# Controlled English for DL-Lite

Camilo Thorne

Bolzano, 26/10/06

# Outline

**A.** Introduction

The problem

Why controlled languages

**B.** Controlled English

Attempto controlled English

Pratt's approach

**C.** DL-Lite and Lite-English

DL-Lite

English-Lite

**D.** Preliminary results

## The Problem

- Natural language (NL) is not, *prima facie*, well-suited for querying and managing data or information stored in a relational database (DB) or a knowledge base (KB), let alone for engaging in knowledge representation (KR) tasks.
- On the other hand, formal query languages (like SQL or Datalog) and KR formalisms like description logics (DL's) have an unambigous syntax and semantics, together with a well-defined expressivity.
- Querying can be seen in this context as evaluating a first order logic (FOL) formula over a finite model under a set of constraints (i.e. query answering can be conceived, *grosso modo* as FOL model-checking).
- It would be desirable to reach a trade-off between this restricted expressivity and the intuitive meaning of everyday speech.

# Why Controlled Languages

- We believe that using a subset of, say, English with a limited vocabulary and a restricted syntax (a controlled language) can tackle this problem.
- A controlled language will disallow or preclude ambiguity and map into some logic.
- However, this approach has to be accompanied by a thorough study of the expressivity of the controlled language, since we want it to be as expressive as the query languages and the KR languages, but not more.
- Let's see why this makes sense.

# Controlled Languages

- A **controlled language** (CL) is a fragment of NL (say, of English), with a limited lexicon and small set of grammar rules.
- A CL sentence is compiled into a logic formula that constitutes its **meaning representation** (MR).
- We retain compositionality and, furthermore, we can enginneer the grammar (and, ultimately the semantics) of the CL in such a way that every sentence has a unique parse and a unique logical form (up to logical equivalence) belonging to some well-known KR formalism or query language.

# Attempto Controlled English (ACE)

▷ Attempto Controlled English (ACE) is a controlled language that has been advanced as a suitable means for KR.

▷ Every sentence of ACE is unambigously compiled into a FOL sentence – its MR.

▷ To achieve this it imposes structural contraints on English.

▷ For instance, quantifiers are allowed one unique reading, namely, surface order.

