

---

---

# Camilo Thorne

---

---

## Postdoctoral Researcher

DWS Group, University of Mannheim  
B6, 26, 68159, Mannheim, Germany

<http://www.linkedin.com/in/camilothorne>: linkedin

<http://www.camilothorne.com>: website

[camilo@informatik.uni-mannheim.de](mailto:camilo@informatik.uni-mannheim.de): email

+49(0)15202380353: mobile

ACM and SIGSEM member, Peruvian citizen

---

## Research Interests

---

- Natural Language Processing
- Formal Semantics
- Knowledge Representation
- Healthcare Informatics
- Ontologies
- Machine Learning

---

## Research Experience

---

**Apr 15** ▷ **pres**     **Post-doctoral Researcher.** DWS Group, University of Mannheim, Mannheim, Germany.

I currently work in a number of topics, such as quantifier distribution and complexity, probabilistic reasoning and entity linking in specialized domains. In addition, I participate in some industrial cooperations of the faculty of business informatics.

**Jun 14** ▷ **Mar 15**     **Computational Linguist.** 10 months. IBM Language and Knowledge Center, Trento RISE, Povo di Trento, Italy.

I worked as an applied natural language processing researcher in an IBM project on domain-specific (troubleshooting) chatterbots and text-based dialog systems. In addition, I actively collaborated in the development of proprietary IBM Java language technologies (e.g., Watson) for Italian.

**Jan 11** ▷ **Jan 14**     **Post-doctoral Researcher.** 3 years. KRDB Research Centre for Knowledge and Data, Free University of Bozen-Bolzano. Bolzano, Italy.

Studied information extraction and entity recognition techniques to automatically extract process fragments from clinical guidelines in the “VeriCliG - Extraction and Verification of Clinical Guidelines” project. A Java-based experimental framework based on WordNet, the Weka data mining framework, the Stanford natural language processing API and the UMLS Metathesaurus was developed. In parallel to this, I studied the distribution of controlled natural language constructs in corpora (via Python-based corpus linguistics techniques).

<http://www.camilothorne.com/vericlig/>

**Jan 06** ▷ **Dec 10**     **Pre-doctoral Researcher.** 5 years. KRDB Research Centre for Knowledge and Data, Free University of Bozen-Bolzano. Bolzano, Italy.

Defined and studied the data complexity of families of controlled natural languages in one-to-one relationship with semantic web (OWL) ontologies and formal query languages (SQL, SQL with aggregations) for ontology-based data access. This work

was partially funded by the projects European FP6-FP7 “NET<sup>2</sup> - Foundations of Networked Knowledge” and “TONES - Thinking Ontologies” projects.

<http://net2.deri.ie/net2/> and <http://www.inf.unibz.it/tones/>

**Jul 05 ▷ Sep 05** **Intern. 3 months.** Knowledge Media Institute (KMi), Open University. Milton Keynes, UK.

Extended the KMi CORDER (COmmunity Relation Discovery by named Entity Recognition) algorithm, to extract domain ontologies using unsupervised machine learning (hierarchical clustering). A Java (SDK 1.5, Collections and JDBC APIs) prototype was developed.

<http://projects.kmi.open.ac.uk/corder>

**Apr 04 ▷ Sep 04** **Intern. 6 months.** CNRS (Centre National de la Recherche Scientifique), Laboratoire d’Informatique de Paris Nord (LIPN). Villetaneuse, France.

Defined the class of categorial module grammars and studied their learnability (in Gold’s model). A small LISP parser was developed.

---

## Education

---

**Dec 05 ▷ Mar 10** **PhD (“Dottorato di Ricerca”) in Computer Science.** Free University of Bozen-Bolzano. Bolzano, Italy. Focus: Formal (Montague) semantics, controlled natural languages, knowledge representation, ontologies and computational logic. Language: English. PhD dissertation: *Query Answering over Ontologies Using Controlled Natural Language*. Supervisors: Prof. Diego Calvanese and Prof. Rafaella Bernardi. Reviewers: Prof. Ian Pratt Hartmann and Prof. Carlos Areces.

