Expressing DL-Lite Ontologies with Controlled English

Camilo Thorne

Co-Authors: Raffaella Bernardi, Diego Calvanese

Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen · Bolzano

DL 07 Workshop, Brixen-Bressanone. June 8, 2007

The Problem

- **Our aim:** To provide natural language (NL) interfaces to efficient ontology based data access systems, to help unskilled users.
- **The background:** Efficient ontology based data access is formally modelled by by the entailment problem:

$$\langle \mathcal{T}, \mathcal{D} \rangle \models q(\vec{c})$$

  where the tuple $\langle \mathcal{T}, \mathcal{D} \rangle$ is a DL-Lite knowledge base (KB) and $q(\vec{c})$ the grounding of a conjunctive query. This problem is known as the query answering problem (QA) for DL-Lite KBs.
- **The questions:** Which NL constructs, i.e., which fragment of English, can capture DL-Lite KBs? Which is the expressive power of DL-Lite w.r.t. other fragments of English?

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

- We have chosen DL-Lite$_{R,\sqcap}$ because of its good computational properties and its closeness to NL when it comes to ontology based data access.
- DL-Lite$_{R,\sqcap}$ **knowledge base** KB $= \langle \mathcal{T}, \mathcal{D} \rangle$ is a set of TBox ($\mathcal{T}$) and ABox ($\mathcal{D}$) **assertions**:

$$A(c), R(c_1, c_2) \qquad \text{(ABox assertions)}$$
$$C_l \sqsubseteq C_r \qquad \text{(TBox concept inclusion assertions)}$$
$$R_1 \sqsubseteq R_2 \qquad \text{(TBox role inclusion assertions)}$$

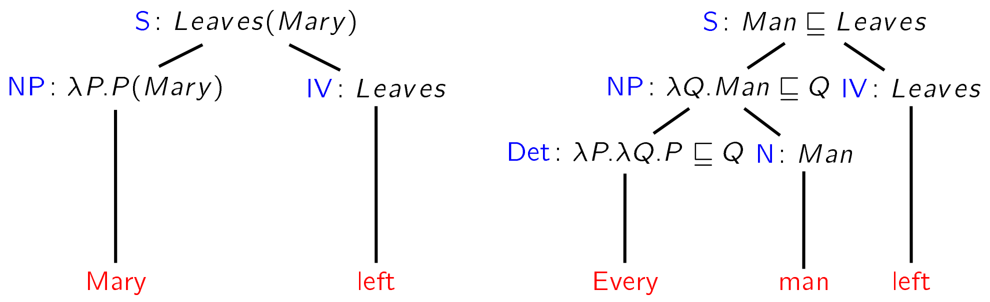
- DL-Lite$_{R,\sqcap}$ left (C_l) and right (C_r) hand side **concepts** are then defined as follows:

$$C_l ::= A \mid \exists R \mid \exists R^- \mid C_l \sqcap C_l.$$
$$C_r ::= C_l \mid \neg A \mid \neg \exists R \mid \neg \exists R^- \mid \exists R.C_r \mid \exists R^-.C_r \mid C_r \sqcap C_r.$$

- DL-Lite$_{R,\sqcap}$ is a fragment of FOL included in the $\forall\exists^*$ prefix class. It is equipped with standard FOL semantics. TBox reasoning is in **P** and QA in **LOGSPACE** w.r.t. **data complexity** (cf. Calvanese *et al.*, 2005 and 2006).

Fragments of English

▷ A **fragment of English** F is a subset of English, defined by a restricted set of grammar rules, that maps, through a compositional translation, onto a fragment of FOL (to which it can be identified) its set Λ_F of **meaning representations** (MRs):



(ABox) (TBox)

▷ Pratt and Third (2005) define a family of fragments of English and study them w.r.t. the **expressive power** of their MRs and the **computational complexity** of satisfiability and entailment among these MRs.

Building the Fragments - COP

- Given the following lexicon:

Copula	=	is a		
Negation	=	is not a		
Determiners/Quantifiers	=	some	every	no
Nouns	=	man, mortal	...	
Proper Names	=	Socrates	...	

- we can build sentences of the following form:

  Every man is a mortal

  Socrates is a man

  Socrates is a mortal.

- We have thus captured the entailment:

$$\{\forall x(Man(x) \rightarrow Mortal(x)), Man(Socrates)\} \models Mortal(Socrates)$$

- The **basic fragment** of sentences built out of this lexicon is called **COP**.

Building the Fragments

- New fragments are built by extending COP with:

  TV = **Transitive verbs**, e.g. "reads" (**binary relations**).

  DTV = **Distransitive verbs**, e.g., "gives" (**ternary relations**).