# An Example - ACE (2)

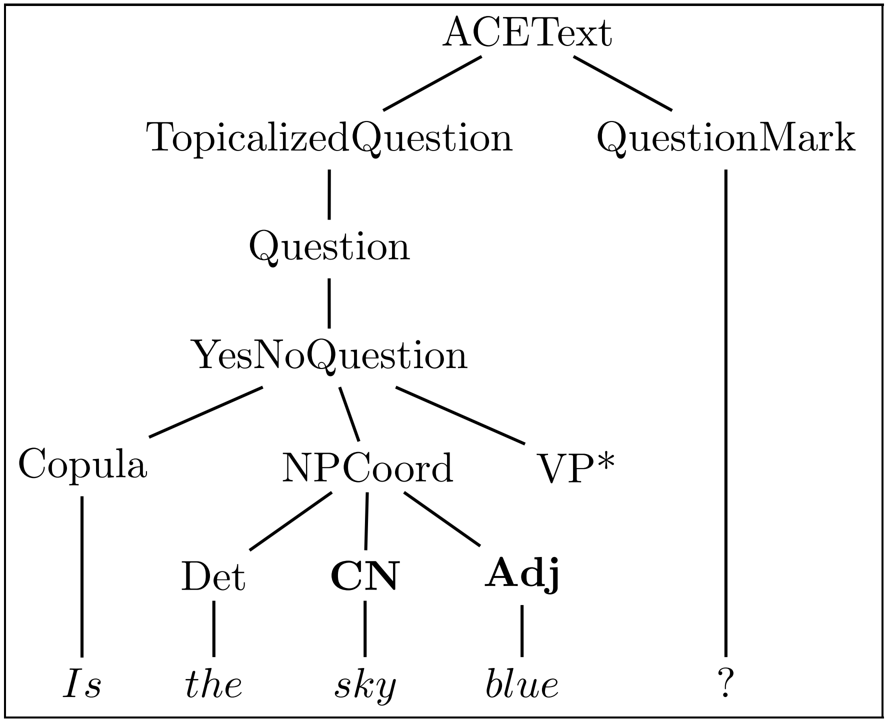



Figure 1: *Parsing an ACE Y/N-question*

## Pratt's Approach

- ▷ Pratt in [16, 17, 18, 15] defines a family of fragments of English and studies them w.r.t. to expressive power.
- ▷ COP, the minimal fragment, deals only with copula, negation, nouns (common and proper) and existential and universal quantification.
- ▷ COP+TV+DTV extends the coverage to transitive verbs and distransitive verbs. COP+REI COP+TV+DTV are their extensions that cover relative pronouns.
- ▷ **Expressive power** is defined in terms of that of the underlying meaning representation logic.

# An Example - COP (1)

| **COP's Phrase-structure Rules** |
|---|
| $\mathbf{IP}/(\phi)\psi \rightarrow \mathbf{NP}/\phi\mathbf{I'}/\psi$ |
| $\mathbf{I'}/\phi \rightarrow$ is a $\mathbf{N'}/\phi$ |
| $\mathbf{I'}/\neg\phi \rightarrow$ is not a $\mathbf{N'}/\phi$ |
| $\mathbf{NP}/\phi \rightarrow \mathbf{PropN}/\phi$ |
| $\mathbf{NP}/(\phi)\psi \rightarrow \mathbf{Det}/\phi\mathbf{N'}/\psi$ |
| $\mathbf{N'}/\phi \rightarrow \mathbf{N}/\phi$ |

# An Example - COP (2)

| COP's Function Lexicon |
|---|
| **Det**/$\lambda P \lambda Q \forall x[(P)x \rightarrow (Q)x] \rightarrow$ every |
| **Det**/$\lambda P \lambda Q \forall x[(P)x \rightarrow \neg(Q)x] \rightarrow$ no |
| **Det**/$\lambda P \lambda Q \exists x[(P)x \wedge (Q)x] \rightarrow$ some |

| COP's Content Lexicon |
|---|
| **N**/$\lambda x(P)x \rightarrow$ woman |
| **N**/$\lambda y(Q)x \rightarrow$ human |
| **PropN**/$\lambda P(P)d \rightarrow$ Diana |

# An example - COP (3)

- The grammar of COP allows some kinds of sentences of NL:

| **Sentence** | **FOL** |
|---|---|
| Diana is a woman | $Woman[d]$ |
| Some woman is a human | $\exists x[Woman[x] \wedge Human[x]]$ |
| Every woman is a human | $\forall x[Woman[x] \rightarrow Human[x]]$ |

- But not all. A sentence like:

  Every woman who is human is human

  Lies outside of the language given that it does not deal with pronouns.

- This is an important issue.

# Complexity of Controlled Languages (Pratt 2003)

| 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** |
| COP+REL+TV+RA | **NEXPTIME-Complete** |
| COP+REL+TV+GA | undecidable |

## Why Description Logics

- Description logics are decidable fragments of FOL specifically conceived for declaring, queryng and reasoning over knowledge bases.
- Their formal properties are well known.
- They have the advantage that for some reasoning tasks they are more efficient than FOL (cf. [1]).
- What we need now is to choose the best-suited for querying – DL-Lite (cf. [6]).

# DL-Lite (1)

- Let $A = \{A_i | i \in \mathbb{N}\}$ and $R = \{R_i | i \in \mathbb{N}\}$ be two countable sets of primitive **concept** and **role** symbols.
- Let $K = \{c_i | i \in \mathbb{N}\}$ be a set of **constants**.
- DL-Lite left and right hand side **concepts** are then defined as follows:

$$B ::= A | \exists R | \exists R^{-}.$$

$$C ::= B | \neg B | C \sqcap C | \exists R \colon C | \exists R^{-} : C.$$

- And DL-Lite A-box and T-Box **assertions** as follows:

$$A ::= B(K) | R(K, K) | A(K).$$

$$T ::= R \sqsubseteq R | B \sqsubseteq C.$$

- A DL-Lite **knowledge base** is a set of A-box and T-Box assertions.

