak subsumee that is not required for the construction of the uniform interpolant. The total maximal depth of function nestings in subsumee terms is then $N \cdot m$. In case of subsumers, the term of the role depth $x + 1$ is obtained by applying at most one rule of type GR3 for each non-terminal, since the corresponding conjunctions in GR3 contain all non-terminals that can be obtained by infinitely many successive applications of GR1-GR3. The total maximal depth of function nestings in subsumer terms is then $N \cdot 2$. Given the maximal function depth $N \cdot m$, the maximal arity n of functions and the number n of different non-terminals, we obtain at most $n^{n^{N \cdot m}}$

different terms. Since in $N \in O(2^n)$, the size of terms is in $O(2^{2^n})$ while the number of terms is in $O(2^{2^{2^n}})$.

These complexity results correspond to the size and number of axioms in Example 4 used to demonstrate the triple-exponential lower bound. □

7. Related Work

In addition to the already discussed results on uniform interpolation in description logics [21, 18, 22, 29, 30, 16, 17], in this section we discuss the work on inseparability and conservative extensions. The latter two notions form the foundation for module extraction, e.g., [31, 17, 32], and decomposition of ontologies into modules, e.g., [33, 34, 35]. The notion of a conservative extension is defined using inseparability: A TBox $\mathcal{T}_1$ is called a Σ-conservative extension of a TBox $\mathcal{T}_2$ if $\mathcal{T}_1$ is Σ-inseparable from $\mathcal{T}_2$ and $\mathcal{T}_1 \subseteq \mathcal{T}_2$.

Ghilardi, Lutz and Wolter [36] investigate modularity of ontologies based on inseparability for the logic $\mathcal{ALC}$, which is defined in the same way as inseparability for $\mathcal{EL}$ in Section 4. They show that deciding if a subontology is a module in the description logic $\mathcal{ALC}$ is 2EXPTIME-complete. In a subsequent work, Lutz, Walther and Wolter [37] show that the same problem is 2EXPTIME-complete for $\mathcal{ALCQI}$, but undecidable for $\mathcal{ALCQIO}$. The authors also investigate a stronger notion of inseparability and conservative extensions defined directly on models instead of entailed consequences: given two TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$, $\mathcal{T}_1$ is a *model-conservative extension* of $\mathcal{T}_2$ iff for every model $\mathcal{I}$ of $\mathcal{T}_2$, there exists a model of $\mathcal{T}_1$ which can be obtained from $\mathcal{I}$ by modifying the interpretation of symbols in $\text{sig}(\mathcal{T}_1) \setminus \text{sig}(\mathcal{T}_2)$ while leaving the interpretation of symbols in $\text{sig}(\mathcal{T}_2)$ fixed. The authors show that the corresponding problem based on the latter notion is undecidable for $\mathcal{ALC}$.

In a more recent work, Konev, Lutz, Walther and Wolter [32] consider the decidability of the above problem based on model-conservative extensions for $\mathcal{ALC}$ under different additional restrictions, e.g., restriction of the relevant signature to concept names, and obtain complexity results ranging from Π_2^p to undecidable. Further, the authors consider the problem for acyclic $\mathcal{EL}$ terminologies. It is interesting that, in contrast to acyclic $\mathcal{ALC}$ terminologies, for which the problem remains undecidable, for acyclic $\mathcal{EL}$ terminologies the complexity goes down to PTIME. In a later work [38], the above authors present a full complexity picture for $\mathcal{ALC}$ and its common extensions. They investigate a broad range of query languages (languages in which the relevant consequences are expressed), starting with the language allowing for expressing inconsistency only and ending with

Second Order Logic. More recently, Lutz and Wolter [27] show that the above notion of model-conservative extensions is undecidable also for such a lightweight logic as $\mathcal{EL}$.