  REL = **Relative pronouns**, e.g., "who", "that", "which", etc. (**conjunction**).

- which yield new FOL sets of MRs:

Sentence	**MR**	**Fragment**
Every man **who** is not dead is alive	$\forall x(Man(x) \wedge \neg Dead(x) \rightarrow Alive(x))$	COP+**REL**
Every soldier **defends** a country	$\forall x((Soldier(x) \rightarrow \exists y(Country(y) \wedge \mathbf{Defends}(\mathbf{x}, \mathbf{y})))$.	COP+**TV**
Every salesman **sells** some merchandise to some customer	$\forall x(Salesman(x) \rightarrow \exists y(Customer(y) \wedge \exists z(Merchandise(z) \wedge \mathbf{Sells}(\mathbf{x}, \mathbf{y}, \mathbf{z})))$.	COP+**DTV**

Complexity of Fragments (Pratt & Third 2005)

Fragment	Decision class for satisfiability
COP	**P**
COP+TV+DTV	**P**
COP+REL	**NP-Complete**
COP+REL+TV	**EXPTIME-Complete**
COP+REL+TV+DTV	**NEXPTIME-Complete**

▷ Only the first two are tractable.

▷ The addition of **relatives** is computationally expensive. Relatives in NL convey boolean **conjunction**. Hence, when combined with boolean **negation**, as in COP+REL, we get a fragment that is propositionally complete and therefore intractable.

Lite English

- **Lite English** (Bernardi, Calvanese and Thorne, 2007) is a fragment of English that compositionally translates into DL-Lite$_{R,\sqcap}$ KB assertions. Utterances respect the pattern:

  Det N VP

- Dets are universal quantifiers (conveying subsumption), the Ns a C_l concept and the VPs a C_r concept, for instance:

Category	**DL-Lite$_{R,\sqcap}$**	**Lite English**	**Semantics**
Det		every	$\lambda P.\lambda Q.P \sqsubseteq Q$
N	C_l concepts	man **that** sleeps	$Man \sqcap Sleeps$
VP	C_r concepts	loves some woman **who** works	$\exists Loves.Woman \sqcap Works$

Sentence	**MR**
Every man **that** sleeps loves some woman **who** works	$Man \sqcap Sleeps \sqsubseteq \exists Loves.Woman \sqcap Works$

- The fact that DL-Lite$_{R,\sqcap}$ has full-boolean conjunction allows for restricted **relative** pronouns. **Negation** is restricted to VPs. **Quantifiers** occur in a fixed order.

DL-Lite$_{R,\sqcap}$ is as Expressive as COP

- **THEOREM** DL-Lite$_{R,\sqcap}$ is as expressive as COP.

  (i) COP expresses containment among sets. DL-Lite$_{R,\sqcap}$ can express this via inclusion assertions of the form $A \sqsubseteq B$.

  (ii) COP expresses disjointness among sets. DL-Lite$_{R,\sqcap}$ can express this via disjointness assertions of the form $A \sqsubseteq \neg B$.

  (iii) COP expresses that an individual belongs to a set or its complement. DL-Lite$_{R,\sqcap}$ can express this via disjointness assertions and ABox assertions.

  (iv) COP expresses that the intersection of two sets is not empty. We can express this in DL-Lite$_{R,\sqcap}$ by the ABox assertions $P(c)$ and $Q(c)$ and by dropping the unique name assumption (UNA).

- But the converse is false.

DL-Lite$_{R,\sqcap}$ Overlaps with COP+TV+DTV

▷ **THEOREM** DL-Lite$_{R,\sqcap}$ overlaps in expressive power with COP+TV+DTV, but neither contains the other.

(i) COP+TV+DTV and DL-Lite$_{R,\sqcap}$ contain COP, whence the overlapping.

(ii) (⇒) DL-Lite$_{R,\sqcap}$ role-typing assertions $\exists R \sqsubseteq A$, yield in FOL, when prenexed, skolemized and clausified:

$$\neg R(x, y) \vee A(x).$$

These clauses lie, however, beyond COP+TV+DTV MRs (Pratt and Third 2005, condition P3).