**Oct 01 ▷ Sep 04** **MSc (“Maîtrise, DEA”) in Computer Science.** Université de Paris 1 Panthéon-Sorbonne, Université de Paris 13 Villetaneuse. Paris and Villetaneuse, France. Focus: logic (first order model and proof theory, set theory), knowledge representation, optimization, text processing (POS tagging, IE and clustering), Java, Prolog, Lisp programming. Language: French. MSc thesis: *Apprentissage de grammaires catégorielles de modules*. Supervisor: Prof. Denis Béchet.

**Oct 00 ▷ Jun 01** **BSc (“Licence”) in Logic.** Université de Paris 1 Panthéon-Sorbonne. Paris, France. Focus: logic, discrete mathematics, theory of computation. Language: French.

**Mar 92 ▷ Jun 97** **BA (“Licenciatura”) in Humanities.** Pontificia Universidad Católica del Perú. Lima, Peru. Focus: Linguistics, philosophy. Language: Spanish.

**Jan 89 ▷ Dec 91** **International Baccalaureate.** Newton College. Lima, Peru. Language: English and Spanish.

---

## Teaching Experience

---

**Materials** <http://www.camilothorne.com/teaching/>

**Aug 16** **Lecturer: Empirical Methods for Linguistics.** 1 week. 2016 NASSLLI Summer School, Rutgers University, NJ, USA.

Will teach an introductory course on statistics and machine learning for corpus analysis.

**Mar 16 ▷ Jun 16** **Teaching Assistant: Knowledge Management.** 6 months. Faculty of Business Informatics, University of Mannheim, Mannheim, Germany.

Teaching the labs of the 2016 MSc (core) in knowledge management.

**Aug 15** **Lecturer: Dialog Systems.** 1 week. 2015 ESLLI Summer School, Pompeu Fabra University, Barcelona, Spain.

Taught an introductory course on dialog systems, with a particular emphasis on state-based and rule-based systems.

**Mar 14 ▷ Feb 15** **Teaching Assistant: Data Structures and Algorithms.** 1 year. Faculty of Computer Science, Free University of Bozen-Bolzano. Bolzano, Italy.

Taught the labs of the 2014 BSc (core) course in data structures and algorithms.

**Aug 13** **Lecturer: Formal Semantics.** 1 week. 2013 ESLLI Summer School. University of Düsseldorf, Düsseldorf, Germany.

Taught a course on aggregation in English, viz., on how SQL-like aggregate functions occur in English, how they fit within generalized quantifier theory, how they can be computed, their computational properties, and their distribution vis-à-vis other generalized quantifiers.

**Mar 11 ▷ Feb 13** **Teaching Assistant: Computational Logic.** 2 years. Faculty of Computer Science, Free University of Bozen-Bolzano. Bolzano, Italy.

Taught the lab of the European MSc course in computational logic, introducing the students to propositional, first-order and description logic modelling and solving tools (e.g., Minisat, SPASS, Pellet).

**Feb 00 ▷ Jul 00** **Lecturer: Introductory Logic.** 6 months. Faculty of Engineering, Universidad Privada Ricardo Palma. Lima, Peru.

Taught a first-year BSc introductory course in logic (propositional and first-order).

---

## Research Projects

---

**Feb 12 ▷ Jan 14** **VeriCliG - Extraction and Verification of Clinical Guidelines (FUB Foundation).** KRDB Research Centre for Knowledge and Data, Bolzano, Italy.

**Jun 10 ▷ May 13** **NET<sup>2</sup>: A Network for Enabling Networked Knowledge (FP7-247601).** KRDB Research Centre for Knowledge and Data, Bolzano, Italy.

**Jan 06 ▷ Aug 08** **TONES: Thinking Ontologies (FP6-7603)** KRDB Research Centre for Knowledge and Data, Bolzano, Italy.

**Jul 05 ▷ Sep 05** **CORDER (KMi internal).** KMi, Milton Keynes, UK.

---

## Publications

---

### Preprints

<http://www.camilothorne.com/publications/>

### 2016

Jenny Copara, Jose Eduardo Ochoa Luna, Camilo Thorne and Goran Glavaš - “Spanish NER with Word Representations and Conditional Random Fields”. 2016, in: Proceedings of the NEWS 2016 workshop (Berlin, Germany). Workshop paper.

