Area #5 : Lexicon, ontologies, semantic interoperability and information extraction

Chair : Nathalie Aussenac-Gilles (IRIT) aussenac@irit.fr
Area #5: Lexicon, ontologies, semantic interoperability and information extraction

- Concern
  - To improve interaction and mutual understanding
  - To explicitly represent the contents of the information exchanged, mainly in its linguistic form.

- Means
  - What is meaning? Investigate semantics and lexical categories
  - How to represent meaning? Representations, lexicon and ontology
  - How to identify meaning in texts? Text analysis, information extraction
  - How to use representations to represent text contents? Semantic annotation for information retrieval
  - How to human negotiate, stabilize or modify meaning? How to detect, anticipate or represent this?

- Involved disciplines:
  - Formal semantics, knowledge engineering, linguistics, natural language processing, information extraction and retrieval, sociology, philosophy.
Participants to Area #5

- **IRIT – LILAC**
  - Nicholas Asher
  - Philippe Muller
  - Laure Vieu
- **IRIT – CSC**
  - Nathalie Aussenac-Gilles
  - Axel Reymonet
  - Mouna Kamel
  - Bernard Rothenburger
  - Nacim Chikhi
- **IRIT – SIG**
  - Josiane Mothe
  - Mohand Boughanem
  - Florence Sedes
- **UNITN - DIT**
  - Paolo Bouquet
  - Massimo Poesio
  - Fausto Giunchiglia
- **UNITN - DISCOF**
  - Marco Cruciani
  - Vincenzo D'Andrea
  - Alexander Lazovik
- **ISTC - LOA**
  - Nicola Guarino
  - Alessandro Oltramari
  - Laure Vieu
  - Aldo Gangemi
  - Eduard Barbu
  - Carola Catenacci
  - Olga Capirici
  - Cristina Caselli
  - Elena Pizzuto
  - Geri Steve
- **IRST**
  - Luciano Serafini
  - Stafano Zanobini
Session overview

• Activities during 2006
  – Lectures in Verona by N. Asher
  – sample joint work: *Bootstrapping semantics on the Web: meaning elicitation from schemas*, P. Bouquet (with L. Serafini and S. Zanobini)

• Work done
  esda Theoretical investigations about semantics and lexical categories
  esda Techniques and tools to go from language to representations
  esda Meaning negotiation and evolution

• Future directions
  – Knowledge dynamics
  – Position talk by A. Oltramari, *LexiPass methodology: a conceptual path from frames to senses and back*
  – Position talk by N. Guarino, *The "Senso Comune" initiative*
1 - Foundational investigations

N. Asher's lectures on Ontology and Language,
Verona (June 20-21-22 + July 3-5-6, 2006)

• Ground work of a theory of lexical meaning and predication requires: lexical semantics, compositional semantics and discourse semantics
• Technical apparatus of this theory of predication and lexical meaning:
  • a type driven theory of predication
  • typed lambda calculus
  • some previous theories of lexical meaning like the Generative Lexicon of Pustejovsky
  • the notion of complex types with underspecification.
• Link to discourse semantics
1 - Foundational investigations

- Well-founded lexica and ontologies
  - Contribution of formal semantics, lexical and discourse semantics
    - Formal grounding for ontology engineering (ISTC-LOA)
    - Speech act and language analysis (IRIT-LILaC)
  - Articulation between lexica and ontologies
    - Conceptual Analysis of Lexical Taxonomies (ISTC-LOA, OntoWordnet project)
    - Interfacing Ontologies and Lexical Resources (ISTC-LOA, IRIT-CSC)
    - Ontology enrichment with lexical resources (UNITN, ISTC-LOA)
    - Clarifying the distinction between lexica and ontologies (UNITN, ISTC-LOA)
Discourse semantics for analyzing speech acts
(IRIT-LILac with ISTC-LOA)

• applications of discourse semantics in a variety of areas
  – lexical semantics, the semantics of modals, questions, evidentials and quantification, and the theory of speech acts.

• Definition of complex types
  – uses in the lexicon (for analyzing copredication, relative predication and depictive clauses)
  – uses at the discourse level (analysis of complex speech acts)

• A dynamic semantics together with a theory of discourse interpretation can give a much more satisfactory analysis of speech acts
1 - Foundational investigations

Lexical semantics based on ontologically well-founded representation theories (IRIT – LILAC and LOA)

- Semantics of space, time and motion, semantics of parthood

- Functional dependence for parthood relations
  - explains the behavior of functional parthood with respect to transitivity.
  - accounts for a variety of phenomena.