Kontchakov, Wolter and Zakharyaschev [39] investigate the above decision problem for two representatives of the DL-Lite family of description logics as ontology languages and existential Σ-queries as a query language. They show that, for DL-Lite$_{\texttt{horn}}$, the problem is coNP-complete, and for DL-Lite$_{\texttt{bool}}$ Π_2^p-complete.

The high complexity results for already rather simple logics have lead to a development of alternative ways to extract modules not requiring checking inseparability. For instance, Cuenca Grau, Horocks, Kazakov and Sattler [31], propose a tractable algorithm for computing modules from OWL DL ontologies based on the notion of *syntactic locality* [40] that defines the locality of an axiom on the syntactic level, i.e., states syntactic conditions for the potential logical relevance of axioms. It is guaranteed that the extracted module preserves all relevant consequences, but the obtained modules are not necessarily minimal.

8. Discussion and Outlook

In this article, we considered the task of uniform interpolation – reformulation of an ontology into an alternative one that uses only a specific subset Σ of the initial signature and preserves all logical consequences about concept and role symbols from Σ. We proposed an approach to computing uniform interpolants of general $\mathcal{EL}$ terminologies based on proof theory and regular tree languages.

One of our results is a representation of $\mathcal{EL}$ TBoxes as regular tree grammars. These grammars reduce the computation of logical consequences of ontologies to replacement of concept symbols within concepts by their subsumees/subsumers, which can serve as a basis for efficient ontology reformulation algorithms as demonstrated, for instance, in a follow-up work within the context of module extraction [23]. While the grammars were designed to enable forgetting-related ontology reformulation, they can also serve as a basis for computing equivalent, but syntactically different ontologies. For instance, they can be used as a starting point for computing structurally simpler, equivalent ontologies, i.e., ontologies containing fewer DL constructs such as conjunction and existential restriction and fewer references to concept and role symbols [41, 42], since the grammars capture all logical consequences that might be part of the less complex ontology.

A further result obtained within this paper is a tight triple exponential bound on the size of uniform $\mathcal{EL}$ interpolants: we showed that, if a finite uniform $\mathcal{EL}$ interpolant exists, then there exists one of at most triple exponential size in terms of the original TBox, and that, in the worst case, no shorter interpolant exists. This insight reveals the effect of structure sharing in the basic logic $\mathcal{EL}$ and demonstrates the worst-case behavior of uniform interpolation algorithms used for the purpose of module extraction. It should be noted that this result does not constitute a fundamental technical obstacle for uniform interpolation in practice as demonstrated by Nikitina and Glimm [23], but rather contributes towards a higher stability of the corresponding tool support through the awareness of the theoretically possible triple exponential blowup. Further, this result reveals the theoretic extent, to which we may be able to increase the succinctness of ontologies by introducing additional vocabulary. Some ideas discussed within this paper inspired a follow-up work on refactoring of large and complex ontologies in order to reduce the maintenance effort by simplifying their structure and eliminating redundancy.

Acknowledgments

This work was supported by the project ExpresST funded by the German Research Foundation (DFG).

Appendix A. Model-Theoretic Properties of $\mathcal{EL}$ Concepts

In Section 2, we characterize $\mathcal{EL}$ concept membership and $\mathcal{EL}$ concept subsumption in the absence of terminological background knowledge. In this section, we include the according proofs.

Lemma 1. For any $\mathcal{EL}$ concept C and any interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ and $x \in \Delta^{\mathcal{I}}$ it holds that $x \in C^{\mathcal{I}}$ if and only if there is a homomorphism from $(\mathcal{I}_C, x_C)$ to $(\mathcal{I}, x)$.

Proof. We prove both directions by structural induction over C.

