Expressing Conjunctive and Aggregate Queries Over DL-Lite Ontologies with Controlled English

Camilo Thorne

KRDB Centre - Free University of Bozen-Bolzano

cthorne@inf.unibz.it

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

KRDB Research Centre
Knowledge Representation meets DataBases

Expressing QA

- We want to measure the difficulty of accessing data stored in knowledge bases and in ontology-mediated database systems with controlled English. Standard approaches (Fuchs et al. [2005]) do not deal with this problem.
- We believe this can be done by expressing *query answering* (QA) with controlled English.
- The decision problem QA is conveyed by the *entailment* $\mathcal{O}, \mathcal{D} \models q(\overline{\mathbf{c}})$ where $\langle \mathcal{O}, \mathcal{D} \rangle$ is a description logic knowledge base (KB) and $q(\overline{\mathbf{c}})$ is the grounding (by a sequence $\overline{\mathbf{c}}$ of n constants) of a formal query of arity n.
- We want to express KBs and queries for which QA's computational complexity is tractable (Calvanese et al. [2007]): DL-Lite$_{R,\sqcap}$ and *graph-shaped conjunctive queries* with (AGCQs) and without (GCQs) aggregations. To do this, we need to define controlled fragments of English and a compositional translation $[\![.]\!]$ that maps declarative sentence into KB assertions and interrogative sentences into queries.

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

Definition 1. Let $\mathcal{P} = \{P_i | i \in \mathbb{N}\}$ and $\mathcal{R} = \{R_i | i \in \mathbb{N}\}$ be two countable sets of primitive concept and role symbols. DL-Lite$_{R,\sqcap}$ *left hand side concepts* $\mathcal{C}_l$ and *right hand side concepts* $\mathcal{C}_r$ are defined as follows:

$$\mathcal{C}_l \ ::= \ \mathcal{P} \mid \exists \mathcal{R} \mid \exists \mathcal{R}^- \mid \mathcal{C}_l \sqcap \mathcal{C}_l.$$
$$\mathcal{C}_r \ ::= \ \neg \mathcal{P} \mid \neg \exists \mathcal{R} \mid \neg \exists \mathcal{R}^- \mid \mathcal{C}_l \mid \mathcal{C}_r \sqcap \mathcal{C}_r \mid \exists \mathcal{R} \colon \mathcal{C}_r \mid \exists \mathcal{R}^- \colon \mathcal{C}_r.$$

Let $\mathcal{C}_o = \{c_i | i \in \mathbb{N}\}$ be a set of constants. DL-Lite$_{R,\sqcap}$ Abox *facts* $\mathcal{D}$ and TBox *terminological assertions* $\mathcal{D}$ are defined as follows:

$$\mathcal{D} \ ::= \ \mathcal{P}(\mathcal{C}_o) \mid \mathcal{R}(\mathcal{C}_o, \mathcal{C}_o) \quad \text{(ABox)}$$
$$\mathcal{O} \ ::= \ \mathcal{C}_l \sqsubseteq \mathcal{C}_r \mid \mathcal{R} \sqsubseteq \mathcal{R} \quad \text{(TBox)}$$

Theorem 1 (Bernardi et al. [2007])**.** *For every sentence S in the* CL Lite-English*, there exists a* DL-Lite$_{R,\sqcap}$ *assertion α s.t. $[\![S]\!] = \alpha$. Conversely, every* DL-Lite$_{R,\sqcap}$ *assertion α is the image by $[\![.]\!]$ of some sentence S lying within* Lite-English*.*

TBox $\mathcal{O}$	**ABox $\mathcal{D}$**
$Man \sqsubseteq Person$	$hasHeight(James, 180)$
$Woman \sqsubseteq Person$	$hasHeight(Mary, 171)$
$Man \sqsubseteq \neg Woman$	$Woman(Mary)$
$Person \sqsubseteq Mortal$	$Man(James)$
$Man \sqsubseteq \exists Loves \colon Woman$	$Odd(171)$
$Person \sqsubseteq \exists hasHeight \colon Height$	$Loves(James, Mary)$
$Height \sqsubseteq Integer$	
$Odd \sqsubseteq Integer$	
$Even \sqsubseteq Integer$	
$Odd \sqsubseteq \neg Even$	

GCQ-English

- GCQ-English is the controlled language that expresses **GCQ**s.
- It is generated by a phrase-structure grammar with *semantic actions*.
- Semantic actions define, by recursion on the syntactic components of an English utterance, a compositional translation $[\![.]\!]$.
- A *compositional translation* is a function that homomorphically maps a fragment of a natural language (Montague [1970]) into, ultimately, first order logic (FOL), by mapping syntactic components into λ-FOL formulas, that is, FOL augmented with the types, the function application, the λ-abstraction and the β-reduction of the symply typed lambda calculus.
- We allow two big classes of questions: (i) Y/N-questions, that are mapped into boolean GCQs and **(ii)** Wh-questions, that are mapped into non-boolean GCQs.