# DL-Lite (2)

- DL-Lite corresponds to a tractable fragment of FOL.
- This is because operators such as, for instance, negation and the universal and existential quantifiers cannot be used freely.
- T-Box satisfiability is in **P** and query answering is **LOGSPACE** in data complexity (i.e. on the size of the KB or, alternatively, of the DB).
- Using DL-Lite as a meaning representation formalism means that reasoning tasks can be carried out efficiently by means of tools such as QONTO (cf. Calvanese *et al.* [7]).

# DL-Lite and Natural Language

▷ DL-Lite can express properties that NL can express:

| Language | Sentence |
| --- | --- |
| NL | Every woman who is human is human |
| FOL | $\forall x[Woman[x] \land Human[x] \rightarrow Human[x]]$ |
| DL-Lite | $Woman \sqcap Human \sqsubseteq Human$ |

▷ But not all:

Diana in not a man.

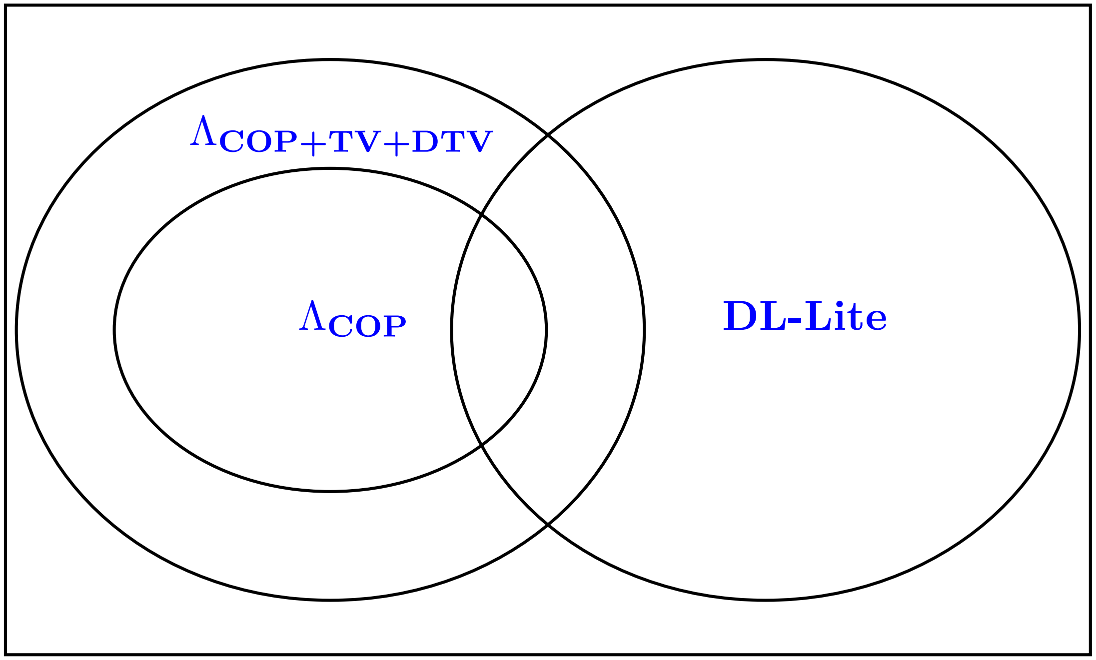
▷ Since its syntax disallows negative facts

# Expressive Power of DL-Lite (1)

**Theorem.** We have proven the following:

1. The expressive power of $\Lambda_{\mathrm{COP}}$ cannot be compared to that of DL-Lite.
2. The expressive power of $\Lambda_{\mathrm{COP+TV+DTV}}$ cannot be compared to that of DL-Lite.
3. That their expressive power nevertheless overlaps.