We start with the if-direction, letting φ be a homomorphism from $(\mathcal{I}_C, x_C)$ to $(\mathcal{I}, x)$:

- For $C = \top$, the case is trivial.
- For $C = A \in N_C$, we find $x_A \in A^{\mathcal{I}_A}$, therefore the existence of the homomorphism ensures that $x = \varphi(x_A) \in A^{\mathcal{I}}$.
- For $C = C_1 \sqcap C_2$, we find that $\varphi_\iota : \Delta^{\mathcal{I}_{C_\iota}} \to \Delta^{\mathcal{I}}$ defined by

$$\varphi_\iota(y) = \begin{cases} x & \text{if } y = x_{C_\iota} \\ \varphi(y') & \text{if } y = (y', \iota) \end{cases}$$

for $\iota \in \{1, 2\}$ are homomorphisms from $(\mathcal{I}_{C_1}, x_{C_1})$ to $(\mathcal{I}, x)$ and $(\mathcal{I}_{C_2}, x_{C_2})$ to $(\mathcal{I}, x)$, respectively. Invoking the induction hypothesis, we conclude that $x \in C_1^{\mathcal{I}}$ as well as $x \in C_2^{\mathcal{I}}$ and thus $x \in C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}} = (C_1 \sqcap C_2)^{\mathcal{I}}$.

- Considering $C = \exists r.C_1$, we find that $\varphi' = \varphi|_{\Delta^{\mathcal{I}_{C_1}}}$ is a homomorphism from $(\mathcal{I}_{C_1}, x_{C_1})$ to $(\mathcal{I}, \varphi(x_{C_1}))$. Invoking the induction hypothesis, we conclude $\varphi'(x_{C_1}) = \varphi(x_{C_1}) \in C_1^{\mathcal{I}}$. On the other hand, by construction of $\mathcal{I}_C$ we find $(x_C, x_{C_1}) \in r^{\mathcal{I}_C}$ and thus, since φ is a homomorphism $(x, \varphi(x_{C_1}) = (\varphi(x_C), \varphi(x_{C_1}) \in r^{\mathcal{I}}$. Together, this allows to conclude $x \in (\exists r.C_1)^{\mathcal{I}}$.

We proceed with the only-if direction.

- For $C = \top$, the case is trivial.

- For $C = A \in N_C$, the mapping $\varphi = \{x_A \mapsto x\}$ is the required homomorphism since by assumption it holds that $x \in A^{\mathcal{I}}$.

- For $C = C_1 \sqcap C_2$, we have by assumption $x \in C^{\mathcal{I}} = C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$ therefore $x \in C_1^{\mathcal{I}}$ and $x \in C_2^{\mathcal{I}}$. Invoking the induction hypothesis we find homomorphisms φ_1 from $(\mathcal{I}_{C_1}, x_{C_1})$ to $(\mathcal{I}, x)$ and φ_2 from $(\mathcal{I}_{C_2}, x_{C_2})$ to $(\mathcal{I}, x)$. Consequently, by construction of $\mathcal{I}_C$, the mapping $\varphi : \Delta^{\mathcal{I}_C} to \Delta^{\mathcal{I}}$ defined by

$$\varphi(y) = \begin{cases} x & \text{if } y = x_C \\ \varphi_1(y') & \text{if } y = (y', 1) \\ \varphi_2(y') & \text{if } y = (y', 2) \end{cases}$$

  is a homomorphism from $(\mathcal{I}_C, x_C)$ to $(\mathcal{I}, x)$.

- For $C = \exists r.C_1$, we find by assumption $x \in (\exists r.C_1)^{\mathcal{I}}$ thus there exists an $x' \in \Delta^{\mathcal{I}}$ with $(x, x') \in r^{\mathcal{I}}$ and $x' \in C_1^{\mathcal{I}}$. Invoking the induction hypothesis, we find a homomorphism φ' from $(\mathcal{I}_{C_1}, x_{C_1})$ to $(\mathcal{I}, x')$. Consequently the mapping $\varphi : \Delta^{\mathcal{I}_C} \to \Delta^{\mathcal{I}}$ with $\varphi = \varphi' \cup \{x_C \mapsto x\}$ is a homomorphism from $(\mathcal{I}_C, x_C)$ to $(\mathcal{I}, x)$.