Grammar rules	**Semantic actions**
$\mathbf{Q}_{wh} \rightarrow$ Intpro $\mathbf{N}_i \ \mathbf{S}_{gap_i}?$	$[\![\mathbf{Q}_{wh}]\!] := [\![\mathbf{Intpro}]\!]([\![\mathbf{N}_i]\!])([\![\mathbf{S}_{gap_i}]\!])$
$\mathbf{Q}_{wh} \rightarrow$ Intpro$_i$ $\mathbf{S}_{gap_i}?$	$[\![\mathbf{Q}_{wh}]\!] := [\![\mathbf{Intpro}_i]\!]([\![\mathbf{S}_{gap_i}?]\!])$
$\mathbf{Q}_{Y/N} \rightarrow$ does $\mathbf{NP}_i^-$ $\mathbf{VP}_i^-?$	$[\![\mathbf{Q}_{Y/N}]\!] := [\![\mathbf{NP}_i^-]\!]([\![\mathbf{VP}_i^-]\!])$
$\mathbf{Q}_{Y/N} \rightarrow$ is $\mathbf{NP}_i$ $\mathbf{VP}_i?$	$[\![\mathbf{Q}_{Y/N}]\!] := [\![\mathbf{NP}_i]\!]([\![\mathbf{VP}_i]\!])$
$\mathbf{S}_{gap_i} \rightarrow \mathbf{NP}_{gap_i}$ $\mathbf{VP}_i$	$[\![\mathbf{S}_{gap_i}]\!] := [\![\mathbf{NP}_{gap_i}]\!]([\![\mathbf{VP}_i]\!])$
$\mathbf{VP}_i \rightarrow \mathbf{VP}_i$ Coord $\mathbf{VP}_i$	$[\![\mathbf{VP}]\!] := [\![\mathbf{Coord}]\!]([\![\mathbf{VP}]\!])([\![\mathbf{VP}]\!])$
$\mathbf{VP}_i^- \rightarrow \mathbf{VP}_i^-$ Coord $\mathbf{VP}_i^-$	$[\![\mathbf{VP}_i^-]\!] := [\![\mathbf{Coord}]\!]([\![\mathbf{VP}^-]\!])([\![\mathbf{VP}_i^-]\!])$
$\mathbf{VP}_i \rightarrow \mathbf{TV}_{i,j}$ $\mathbf{NP}_j$	$[\![\mathbf{VP}_i]\!] := [\![\mathbf{TV}_{i,j}]\!]([\![\mathbf{NP}_j]\!])$
$\mathbf{VP}_i \rightarrow$ is $\mathbf{Adj}_i$	$[\![\mathbf{VP}_i]\!] := [\![\mathbf{Adj}_i]\!]$
$\mathbf{VP}_i \rightarrow$ is a $\mathbf{N}_i$	$[\![\mathbf{VP}_i]\!] := [\![\mathbf{N}_i]\!]$
$\mathbf{VP}_i^- \rightarrow \mathbf{IV}_i^-$	$[\![\mathbf{VP}_i^-]\!] := [\![\mathbf{IV}_i^-]\!]$
$\mathbf{VP}_i \rightarrow \mathbf{IV}_i$	$[\![\mathbf{VP}_i]\!] := [\![\mathbf{IV}_i]\!]$
$\mathbf{VP}_i^- \rightarrow \mathbf{TV}_{i,j}^-$ $\mathbf{NP}_j$	$[\![\mathbf{VP}_i^-]\!] := [\![\mathbf{TV}_{i,j}^-]\!]([\![\mathbf{NP}_j]\!])$
$\mathbf{VP}_i \rightarrow \mathbf{VP}_i^p$	$[\![\mathbf{VP}_i]\!] := [\![\mathbf{VP}_i^p]\!]$
$\mathbf{VP}_i^p \rightarrow \mathbf{TV}_{i,j}^p$ $\mathbf{NP}_j$	$[\![\mathbf{VP}_i^p]\!] := [\![\mathbf{TV}_{i,j}^p]\!]([\![\mathbf{NP}_j]\!])$
$\mathbf{NP}_i^- \rightarrow \mathbf{Det}^-$ $\mathbf{N}_i$	$[\![\mathbf{NP}_i^-]\!] := [\![\mathbf{Det}^-]\!]([\![\mathbf{N}_i]\!])$
$\mathbf{NP}_i \rightarrow \mathbf{Pro}_i$	$[\![\mathbf{NP}_i]\!] := [\![\mathbf{Pro}_i]\!]$
$\mathbf{NP}_i \rightarrow \mathbf{Det}$ $\mathbf{N}_i$	$[\![\mathbf{NP}_i]\!] := [\![\mathbf{Det}]\!]([\![\mathbf{N}]\!])$
$\mathbf{NP}_i \rightarrow \mathbf{Pn}_i$	$[\![\mathbf{NP}_i]\!] := [\![\mathbf{Pn}_i]\!]$
$\mathbf{NP}_i \rightarrow \mathbf{Pro}_i$	$[\![\mathbf{NP}_i]\!] := [\![\mathbf{Pro}_i]\!]$
$\mathbf{N}_i \rightarrow \mathbf{Adj}$ $\mathbf{N}_i$	$[\![\mathbf{N}_i]\!] := [\![\mathbf{Adj}]\!]([\![\mathbf{N}_i]\!])$
$\mathbf{N}_i \rightarrow \mathbf{N}_i$ $\mathbf{Relpro}_i$ $\mathbf{S}_{gap_i}$	$[\![\mathbf{N}_i]\!] := [\![\mathbf{Relpro}_i]\!]([\![\mathbf{N}_i]\!])([\![\mathbf{S}_{gap_i}]\!])$