Jenny Copara, Jose Eduardo Ochoa Luna, Camilo Thorne and Goran Glavaš - “Conditional Random Fields for Spanish Named Entity Recognition Using Unsupervised Features”. 2016, in: Proceedings of the IBERAMIA 2016 conference (San José, Costa Rica). Conference paper.

Jenny Copara, Jose Eduardo Ochoa Luna, Camilo Thorne and Goran Glavaš - “Exploring unsupervised features in Conditional Random Fields for Spanish Named Entity Recognition”. 2016, in: Proceedings of the BRACIS 2016 symposium (Recife, Brazil). Conference paper.

Camilo Thorne, Stefano Faralli and Heiner Stuckenschmidt. “Entity Linking for Clinical Text Annotation and Disambiguation”. 2016, in: Proceedings of the SEMANTiCS 2016 conference (Leipzig, Germany). Conference paper.

Erman Acar, Manuel Fink, Christian Meilicke, Camilo Thorne and Heiner Stuckenschmidt - “Multi-attribute Decision Making with Weighted Description Logics”. 2016, In: IFCoLog. Journal paper. (under review)

Jakub Szymanik, Camilo Thorne. “Semantic Complexity Influences Quantifier Distribution in Corpora”. 2016, in: Language Sciences – Special Issue on Complexity in Human Languages. Journal paper. (under review)

### 2015

Erman Acar, Camilo Thorne and Heiner Stuckenschmidt. “Towards Decision Making via Expressive Probabilistic Ontologies”. 2015, in: Proceedings of the ADT2015 conference (Lexington, Kentucky). Conference paper.

Camilo Thorne, Jakub Szymanik. “Semantic Complexity of Quantifiers and their Distribution in Corpora”. 2015, in: Proceedings of the IWCS 2015 conference (London, UK). Conference paper.

### 2014

Camilo Thorne, Raffaella Bernardi, Diego Calvanese. “Designing Efficient Controlled Languages for Ontologies”. 2014, in: Computing Meaning, Vol. 4. Book chapter.

### 2013

Camilo Thorne, Elena Cardillo, Claudio Eccher, Marco Montali, Diego Calvanese. “Process Fragment Recognition in Clinical Documents”. 2013, in: Proceedings of the 2013 AI\*IA Conference (Turin, Italy). Conference paper.

Camilo Thorne, Elena Cardillo, Claudio Eccher, Marco Montali, Diego Calvanese. “Clinical Activity Recognition in Clinical Documents”. 2013, in: Proceedings the IJCNLP 2013 conference (Nagoya, Japan). Conference paper.

Camilo Thorne, Jakub Szymanik. “Generalized Quantifier Distribution and Semantic Complexity”. 2013, in: Pre-Proceedings of the TbiLLC2013 symposium (Tbilisi, Georgia). Extended abstract.