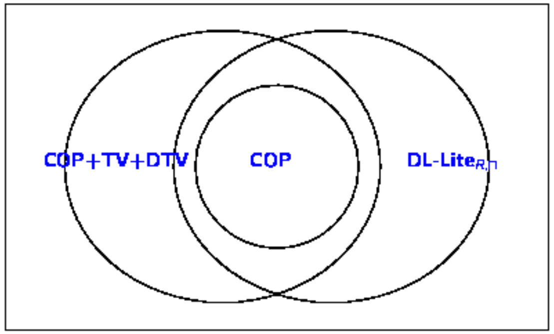
DL-Lite$_{R,\sqcap}$ Overlaps with COP+TV+DTV

(ii) ($\Leftarrow$) The COP+TV sentence $\psi = \exists x \forall y (P(x) \wedge Q(x) \rightarrow R(x,y))$ that says "there is a least element" is not closed under **union of chains**. We can build a non well-founded model $\mathcal{M}_\omega$ from the chain $\mathcal{M}_1 \preccurlyeq \mathcal{M}_2 \preccurlyeq \ldots \preccurlyeq \mathcal{M}_i \preccurlyeq \ldots$, for $i \in \mathbb{N}^+$, of models of ψ:

$$
\begin{array}{llll}
\mathcal{M}_1 : & {}_0 \longrightarrow {}_{+n} & (\mathbb{N}) & \mathcal{M}_1 \models \psi \\
\mathcal{M}_2 : & {}_{-1} \longrightarrow {}_{+n} & (\mathbb{N} \cup \{-1\}) & \mathcal{M}_2 \models \psi \\
\vdots & \vdots & \vdots & \vdots \\
\mathcal{M}_i : & {}_{-(i-1)} \longrightarrow {}_{+n} & (\mathbb{N} \cup \{-1, \ldots, -(i-1)\}) & \mathcal{M}_i \models \psi \\
\vdots & \vdots & \vdots & \vdots \\
\mathcal{M}_\omega = \bigcup_{i \in \mathbb{N}} \mathcal{M}_i : & {}_{-n} \longleftrightarrow {}_{+n} & (\mathbb{Z}) & \mathcal{M}_\omega \not\models \psi
\end{array}
$$

But DL-Lite$_{R,\sqcap}$ is closed under this property, since all of its assertions belong (in FOL) to the $\forall\exists^*$ prefix class (Cori and Lascar 2003).

Summarizing



- ▷ Lite English contains **relatives**, which yield an exponential blowup in Pratt and Third's fragments.
- ▷ But in Lite English relatives are harmless to computational complexity, due to the **controlled behaviour of negation**, which cannot occur in N position (only in VPs). In DL-Lite$_{R,\sqcap}$ negation is limited to the right hand side of $\sqsubseteq$ (i.e. to C_r concepts).

An Example - COP (1)

Syntax Rules	**MR** $(= \phi)$
IP $\to$ **NP I′**	$\phi(\mathbf{NP})(\phi(\mathbf{I'})) \rhd_\beta \phi(\mathbf{IP})$
I′ $\to$ is a **N**	$\phi(\mathbf{I}) = \phi(\mathbf{N})$
I′ $\to$ is not a **N**	$\phi(\mathbf{I}) = \neg\phi(\mathbf{N})$
NP $\to$ **PropN**	$\phi(\mathbf{NP}) = \phi(\mathbf{PropN})$
NP $\to$ **Det N**	$\phi(\mathbf{Det})(\phi(\mathbf{N})) \rhd_\beta \phi(\mathbf{NP})$

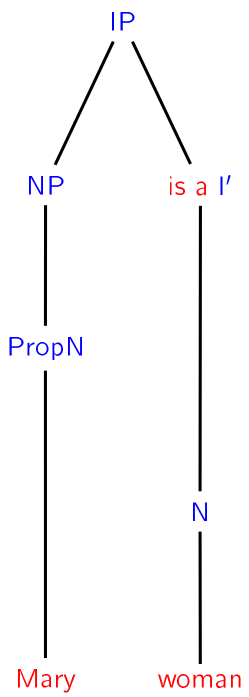
▷ ϕ is a compositional translation.

An Example - COP (2)

Lexicon	**MR** $(= \phi)$
N $\to$ woman	$\phi(\textbf{N}) = \lambda x.Woman(x)$
N $\to$ man	$\phi(\textbf{N}) = \lambda x.Man(x)$
N $\to$ human	$\phi(\textbf{N}) = \lambda y.Human(x)$
PropN $\to$ Mary	$\phi(\textbf{PropN}) = \lambda P.P(Mary)$
Det $\to$ every	$\phi(\textbf{Det}) = \lambda P.\lambda Q.\forall x(P(x) \to Q(x))$
Det $\to$ no	$\phi(\textbf{Det}) = \lambda P.\lambda Q.\forall x(P(x) \to \neg Q(x)$
Det $\to$ some	$\phi(\textbf{Det}) = \lambda P.\lambda Q.\exists x(P(x) \wedge Q(x))$