# Expressive Power of DL-Lite (2)



# Lite-English

- Lite-English is a controlled language where sentences can be translated at runtime, during parsing, into a DL-Lite meaning representation with little or no ambiguity. It has a compositional semantics based on Montague's approach (i.e. on lambda expressions, type theory and the homeomorphism principle, cf. [12]).
- Syntactic composition mirrors semantic composition.
- This implies (trivially) that Lite English and DL-Lite share the same expressive power.

## Expressive Power of Lite English

- Since Lite English inherits the expressiveness of DL-Lite itself, its satisfiability problem is in **P**.
- Moreover, it allows for a restricted use of relative clauses without yielding a state explosion.
- This is relevant, since the addition of relatives to COP or COP+TV+DTV yields untractability.
- This comes from the fact that relative pronouns are constrained in their use both in subject and predicate position, thus precluding full-blown gap-filler dependencies.

## Lite English's Grammar

- A grammar for Lite English has to constrain parse trees in such a way that sentences compile into DL-Lite.
- For instance, it must allow negation only over nouns, adjectives and verbs in predicate position (but not over VP's or any other recursive component).
- this can be achived by restricting the syntax rules related to the function lexicon.
- This characteristic is independant of the grammar formalism. We have tried categorial grammar (CG) (cf. Bernardi *et al.* [2]) and implemented a sample parser based on phrase structure grammar (PSG).

## Conclusions

- We have defined a controlled language, Lite-English, based on the KR logic DL-Lite sharing its expressive power.
- Its expressiveness is therefore suitable for representing and managing data in restricted domains, just as it is the case for DL-Lite.
- Its satisfiablity problem is in **P**.
- It provides, moreover, coverage for (restricted) long distance dependencies without compromising these complexity results, as it is the case with Pratt's fragments or ACE.

# References


[1] F. Baader, D. Calvanese, D. Nardi, P. Patel-Schneider, and D. McGuinness. *The Description Logic Handbook*. Cambridge University Press, 2003.

[2] Raffaella Bernardi, Diego Calvanese, and Camilo Thorne. Lite natural language. Submitted to IWCS-7, 2006.

[3] Alex Borgida. On the relative expressiveness of description logics and predicate logics. *Artificial Intelligence*, (82), 1996.

[4] Johan Bos and Patrick Blackburn. *Representation and Inference for Natural Language: A First Course in Computational Semantics*. CSLI, 2005.

[5] Carlos Caleiro and Ricardo Gonçalves. Equipollent logical systems. In Jean-Yves Beziau, editor, *Logica Universalis*, pages 99–112. Birkhauser, 2005.

[6] Diego Calvanese, Giusseppe de Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Dl-lite: Tractable description logics for ontologies. In *Proceedings*

*of the 20th National Conference on Artificial Intelligence (AAAI 2005)*, 2005.

[7] Diego Calvanese, Giuseppe de Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Efficiently managing data intensive ontologies. In *Proceedings of the 2nd Italian Semantic Web Workshop: Semantic Web Applications and Perspectives (SWAP 2005)*, 2005.

[8] Bob Carpenter. *Type-Logical Semantics.* The MIT Press, 1997.

[9] René Cori and Daniel Lascar. *Logique Mathématique (2 vols).* Dunod, 2003.

[10] René David, Karim Nour, and Christophe Rafalli. *Introduction à la Logique. Théorie de la Démonstration.* Dunod, 2001.

[11] 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.

[12] L. T. F. Gamut. *Logic, Language and Meaning (2 vols.).* University of Chicago Press, 1991.

[13] R. Hull, S. Abiteboul, and V. Vianu. *Foundations of Databases.* Addison-Welsey,

1995.

[14] René LALÉMENT. *Logique, Réduction, Résolution.* Dunod, 1997.

[15] Ian PRATT. On the semantic complexity of some fragments of english. Technical report, Department of Computer Science – University of Manchester, 2001.

[16] Ian PRATT. A two-variable fragment of english. *Journal of Logic, Language and Information*, (12), 2003.

[17] Ian PRATT. Fragments of language. *Journal of Logic, Language and Information*, (13), 2004.

[18] Ian PRATT and Allan THIRD. More fragments of language. *Notre Dame Journal of Formal Logic*, 2005.