Lexicon	**Semantic actions**
$\mathbf{Det} \rightarrow$ some	$\lambda P.\lambda Q.\exists x(P(x) \wedge Q(x)) \colon (e \rightarrow t) \rightarrow ((e \rightarrow t) \rightarrow (e \rightarrow t))$
$\mathbf{Pro}_i \rightarrow$ somebody	$\lambda P.\exists x P(x) \colon (e \rightarrow t) \rightarrow t$
$\mathbf{Pro}_i^- \rightarrow$ anybody	$\lambda P.\exists x P(x) \colon (e \rightarrow t) \rightarrow t$
$\mathbf{Coord} \rightarrow$ and	$\lambda P.\lambda Q.\exists x(P(x) \wedge Q(x)) \colon (e \rightarrow t) \rightarrow ((e \rightarrow t) \rightarrow (e \rightarrow t))$
$\mathbf{Relpro}_i \rightarrow$ who	$\lambda P.\lambda x.P(x) \colon (e \rightarrow t) \rightarrow (e \rightarrow t)$
$\mathbf{Pro}_i \rightarrow$ him	$\lambda P.P(x) \colon (e \rightarrow t) \rightarrow t$
$\mathbf{Pro}_i \rightarrow$ himself	$\lambda P.P(x) \colon (e \rightarrow t) \rightarrow t$
$\mathbf{Intpro} \rightarrow$ which	$\lambda P.\lambda Q.\lambda x.P(x) \wedge Q(x) \colon (e \rightarrow t) \rightarrow (e \rightarrow t)$
$\mathbf{Intpro}_i \rightarrow$ who$_i$	$\lambda P.\lambda x.P(x) \colon (e \rightarrow t) \rightarrow (e \rightarrow t)$
$\mathbf{NP}_{gap_i} \rightarrow \epsilon$	$\lambda P.P(x) \colon (e \rightarrow t) \rightarrow t$
$\mathbf{N}_i \rightarrow$ man,...	$\lambda x.man(x) \colon e \rightarrow t$,...
$\mathbf{IV}_i \rightarrow$ runs,...	$\lambda x.run(x) \colon e \rightarrow t$,...
$\mathbf{IV}_i^- \rightarrow$ run,...	$\lambda x.run(x) \colon e \rightarrow t$,...
$\mathbf{TV}_{i,j} \rightarrow$ loves,...	$\lambda \alpha.\lambda x.\alpha(\lambda y.loves(x,y)) \colon ((e \rightarrow t) \rightarrow t) \rightarrow (e \rightarrow t)$,...
$\mathbf{TV}_{i,j}^- \rightarrow$ love,...	$\lambda \alpha.\lambda x.\alpha(\lambda y.loves(x,y)) \colon ((e \rightarrow t) \rightarrow t) \rightarrow (e \rightarrow t)$,...
$\mathbf{TV}_{i,j}^p \rightarrow$ loved,...	$\lambda \alpha.\lambda x.\alpha(\lambda y.loves(x,y)) \colon ((e \rightarrow t) \rightarrow t) \rightarrow (e \rightarrow t)$,...
$\mathbf{Adj}_i \rightarrow$ mortal,...	$\lambda x.mortal(x) \colon e \rightarrow t$,...
$\mathbf{Pn}_i \rightarrow$ Mary,...	$\lambda P.P(Mary) \colon (e \rightarrow t) \rightarrow t$,...