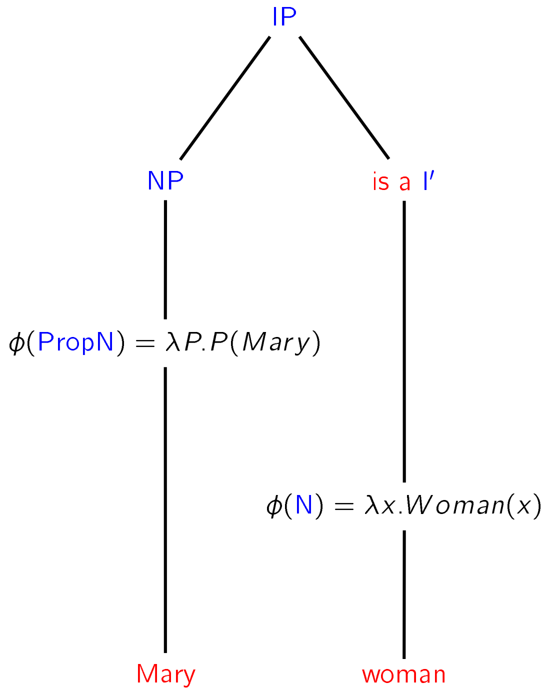
- ϕ is a compositional translation.

An Example - COP (3)



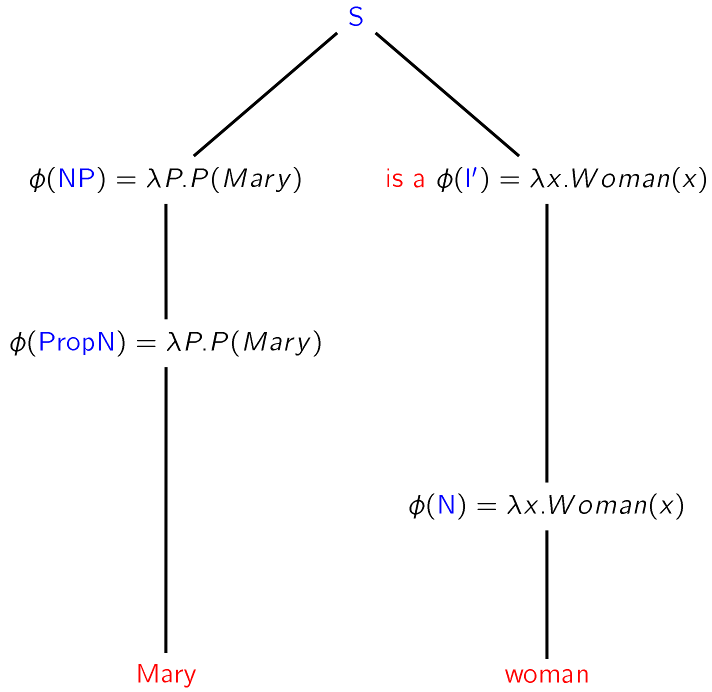
▷ ϕ is computed bottom-up, from leaves to root.

An Example - COP (4)



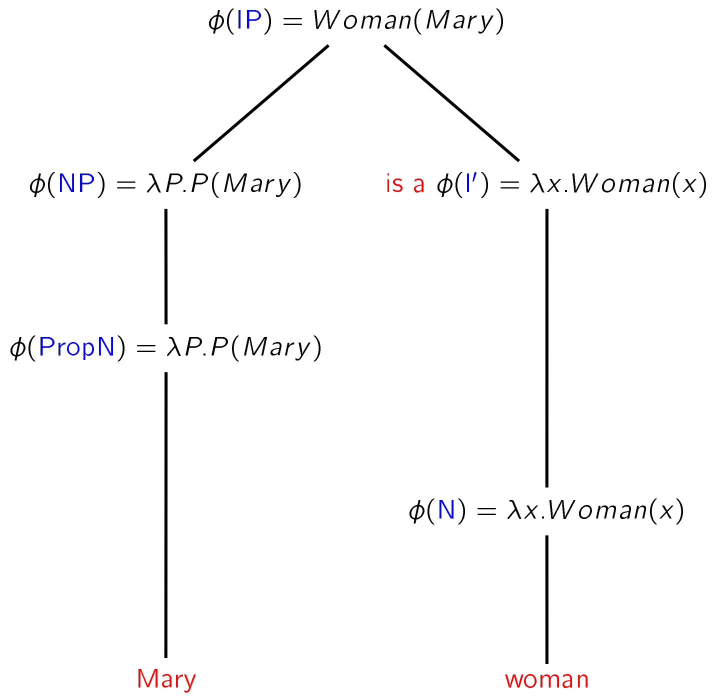
▷ ϕ is computed bottom-up, from leaves to root.

An Example - COP (5)



- ϕ is computed bottom-up, from leaves to root.

An Example - COP (6)



- ϕ is computed bottom-up, from leaves to root.