□

Lemma 2. *Let C and C' be two $\mathcal{EL}$ concepts. Then $\emptyset \models C \sqsubseteq C'$ if and only if there is a homomorphism from $(\mathcal{I}'_C, x'_C)$ to $(\mathcal{I}_C, x_C)$.*

Proof. For the if-direction, let φ be a homomorphism from $(\mathcal{I}'_C, x'_C)$ to $(\mathcal{I}_C, x_C)$. Now let $\mathcal{I}$ be an interpretation and pick an arbitrary $x \in \Delta^{\mathcal{I}}$ with $x \in C^{\mathcal{I}}$. By

Lemma 1, there exists a homomorphism φ' from $(\mathcal{I}_C, x_C)$ to $(\mathcal{I}, x)$. Then $\varphi' \circ \varphi$ is a homomorphism from $(\mathcal{I}_{C'}, x_{C'})$ to $(\mathcal{I}, x)$ and by the other direction of Lemma 1, we can conclude $x \in C'$. Thus $C^{\mathcal{I}} \subseteq C'^{\mathcal{I}}$ for all interpretations $\mathcal{I}$ and therefore $\emptyset \models C \sqsubseteq C'$.
For the only-if-direction, assume $\emptyset \models C \sqsubseteq C'$. Now consider the pointed interpretation $(\mathcal{I}_C, x_C)$. As the identity on $\Delta^{\mathcal{I}_C}$ is a homomorphism from $(\mathcal{I}_C, x_C)$ to itself, we use Lemma 1 to conclude $x_C \in C^{\mathcal{I}_C}$. By $\emptyset \models C \sqsubseteq C'$ we can infer that $x_C \in C'^{\mathcal{I}_C}$. Invoking the if-direction of Lemma 1, we find that there must be a homomorphism from $(\mathcal{I}'_C, x'_C)$ to $(\mathcal{I}_C, x_C)$. □

Appendix B. $\mathcal{EL}$ Automata

In this appendix section, we recall core notions on $\mathcal{EL}$ automata [22] before giving the proof of Lemma 4.

Definition 11 [22]. *An $\mathcal{EL}$* automaton (EA) *is a tuple $\mathcal{A} = (Q, P, \Sigma_N, \Sigma_E, \delta)$, where Q is a finite set of* bottom up states, *P is a finite set of* top down states, *$\Sigma_N \subseteq N_C$ is the finite* node alphabet, *$\Sigma_E \subseteq N_R$ is the finite* edge alphabet, *and δ is a set of transitions of the following form:*

$$\begin{array}{rclcrcl}
true & \rightarrow & q & \quad & p & \rightarrow & p_1 \\
A & \rightarrow & q & & p & \rightarrow & \langle r\rangle p1 \\
q_1 \wedge \ldots \wedge q_n & \rightarrow & q & & p & \rightarrow & A \\
\langle r\rangle q1 & \rightarrow & q & & p & \rightarrow & false \\
q & \rightarrow & p & & & &
\end{array} \quad \begin{array}{r}(B.1)\\(B.2)\\(B.3)\\(B.4)\\(B.5)\end{array}$$

where $q, q_1, ..., q_n$ range over Q, p, p_1 range over P, A ranges over Σ_N, and r ranges over Σ_E.

Definition 12 [22]. *Let $\mathcal{I}$ be an interpretation and $\mathcal{A} = (Q, P, \Sigma_N, \Sigma_E, \delta)$ an EA. A* run *of $\mathcal{A}$ on $\mathcal{I}$ is a map $\rho : \delta \rightarrow 2^{Q \cup P}$ such that for all $d \in \Delta^{\mathcal{I}}$, we have:*