- Committed to the existence of some sort of universals.
  - Linguistic universals were used
  - socially-dependant categories could have been used as well.
  - what exactly are the categories involved? Need for an ontological point of view

- Semantics of time: Formally differences between categories of concrete objects
  - mass terms and singular nouns, singular and collective entities,
  - objects and events, with respect to their spatio-temporal properties.
2 – Tools and methods

Relation identification from texts

• Identification of temporal relations (IRIT-LILaC)
• Automated processing of temporal information in written texts
  – the extraction of temporal adjuncts,
  – the computation of their reference,
  – the extraction of events descriptions and computation of their respective relations.
• Extraction of lexical relations from dictionaries (IRIT-LILaC)
  – using a semantic distance to relate lexical items
  – extracting synonyms
• Towards lexical relation extraction for discourse segmentation
2 – Tools and methods

Attribute and relation extraction from texts

• Attribute acquisition from texts (UNITN)
  – Evaluating representations for attributes or properties
  – Similarities/ differences when building lexicon or ontology
  – what makes for a good 'attribute'? How can that information be learned?

• Evaluation of pattern-based relation extraction from various types of corpora (IRIT-CSC)
  – Dependency of a pattern efficiency and meaning on textual genre
  – Evaluation of pattern-matching results
  – Need for pattern comment to decide how to reuse it
  – Definition of domain specific patterns involving concepts for information extraction
    • Application to bio-medicine, relation between genes and pathologies
2 – Tools and methods

Concept and relation acquisition from texts

• Meaning elicitation from schemas
  – UNITN and IRST, 7 joint papers, P. Bouquet’s talk
  – Lexical resources and ontologies as means to interpret word categories in word hierarchies

• Concept extraction cycle in keeping with document collections (IRIT-CSC and IRIT-SIG)
  – Identification of new concepts / relations from corpora
  – Consistency issues when adding new concepts and relations
2 – Tools and methods

Ontologies for representing « meaning » in documents (IRIT-CSC with IRIT-SIG)

• Document management with ontologies
  – Document classification with a domain ontology (OntoExplo)
  – Document management with concept hierarchies
  – Content description with semantic annotation

• Information retrieval using ontologies or lexical models (i.e. WordNet)
  – Query expansion
  – Document representation with concepts

• Ontologies for querying structured documents
  – when document structure reflects problem solving stages
    • Diagnosis procedures data base
    • Scientific papers and PhD. in human sciences (archaeology)
  – Taking advantage of text structure for automatic annotation with concepts
3 – Meaning negotiation and evolution

Direct negotiation through interaction between agents (cf “agents” area)

• Meaning negotiation process and interests (UNITN)
  – Related linguistic and social dimensions (Actor Network Theory)
  – Goal of negotiation
    • plausible meaning for ambiguous clauses, semantic equilibrium
  – Underlying interests force meaning negotiation within a social network

• Take into account the embodied nature of meaning (Position talk, A. Oltramari (INST-LOA))
  – analysis of static and dynamic dimensions of human cognition
    • meaning is embodied (dynamic dim;)
    • physical features of experiences help to structure cognition and language
  – Application to human-computer interaction and to reduce ambiguity
    • static and dynamic dimensions of cognition need to be encoded
3 – Meaning negotiation and evolution

Mediated “negotiation” through documents, data or knowledge models (ontologies)

how people with different conceptualizations may succeed in exchanging knowledge about a specific domain.

• Semantic coordination (UNITN)
  – Assumption
    • Communication language (e.g. English for humans) works as a coordination tool,
    • It can be used to convey potentially non aligned meanings
  – As a technological solution,
    • reuse the results of meaning elicitation
    • use logical reasoning to derive semantic relations across WDL formulae built by agents equipped with different ontologies.
3 – Meaning negotiation and evolution

Mediated “negotiation” through documents, data or knowledge models (ontologies)

• Focus: Changes that lead to distinct conceptualizations and terminologies in specific domain (IRIT-CSC)
• Method
  – analyzing natural language used in texts
  – Mapping texts and domain models (ontologies or terminologies)
  – Comparing domain ontologies and terminologies
• Technical solutions: NLP + model and text mapping
• Application
  – Localize risk and critical situations in large documentations
  – Linking scientific papers to scientific data through meta-data
  – Reduce mis-interpretations of documents and data in large and long duration projects in space like Rosetta (CNES)
Outlook: knowledge dynamics

• Re-examine right-frontier constraints in dialogue with a dynamic semantics and a theory of discourse interpretation (IRIT - ISTC)

• Meaning evolution in natural language, lexica, cognitive representations and ontologies (UNITN, IRIT, ISTC)
  – Identification in language
  – Representation management

• Meaning negotiation or confrontation (IRIT, UNITN, ISTC)
  – Within human communities
  – In man-system interaction