[19] Rolf SCHWITTER and Marc TILBROOK. Let's talk in description logic via controlled language. In *Proceedings of the Third International Workshop on Logic and Engineering of Natural Language Semantics (LENLS2006)*, 2006.

[20] SHAPIRO and STERLING. *The Art of Prolog.* The MIT Press, 1994.

[21] John SOWA. *Knowledge Representation: Logical, Philosophical and Computational*

*Foundations*. Brooks Cole Publishing Co., 1999.

[22] Lutz Strassburger. What is a logic, and what is a proof? In Jean-Yves Beziau, editor, *Logica Universalis*, pages 135–145. Birkhauser, 2005.

[23] Alfred Tarski. The semantic conception of truth and the foundation of semantics. *Philosophy and Phenomenological Research*, (4), 1944.

[24] Dirk Van Dalen. *Logic and Structure*. Springer, 1997.

[25] Pierre Wolper. *Introduction à la calculabilité*. Dunod, 2001.

# A Sample DL-Lite Knowledge Base

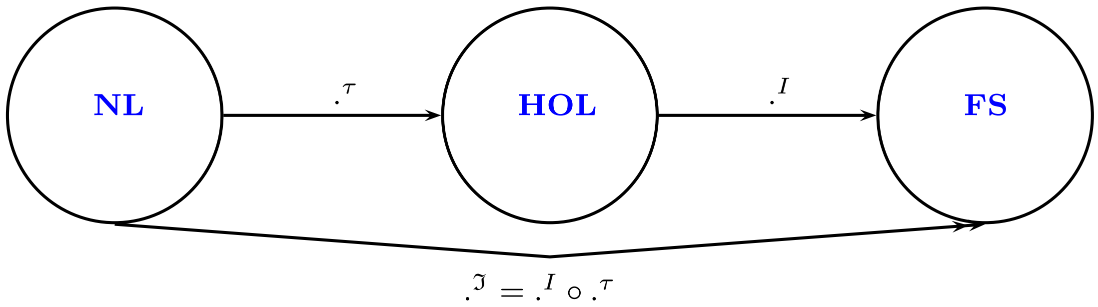
**T-Box**

$$MarriedMan \sqsubseteq Man$$
$$Man \sqsubseteq \exists Loves$$
$$Man \sqsubseteq \exists Loves\colon Woman$$
$$Man \sqsubseteq LivingCreature$$

**A-Box**

$$Man(James)$$
$$MarriedMan(John)$$
$$Loves(John, Mary)$$
$$(\exists Loves\colon Woman)(John)$$