1. *if $true \rightarrow q \in \delta$, then $q \in \rho(d)$;*
2. *if $A \rightarrow q \in \delta$, and $d \in A^{\mathcal{I}}$, then $q \in \rho(d)$;*
3. *if $q_1, ..., q_n \in \rho(d)$ and $q_1 \wedge \ldots \wedge q_n \rightarrow q \in \delta$, then $q \in \rho(d)$;*
4. *if $(d, e) \in r^{\mathcal{I}}$, $q_1 \in \rho(e)$ and $\langle r\rangle q_1 \rightarrow q \in \delta$, then $q \in \rho(d)$;*
5. *if $q \in \rho(d)$ and $q \rightarrow p \in \delta$, then $p \in \rho(d)$;*

6. *if* $p \in \rho(d)$ *and* $p \rightarrow p_1 \in \delta$*, then* $p_1 \in \rho(d)$*;*
7. *if* $p \in \rho(d)$ *and* $p \rightarrow \langle r \rangle p_1 \in \delta$*, then there is an* $(d, e) \in r^{\mathcal{I}}$ *with* $p_1 \in \rho(e)$*;*
8. *if* $p \in \rho(d)$ *and* $p \rightarrow A \in \delta$*, then* $d \in A^{\mathcal{I}}$*;*
9. *if* $p \rightarrow \mathit{false} \in \delta$*, then* $p \notin \rho(d)$.

The following Proposition specifies how the corresponding EA $\mathcal{A}$ for any TBox $\mathcal{T}$ can be constructed such that $\mathcal{T}_\Sigma(\mathcal{A}) \equiv_\Sigma^{\mathcal{EL}} \mathcal{T}$ for any Σ.

Construction from Proposition 13 [22] *Let* $\mathcal{T}$ *be a TBox,* $s(\mathcal{T})$ *subconcepts of* $\mathcal{T}$ *and* $\mathcal{A} = (Q, P, \mathit{sig}_C(\mathcal{T}), \mathit{sig}_R(\mathcal{T}), \delta)$ *with* $Q = \{q_C | C \in s(\mathcal{T})\}, P = \{p_C | C \in s(\mathcal{T})\}$ *and* δ *given by*

- $\mathit{true} \rightarrow q_\top$ *if* $\top \in s(\mathcal{T})$*;*
- $A \rightarrow q_A$ *and* $q_A \rightarrow p_A$ *for all* $A \in \mathit{sig}_C(\mathcal{T})$*;*
- $q_C \wedge q_D \rightarrow q_{C \sqcap D}$*;*
- $\langle r \rangle q_C \rightarrow q_{\exists r.C}$ *and* $q_{\exists r.C} \rightarrow \langle r \rangle p_C$ *for all* $\exists r.C \in s(\mathcal{T})$*;*
- $q_C \rightarrow q_D$ *for all* $C, D \in s(\mathcal{T})$ *with* $\mathcal{T} \models C \sqsubseteq D$*;*
- $p_A \rightarrow A$ *for all* $A \in \mathit{sig}_C(\mathcal{T})$*;*
- $p_{\exists r.C} \rightarrow \langle r \rangle p_C$ *for all* $\exists r.C \in s(\mathcal{T})$*;*
- $p_C \rightarrow p_D$ *for all* $C, D \in s(\mathcal{T})$ *with* $\mathcal{T} \models C \sqsubseteq D$*;*
- $p_\bot \rightarrow \mathit{false}$ *if* $\bot \in s(\mathcal{T})$.

An EA $\mathcal{A}$ is said to entail a subsumption $C \sqsubseteq D$ if every model accepted by $\mathcal{A}$ satisfies $C \sqsubseteq D$. Subsequently, an EA $\mathcal{A}$ and a TBox $\mathcal{T}$ are $\mathcal{EL}$ Σ-inseparable, in symbols $\mathcal{A} \equiv_\Sigma^{\mathcal{EL}} \mathcal{T}$, if $\mathcal{A} \models C \sqsubseteq D$ iff $\mathcal{T} \models C \sqsubseteq D$ for all $\mathcal{EL}$ Σ-inclusions $C \sqsubseteq D$. Further, for a signature Σ, $\mathcal{T}_\Sigma(\mathcal{A}) = \{C \sqsubseteq D \mid \mathcal{A} \models C \sqsubseteq D, \mathrm{sig}(C) \cup \mathrm{sig}(D) \subseteq \Sigma\}$. For a natural number m, $\mathcal{T}_\Sigma^m(\mathcal{A}) = \{C \sqsubseteq D \mid C \sqsubseteq D \in \mathcal{T}_\Sigma(\mathcal{A}), d(C) \leq m$ and $d(D) \leq m\}$.