- Camilo Thorne, Elena Cardillo, Claudio Eccher, Marco Montali, Diego Calvanese. “The VeriCliG Project: Extraction of Computer Interpretable Guidelines via Syntactic and Semantic Annotation”. 2013, in: Proceedings of the ClinText2013 workshop (Potsdam, Germany). Workshop paper.
- 2012** Camilo Thorne. “Studying the Distribution of Fragments of English Using Deep Semantic Annotation”. 2012, in: Proceedings of the ISA8 workshop (Pisa, Italy). Workshop paper.
- Camilo Thorne, Diego Calvanese. “Tractability and Intractability of Controlled Languages for Data Access”. 2012, in: Studia Logica – Special Issue on Logic and Language. Journal paper.
- 2011** Camilo Thorne. “Querying the Fragments of English”. 2011, in: Proceedings of the WoLLIC 2011 workshop (Philadelphia, USA). Workshop paper.
- 2010** Camilo Thorne. “Query Answering over Ontologies Using Controlled Natural Languages”. 2010, Free University of Bozen-Bolzano, Italy. PhD Thesis.
- 2009** Camilo Thorne, Diego Calvanese. “The Data Complexity of the Syllogistic Fragments of English”. 2009, in: Proceedings of the AC 2009 colloquium. Workshop paper.
- Camilo Thorne, Diego Calvanese. “Controlled English Ontology-Based Data Access”. 2009, in: Proceedings of the CNL 2009 workshop. Workshop paper.
- Camilo Thorne, Diego Calvanese. “Tree Shaped Aggregate Questions over Ontologies”. 2009, in: Proceedings of the FQAS 2009 conference. Conference paper.
- Camilo Thorne, Diego Calvanese. “Exploring Controlled English Ontology-Based Data Access”. 2009, in: Pre-Proceedings of the CNL 2009 workshop. Extended abstract.
- 2008** Diego Calvanese, Evgeny Kharlamov, Werner Nutt, Camilo Thorne. “Aggregate Queries over Ontologies”. 2008, in: Proceedings of the ONISW 2008 workshop. Workshop paper.
- Camilo Thorne. “Expressing Conjunctive and Aggregate Queries over Ontologies with Controlled English”. 2008, in: Proceedings of the ESSLLI 2008 student session. Workshop paper.
- 2007** Raffaella Bernardi, Francesca Bonin, Diego Calvanese, Domenico Carbotta, Camilo Thorne. “English Querying over Ontologies: E-QuOnto. 2007, in: Proceedings of the AI\*IA 2007 conference. Conference paper.
- Camilo Thorne. “Managing Structured Data with Controlled English - An Approach based on Description Logics”. 2007, in: Proceedings of the ESSLLI 2007 student session. Workshop paper.
- Raffaella Bernardi, Diego Calvanese, Camilo Thorne. “Expressing DL-Lite Ontologies with Controlled English”. 2007, in: Proceedings of the DL 2007 workshop. Workshop paper.
- Raffaella Bernardi, Diego Calvanese, Camilo Thorne. “Lite Natural Language”. 2007, in: Proceedings of the IWCS 2007 workshop. Workshop paper.

Camilo Thorne. “Categorical Module Grammars Have Finite Bounded Density”. 2007, in: Proceedings of the LATA 2007 conference. [Conference paper](#).

**2005** Camilo Thorne, Victoria Uren, Jianhan Zhu. “Extracting Domain Ontologies with CORDER”. 2005, KMi. [KMi report](#).

---

## Programming Skills

---

**Code Repos** <https://github.com/camilothorne>

**Languages** Java SDK 1.7+, Python 2.7+, Prolog, Lisp, R.

**NLP & AI** Python: Pyplot, Scipy, Numpy, Sklearn, NLTK APIs.  
Java: Stanford NLP, OpenNLP, WEKA, JENA, Lucene APIs.  
Thesauri: WordNet, BabelNet, UMLS Metathesaurus.  
Reasoners: SPASS, Minisat, Pellet, Blackbox.  
Prediction & classification models: Maximum Entropy, Logistic and Linear Regression Models, Conditional Random Fields; Unsupervised models: K-means, agglomerative clustering.

**APIs & Tools** IDEs: Eclipse, Protégé (ontology development), RStudio.  
Structured datastores: PostgreSQL (relational), Virtuoso (graph); SQL 3, Sparql 1.1, RDF & relational data models.  
Versioning: SVN, Git, Mercurial (protocols); JIRA, Kanban (management).  
Other: L<sup>A</sup>T<sub>E</sub>X. MS Office

**Modeling** OWL 1.1, UML class diagrams/ER diagrams.

**Platforms** Linux/UNIX, Windows, MacOS X.

---

## Language Skills & Interests

---

**Languages** Spanish (native), English (good), French (good), Italian (good), German (average).

**Interests** Creative writing, jogging, swimming, trekking.

---

## Selected Presentations & Talks

---

**Slides** <http://www.camilothorne.com/talks/>

**2016** “Data Science for Deep Natural Language Understanding”. NEC Labs of Europe, Heidelberg, Germany.  
“Cross-Evaluation of Entity Linking and Disambiguation Systems for Clinical Text Annotation”, SEMANTICS 2016 conference, Leipzig, Germany.  
“Spanish NER with Word Representations and CRFs”. NEWS 2016 workshop, Berlin, Germany.  
“Generalized Quantifiers in Corpora”. Data and Web Science Group, Mannheim University, Mannheim, Germany.