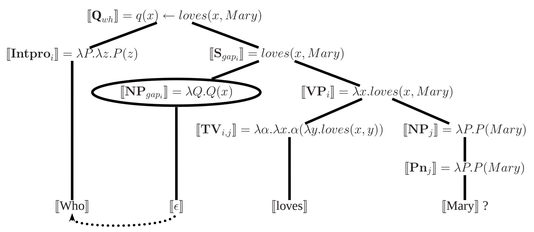
Expressing Graph-Shaped Conjunctive Queries

Definition 2. A non-boolean *graph-shaped simple conjunctive query* (GCQ) of arity ≤ 1 is a CQ over a schema $\mathbf{R}$ composed of relation symbols of arity ≤ 2 of the form $q := q(x) \leftarrow \Phi[x]$ where the body $\Phi[x]$ is inductively defined as:

$$\Phi[x] := A_{i_0}(x) \wedge ... \wedge A_{i_m}(x) \wedge R_{j_0}(x,x) \wedge ... \wedge R_{j_m}(x,x) \wedge \wedge R_{j_0}(x,c) \wedge R_{j_m}(x,c).$$
$$\Phi[x] := \Phi'[x] \wedge \exists y(A_{i_0}(x) \wedge ... \wedge A_{i_m}(x) \wedge R_{j_0}(x,y) \wedge ... \wedge R_{j_m}(x,y) \wedge R_{j_0}(y,x) \wedge R_{j_m}(y,x) \wedge \Phi''[y]).$$

A *boolean* GCQ is a query of the form $q := q() \leftarrow \exists y \Phi[y]$, where $\Phi[y]$ is the body of a non-boolean GCQ.

Theorem 2 (Expressing GCQs)**.** *For every question Q in* GCQ-English*, there exists a* GCQ *q s.t. $[\![Q]\!] = q$. Conversely, every* GCQ *q is the image by $[\![.]\!]$ of some question Q in* GCQ-English*.*



AGCQ-English

- AGCQ-English is a controlled fragment of English that extends the coverage of GCQ-English to graph-shaped aggregate conjunctive queries. It extends the set of Wh-questions of GCQ-English (there are no boolean AGCQs).
- We introduce a numerical type, $\mathbb{Q}$ together with *aggregate determiners* of type $(\mathbb{Q} \rightarrow t) \rightarrow \mathbb{Q}$ to cover **min** and **max**.

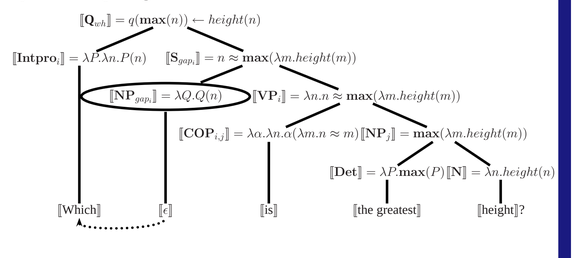
Grammar rule	**Semantic actions**
$\mathbf{VP}_{i,j} \rightarrow \mathbf{COP}$ $\mathbf{NP}_j$	$[\![\mathbf{VP}_{i,j}]\!] := [\![\mathbf{COP}]\!]([\![\mathbf{NP}_j]\!])$