Excerpt from Lemma 55 [22]. *Let* $\mathcal{A}$ *be an EA and* $M_\mathcal{A} = 2^{|P \cup Q|}$*. The following conditions are equivalent:*

1. *There exists* $k > M_\mathcal{A}^2 + 1$ *such that* $\mathcal{T}_\Sigma^{M_\mathcal{A}^2+1} \not\models \mathcal{T}_\Sigma^k$*;*
4. *There does not exists an* $\mathcal{EL}$ *TBox* $\mathcal{T}$ *with* $\mathcal{A} \equiv_\Sigma^{\mathcal{EL}} \mathcal{T}$.

Lemma 4. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, Σ a signature. The following statements are equivalent:*

1. *There exists a uniform $\mathcal{EL}$ Σ-interpolant of $\mathcal{T}$.*
2. *There exists a uniform $\mathcal{EL}$ Σ-interpolant $\mathcal{T}'$ of $\mathcal{T}$ for which it holds that $d(\mathcal{T}') \leq 2^{4\cdot(|sub(\mathcal{T})|)} + 1$.*

Proof. Assume that a uniform $\mathcal{EL}$ Σ-interpolant of $\mathcal{T}$ exists and let $M = 2^{(2\cdot|\mathrm{sub}(\mathcal{T})|)}$. Then, by Lemma 55 [22], there is no $k > M^2+1$ such that $\mathcal{T}_\Sigma^{M^2+1}(\mathcal{A}) \not\models \mathcal{T}_\Sigma^{k}(\mathcal{A})$, where $\mathcal{A}$ is the corresponding $\mathcal{EL}$ automaton for $\mathcal{T}$. Then $\mathcal{T}_\Sigma^{M^2+1}(\mathcal{A}) \models \mathcal{T}_\Sigma(\mathcal{A})$. Therefore, $\mathcal{T}_\Sigma^{M^2+1}(\mathcal{A}) \equiv_\Sigma^{\mathcal{EL}} \mathcal{T}$, i.e., $\mathcal{T}_\Sigma^{M^2+1}(\mathcal{A})$ is a uniform $\mathcal{EL}$ Σ-interpolant $\mathcal{T}'$ of $\mathcal{T}$ with $d(\mathcal{T}') \leq M^2+1$. We can replace M^2+1 by $2^{4\cdot(|\mathrm{sub}(\mathcal{T})|)}+1$ and obtain $d(\mathcal{T}') \leq 2^{4\cdot(|\mathrm{sub}(\mathcal{T})|)} + 1$. □

References

[1] N. Nikitina, S. Rudolph, ExpExpExplosion: Uniform interpolation in general $\mathcal{EL}$ terminologies, in: Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012), pp. 618–623. (Shortlisted for best student paper award).

[2] W3C OWL Working Group, OWL 2 Web Ontology Language: Document Overview, W3C Recommendation, 27 October 2009. Available at `http://www.w3.org/TR/owl2-overview/`.

[3] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P. Patel-Schneider (Eds.), The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, second edition, 2007.

[4] S. Rudolph, Foundations of Description Logics, in: Reasoning Web. Semantic Technologies for the Web of Data - 7th International Summer School 2011, Tutorial Lectures, pp. 76–136.

[5] E. Sirin, B. Parsia, B. Cuenca Grau, A. Kalyanpur, Y. Katz, Pellet: A practical OWL-DL reasoner, Web Semantics: Science, Services and Agents on the World Wide Web 5 (2007) pp. 51–53.