- 2014** “Extracting Process Models from Clinical Guidelines”. Data and Web Science Group, Mannheim University, Mannheim, Germany.  
“Dialog Models for Troubleshooting”. IBM meeting, Venice, Italy.
- 2013** “Process Fragment Recognition in Clinical Documents”. AI\*IA 2013 conference, Turin, Italy.  
“Generalized Quantifier Distribution and Semantic Complexity”. TbiLLC 2013 symposium, Gudauri, Georgia.  
“The VeriCliG Project: Extraction of Computer Interpretable Guidelines via Syntactic and Semantic Annotation”. CSLT 2013 workshop, Potsdam, Germany.
- 2012** “Studying the Distribution of Fragments of English Using Deep Semantic Annotation”. ISA8 workshop, Pisa, Italy.
- 2011** “Semantic Complexity of Controlled Languages”. TiLPS, Tilburg, Netherlands.  
“Data-scalable Controlled Languages”. SRI International, Menlo Park, USA.  
“Querying the Fragments of English”. WoLLIC 2011 workshop, Philadelphia, USA.
- 2009** “The Data Complexity of the Syllogistic Fragments of English”. AC 2009 colloquium, Amsterdam, Netherlands.  
“Tree Shaped Aggregate Questions over Ontologies”. FQAS 2009 conference, Roskilde, Denmark.  
“Exploring Controlled English Ontology-Based Data Access”. CNL 2009 workshop, Marettimo, Italy.
- 2008** “Aggregate Queries over Ontologies”. ONISW 2008 workshop, Napa, USA.
- 2007** “Expressing DL-Lite Ontologies with Controlled English”. DL 2007 workshop. Bressanone, Italy.

---

## Collaborations & Activities

---

- October 2014** Visitor, Web and Data Science Group, Mannheim University. Mannheim, Germany.
- Dec 13 ▷ Aug 14** PC member, 4th Workshop on Controlled Natural Language (CNL 2014).
- Dec 12 ▷ pres** Collaborator, Institute for Logic, Language and Computation (ILLC), University of Amsterdam. Amsterdam, Netherlands.
- Feb 12 ▷ Jan 14** Collaborator, Fondazione Bruno Kessler (FBK). Trento, Italy.
- Aug 11** Visitor, computer science department, University of Concepción. Concepción, Chile.
- May 11** Visitor, SRI International. Menlo Park, USA.
- Apr 11 ▷ May 11** Visitor, TiLPS, Tilburg University. Tilburg, Netherlands.
- Jan 11 ▷ pres** Reviewer, Journal of Logic, Language and Information (JoLLI).
- Sep 10** Visitor, computer science department, University of Chile. Santiago, Chile.

**Feb 08** ▷ **Jan 10** Student representative, Academic Senate, Free University of Bozen-Bolzano. Bolzano, Italy.

Student representative (computer science students) at the university senate. Promoted enhancements to the status of PhD students/fellows/researchers.

---

## References

---

**Prof. Heiner Stuckenschmidt** University of Mannheim  
B6, 26, 68159, Mannheim, Germany  
Tel: +49 6211812530  
email: [heiner@informatik.uni-mannheim.de](mailto:heiner@informatik.uni-mannheim.de)

**Mr. Guido Vetere** IBM CAS Trento - IBM Italia  
17 Piazza Mancini, 38123, Povo di Trento, Italy  
Tel: +39 0461312345  
email: [gvetere@ibm.it](mailto:gvetere@ibm.it)

**Prof. Diego Calvanese** Free University of Bozen-Bolzano  
3 Piazza Domenicani, 39100, Bolzano, Italy  
Tel: +39 0471016160  
email: [calvanese@inf.unibz.it](mailto:calvanese@inf.unibz.it)