Lexicon	**Semantic actions**
$\mathbf{Det} \rightarrow$ the greatest	$\lambda P.\mathbf{max}(P) \colon (\mathbb{Q} \rightarrow t) \rightarrow \mathbb{Q}$
$\mathbf{Det} \rightarrow$ the smallest	$\lambda P.\mathbf{min}(P) \colon (\mathbb{Q} \rightarrow t) \rightarrow \mathbb{Q}$
$\mathbf{Det} \rightarrow$ some	$\lambda P.\lambda Q.\exists n(P(n) \wedge Q(n)) \colon (\mathbb{Q} \rightarrow t) \rightarrow ((\mathbb{Q} \rightarrow t) \rightarrow (\mathbb{Q} \rightarrow t))$
$\mathbf{Pro}_i \rightarrow$ something	$\lambda P.\exists n P(n) \colon (\mathbb{Q} \rightarrow t) \rightarrow t$
$\mathbf{Pro}_i^- \rightarrow$ anything	$\lambda P.\exists n P(n) \colon (\mathbb{Q} \rightarrow t) \rightarrow t$
$\mathbf{Pro}_i \rightarrow$ it	$\lambda P.P(n) \colon (\mathbb{Q} \rightarrow t) \rightarrow t$
$\mathbf{Pro}_i \rightarrow$ itself	$\lambda P.P(n) \colon (\mathbb{Q} \rightarrow t) \rightarrow t$
$\mathbf{Coord} \rightarrow$ and	$\lambda P.\lambda Q.\exists n(P(n) \wedge Q(n)) \colon (\mathbb{Q} \rightarrow t) \rightarrow ((\mathbb{Q} \rightarrow t) \rightarrow (\mathbb{Q} \rightarrow t))$
$\mathbf{Relpro}_i \rightarrow$ that	$\lambda P.\lambda n.P(n) \colon (\mathbb{Q} \rightarrow t) \rightarrow (\mathbb{Q} \rightarrow t)$
$\mathbf{Intpro}_i \rightarrow$ which	$\lambda P.\lambda n.P(n) \colon (\mathbb{Q} \rightarrow t) \rightarrow (\mathbb{Q} \rightarrow t)$
$\mathbf{COP}_{i,j} \rightarrow$ is	$\lambda n.\lambda m.n \approx m \colon \mathbb{Q} \rightarrow (\mathbb{Q} \rightarrow t)$
$\mathbf{NP}_{gap_i} \rightarrow \epsilon$	$\lambda P.P(n) \colon (\mathbb{Q} \rightarrow t) \rightarrow t$
$\mathbf{N}_i \rightarrow$ height,...	$\lambda n.height(n) \colon \mathbb{Q} \rightarrow t$,...
$\mathbf{Adj} \rightarrow$ odd,...	$\lambda n.odd(n) \colon \mathbb{Q} \rightarrow t$,...

Expressing Aggregate Graph-Shaped Conjunctive Queries

Definition 3. A *graph-shaped conjunctive aggregate query* (AGCQ) over a relational schema $\mathbf{R}$ is a query of the form $q(\alpha(n)) \leftarrow \Phi[n]$, where $\alpha \in \{\mathbf{min}, \mathbf{max}\}$, n is q's distinguished variable, a numerical variable, and $\Phi[n]$ is the body of a non boolean GCQ. Note that there are no boolean AGCQs.

Theorem 3 (Expressing AGCQs)**.** *For every question Q in AGCQ-English, there exists a AGCQ q s.t. $[\![Q]\!] = q$. Conversely, every ATCQ q is the image by $[\![.]\!]$ of some question Q in AGCQ-English.*



What We Can and Cannot Express

Question	**Type of Question**
Who loves Mary?	Wh
Which is the greatest height that is odd?	Wh
Who loves some woman that loves him?	Wh
Who loves himself?	Wh
Which man loves somebody who is mortal and who loves?	Wh
Does anybody love somebody?	Y/N
* Is anybody higher than 190cm?	Y/M
* What is the total number of men in China?	Wh
* Which is the meaning and purpose of life?	Wh

References

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

Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable Reasoning and Efficient Query Answering in Description Logics: The DL-Lite Family. *JAR*, 2007.

Norbert E. Fuchs, Kaarel Kaljurand, and Gerold Schneider. Attempto Controlled English Meets the Challenges of Knowledge Representation, Reasoning, Interoperability and User Interfaces. http://www.ifi.unizh.ch/attempto/publications, 2005.

Richard Montague. Universal Grammar. *Theoria*, (36), 1970.