[6] R. Shearer, B. Motik, I. Horrocks, HermiT: A Highly-Efficient OWL Reasoner, in: Proceedings of the 5th International Workshop on OWL: Experiences and Directions (OWLED 2008).

[7] D. Tsarkov, I. Horrocks, FACT++ Description Logic Reasoner: System Description, in: Proceedings of the 3rd International Joint Conference on Automated Reasoning (IJCAR 2006), pp. 292–297.

[8] V. Haarslev, K. Hidde, R. Möller, M. Wessel, The RacerPro knowledge representation and reasoning system, Semantic Web 3 (2012) pp. 267–277.

[9] B. Motik, B. Cuenca Grau, I. Horrocks, Z. Wu, A. Fokoue, C. Lutz (Eds.), OWL 2 Web Ontology Language: Profiles, W3C Recommendation, 27 October 2009. Available at `http://www.w3.org/TR/owl2-profiles/`.

[10] S. Brandt, Polynomial Time Reasoning in a Description Logic with Existential Restrictions, GCI Axioms, and - What Else?, in: Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2004), pp. 298–302.

[11] F. Baader, S. Brandt, C. Lutz, Pushing the $\mathcal{EL}$ envelope, in: Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005), pp. 364–369.

[12] M. Ashburner, Gene ontology: Tool for the unification of biology, Nature Genetics 25 (2000) pp.25–29.

[13] N. Sioutos, S. d. Coronado, M. W. Haber, F. W. Hartel, W.-L. Shaiu, L. W. Wright, NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information, Journal of Biomedical Informatics 40 (2007) pp.30–43.

[14] K. A. Spackman, K. E. Campbell, R. A. Cote, SNOMED RT: A reference terminology for health care, in: Proceedings of the AIMA Fall Symposium, pp. 640–644.

[15] N. Nikitina, S. Rudolph, B. Glimm, Reasoning-supported interactive revision of knowledge bases, in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011), pp. 1027–1032.

[16] Z. Wang, K. Wang, R. W. Topor, J. Z. Pan, Forgetting Concepts in DL-Lite, in: Proceedings of the 5th European Semantic Web Conference (ESWC 2008), pp. 245–257.

[17] R. Kontchakov, F. Wolter, M. Zakharyaschev, Logic-based ontology comparison and module extraction, with an application to DL-Lite, Artificial Intelligence 174 (2010) pp.1093–1141.

[18] C. Lutz, F. Wolter, Foundations for uniform interpolation and forgetting in expressive description logics, in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011), pp. 989–995.

[19] P. Koopmann, R. A. Schmidt, Forgetting concept and role symbols in $\mathcal{ALCH}$-ontologies, in: Proceedings of the 19th International Conference on Logic for Programming Artificial Intelligence and Reasoning (LPAR 2013), volume 8312 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 552–567.

[20] M. Ludwig, B. Konev, Practical uniform interpolation and forgetting for alc tboxes with applications to logical difference, in: Proceedings of the

14th International Conference on Principles of Knowledge Representation and Reasoning (KR'14), AAAI Press, 2014. To appear.

[21] B. Konev, D. Walther, F. Wolter, Forgetting and uniform interpolation in large-scale description logic terminologies, in: Proceedings of the 21st International Joint Conference on Artifical Intelligence (IJCAI 2009), pp. 830–835.

[22] C. Lutz, I. Seylan, F. Wolter, An Automata-Theoretic Approach to Uniform Interpolation and Approximation in the Description Logic $\mathcal{EL}$, in: Proceedings of the 13th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2012).

[23] N. Nikitina, B. Glimm, Hitting the sweetspot: Economic rewriting of knowledge bases, in: Proceedings of the 11th International Semantic Web Conference (ISWC 2012), pp. 394–409.

[24] M. Hofmann, Proof-theoretic approach to description-logic, in: Proceedings of the 20th Annual IEEE Symposium on Logic in Computer Science (LICS 2005), pp. 229–237.