## Homeomorphism Principle (Montague)

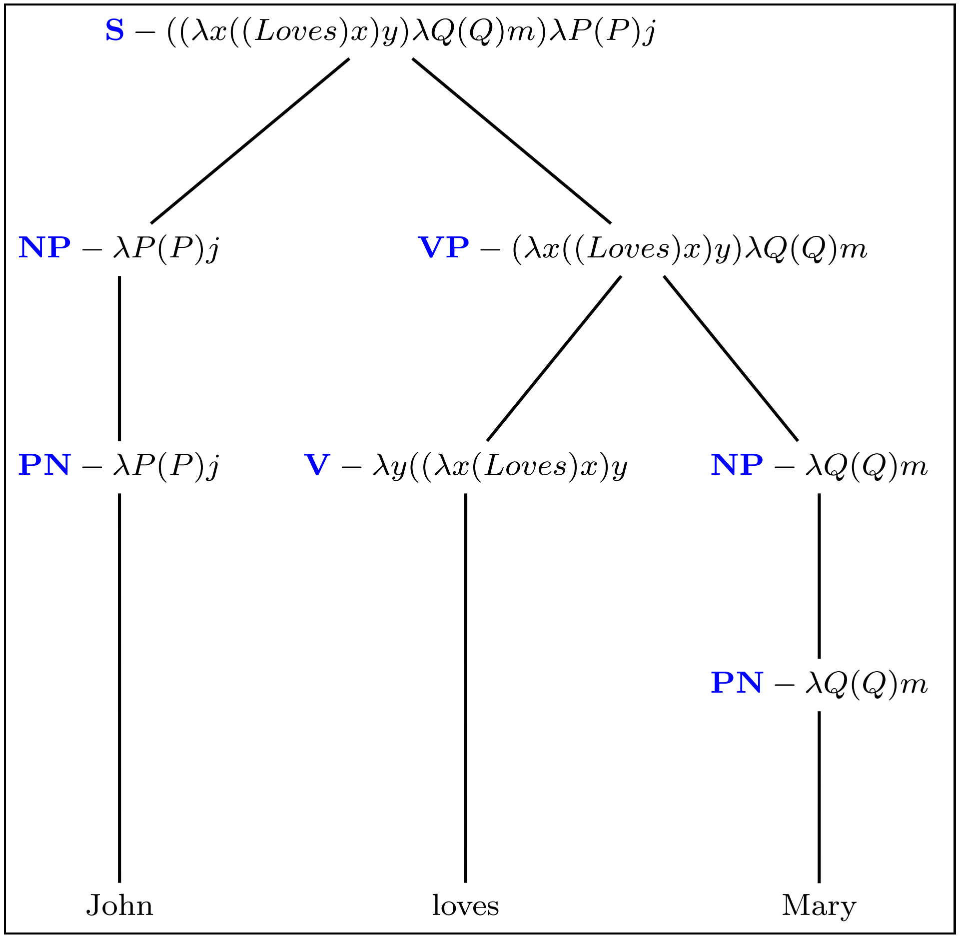



- Mappings $.^{I}$ and $.^{\tau}$ are, respectively. homeomorphims from NL to HOL and from HOL to FS.
- Therefore $.^{\mathfrak{I}} = .^{I} \circ .^{\tau}$ is an homeomorphism from NL to FS.
- Syntactic composition is thus homeomorphic to semantic composition and $.^{\mathfrak{I}}$ provides a compositional semantics for NL.

# Lite-English - Logical Form of Categories

| **POS** | **Word** | **Lexicon** | **MR** $(= .^{\tau})$ | **Type** |
|---|---|---|---|---|
| Determiner | every | Function | $\lambda P \lambda Q \forall x[(P)x \rightarrow (Q)x]$ | $(e \rightarrow t) \rightarrow t$ |
| Determiner | some | Function | $\lambda P \lambda Q \exists x[(P)x \wedge (Q)x]$ | $(e \rightarrow t) \rightarrow t$ |
| Proper noun | Julian | Content | $\lambda P(P)j$ | $(e \rightarrow t) \rightarrow t$ |
| Intransitive Verb | rules | Content | $\lambda x(rules)x$ | $e \rightarrow t$ |
| Transitive verb | loves | Content | $\lambda P \lambda x(P) \lambda y((loves)x)y$ | $((e \rightarrow t) \rightarrow t) \rightarrow (e \rightarrow t)$ |
| Common noun | man | Content | $\lambda x(man)x$ | $e \rightarrow t$ |
| Relative pronoun | who | Function | $\lambda P \lambda Q \lambda x[(P)x \wedge (Q)x]$ | $(e \rightarrow t) \rightarrow (e \rightarrow t)$ |
| Pronoun | somebody | Function | $\lambda P \exists x(P)x$ | $(e \rightarrow t) \rightarrow t$ |

# A Sample Parse Tree of Lite English (1)



# A Sample Parse Tree of Lite-English (2)

$$\begin{aligned}(\lambda P(P)j)(\lambda Q(Q)m)\lambda y\lambda x((Loves)x)y &\rhd_\beta (\lambda P(P)j)(\lambda y\lambda x((Loves)x)y)m\\ &\rhd_\beta (\lambda P(P)j)\lambda x((Loves)x)m\\ &\rhd_\beta (\lambda x((Loves)x)m)j\\ &\rhd_\beta ((Loves)j)m\end{aligned}$$

▷ This HOL normalized expression is actually a DL-Lite A-Box assertion, namely:

$$Loves(John, mary).$$

▷ It has been computed compositionally in the sense that syntactic composition is mirrored by lambda-calculus application (with or without reduction).