Using OWL for Rhetorical Ontology Building

Chrysanne Di Marco[1][2]

[1]Cheriton School of Computer Science

[2]Department of English Language and Literature

University of Waterloo

Computational Rhetoric Workshop II
August 13, 2016

What is a computational ontology?

- **Ontological engineering** is used in modelling a domain:
  - Taxonomies of objects and relations between them.
- In a **formal** ontology, relations have precise semantics, e.g.,:
  - Hypernymy/hyponymy—*IsA*.
  - Meronymy—*PartOf*.
- **Description logics** used for many knowledge representation languages:
  - Descriptions of concepts (classes) and roles (relations).
  - Definitions of individuals (instances).
  - **Subsumption** (*IsA*) orders hierarchies of concepts and roles.
- **OWL** (Web Ontology Language) based on description logic:
  - Auto-classification of new instances using subsumption hierarchies.
  - Desirable properties of a formal logic: Soundness, completeness.

Why create an ontology of rhetoric?
(adapted from Noy & McGuinness "Ontology Development 101")

- Share understanding about structure of rhetorical knowledge.
- Enable reuse of rhetorical domain knowledge.
- Make assumptions about rhetoric explicit: Declarative knowledge versus hard-coding in computer program.
- *Analyze domain knowledge:* Formalizing Cognitive Rhetoric.

What suggests rhetoric is "ontological":

Can observe figures as related by familiar semantic relations

Hypernymy/hyponymy:

- **Ploche:** Repetition of a lexeme or lexemes.
    O <u>villain</u>, <u>villain</u>, smiling, damned <u>villain</u>!
- **Epanaphora *IsA* Ploche:** Repetition of a lexeme or lexemes at the **beginnings** of successive phrases.
    <u>To you</u> must go the credit for this, <u>to you</u> are thanks due, <u>to you</u> will this act of yours bring glory.

Meronymy:

- **Anadiplosis:** Starting a clause or phrase with the word or phrase that ended the preceding unit.
    Drake covets <u>loons</u>, <u>loons</u> with cash.
- **Anadiplosis *PartOf* Gradatio:** Gradatio is iterated series of **interconnected** Anadiploses.
    He who <u>controls Berlin</u>, <u>controls Germany</u>, and he who <u>controls Germany</u>, <u>controls Europe</u>, and he who <u>controls Europe</u>, controls the world.

Challenges of modeling rhetoric as an ontology

- **Easy to model:**
  - Natural class/subclass (*IsA*) relationships among figures.
  - Defining characteristic properties of each figure class.
  - Classifying instances (i.e., examples of figures):
    - Can explicitly specify an instance as member of certain classes.
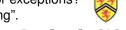   - Or, leave to OWL to auto-classify based on instance's properties.

- **Difficult to model:**
  - Sequence—e.g., Epanaphora.
  - Indexing—e.g., Anadiplosis.
  - Position—e.g., Mesodiplosis.
  - Extra-ontology knowledge—e.g., grammatical, morphological.

- **Classic AI trade-off:** Expressiveness versus tractability:
  - Limited relations make auto-classification computationally feasible.
  - Enhanced OWL, e.g., adding rules to class definition (SWRL).
    (O'Reilly & Paurobally, "Lassoing Rhetoric with OWL and SWRL")

What we need: A methodology for ontology creation

So far only tips and informal guidelines used for building ontologies.
We have been developing a formal methodology. What learned so far.

- **Method 1: Top-down:** Rule-based, rhetorician knowledge.
  - Problem: Experts disagree.
  - "Essence of a lemon" dilemma: What exactly *is* a certain figure?
  - "Forest vs woods" classification issue: Boundary of a figure?

- **Method 2-1: Middle-out:** Work from a "cluster" in the ontology.
  - We chose the **Chiastic Suite**.

- **Method 2-2: Bottom-up:** Build up ontology from real instances.
  In effect, "learning from data" (human as Machine Learning!!).
  - Worked very well up to a point.
  - Challenge: What is prototypical instance of a figure?
  - Challenge: What to do with "pseudo" instances or exceptions?
  - Challenge: Stopping threshold?—avoid "overfitting".

Discussion: Formalize rhetoric ontology engineering?

- Collaborative development?:
  - Creation of standardized annotated corpora of figures.
  - Coordination among ontology builders working on different clusters of figures—**Web Protege** supports group collaboration.
- **"Semantic"** Machine Learning?:
  - Informed feature engineering.
  - Integration of heterogeneous Semantic Web knowledge sources —**"Linguistic Linked Open Data" (LLOD)**—with our ontology.
- Emerging AI paradigm—**"Cognitive Computing"**:
  - Cooperation between rhetoricians **and** Machine Learning.
- **???**