[25] Y. Kazakov, Consequence-driven reasoning for $\mathcal{H}$orn $\mathcal{SHIQ}$ ontologies, in: Proceedings of the 21st International Joint Conference on Artifical Intelligence (IJCAI 2009), pp. 2040–2045.

[26] C. Lutz, F. Wolter, Conservative extensions in the lightweight description logic EL, in: In Proceedings of the 21st International Conference on Automated Deduction (CADE 2007), pp. 84–99.

[27] C. Lutz, F. Wolter, Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$, Journal of Symbolic Computation 45 (2010) pp.194–228.

[28] H. Comon, F. Jacquemard, M. Dauchet, R. Gilleron, D. Lugiez, C. Loding, S. Tison, M. Tommasi, Tree automata techniques and applications, 2008.

[29] Z. Wang, K. Wang, R. Topor, J. Z. Pan, G. Antoniou, Uniform interpolation for $\mathcal{ALC}$ revisited, in: Proceedings of the 22nd Australasian Joint Conference on Advances in Artificial Intelligence (AI 2009), pp. 528–537.

[30] Z. Wang, K. Wang, R. Topor, X. Zhang, Tableau-based forgetting in $\mathcal{ALC}$ ontologies, in: Proceedings of the 19th European Conference on Artificial Intelligence (ECAI 2010), pp. 47–52.

[31] B. Cuenca Grau, I. Horrocks, Y. Kazakov, U. Sattler, Extracting modules from ontologies: A logic-based approach, in: Proceedings of OWL: Experiences and Directions (OWLED 2007), pp. 159–186.

[32] B. Konev, C. Lutz, D. Walther, F. Wolter, Semantic modularity and module extraction in description logics, in: Proceedings of the 18th European Conference on Artificial Intelligence (ECAI 2008), pp. 55–59.

[33] B. Cuenca Grau, B. Parsia, E. Sirin, A. Kalyanpur, Modularity and web ontologies, in: Proceedings of the 10th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2006), pp. 198–209.

[34] B. Konev, C. Lutz, D. Ponomaryov, F. Wolter, Decomposing description logic ontologies, in: Proceedings of the 12th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010), pp. 236–246.

[35] C. Del Vescovo, B. Parsia, U. Sattler, T. Schneider, The modular structure of an ontology: Atomic decomposition, in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011), pp. 2232–2237.

[36] S. Ghilardi, C. Lutz, F. Wolter, Did i damage my ontology? a case for conservative extensions in description logics, in: Proceedings of the 10th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2006), pp. 187–197.

[37] C. Lutz, D. Walther, F. Wolter, Conservative extensions in expressive description logics, in: Proceedings of the 20th International Joint Conference on Artifical Intelligence (IJCAI 2007), pp. 453–458.

[38] B. Konev, C. Lutz, D. Walther, F. Wolter, Formal properties of modularisation, in: H. Stuckenschmidt, C. Parent, S. Spaccapietra (Eds.), Modular Ontologies, Springer-Verlag, 2009, pp. 25–66.

[39] R. Kontchakov, F. Wolter, M. Zakharyaschev, Can you tell the difference between DL-Lite ontologies?, in: Proceedings of the 11th International

Conference on Principles of Knowledge Representation and Reasoning (KR 2008), pp. 285–295.

[40] B. Cuenca Grau, I. Horrocks, Y. Kazakov, U. Sattler, A logical framework for modularity of ontologies, in: Proceedings of the 20th International Joint Conference on Artifical Intelligence (IJCAI 2007), pp. 298–304.

[41] N. Nikitina, S. Schewe, Simplifying Description Logic Ontologies, in: Proceedings of the 12th International Semantic Web Conference (ISWC 2013), pp. 411–426.

[42] N. Nikitina, S. Schewe, More is Sometimes Less: Succinctness in $\mathcal{EL}$, in: Proceedings of the 26th International Workshop on Description Logics (DL 2013), pp. 403–414.