

Complex categories in ontologies

Alexandra Arapinis & Laure Vieu

LOA-ISTC-CNR, Trento & IRIT-CNRS, Toulouse

`aarapinis@gmail.com & vieu@irit.fr`

LogOnto Workshop at FOIS 2014

1 Inherent polysemy, a linguistic phenomenon

- **inherent nominal polysemy**³ \subset **logical polysemy**² \subset **systematic polysemy**¹
(Pustejovsky's terminology)

¹ Systematic patterns of conceptual relations (e.g. metonymy)

CLIENT/ORDER: the ham-sandwich left without paying

² Lexicalized complementary senses with overlapping, dependent, or shared meanings (e.g. selectional polysemy)

PHYSICAL OBJECT/SUBSTANCE: I bought an apple vs. You have apple on your shirt

³ Inherent senses constitutive of the complex word-meaning, definitional of the kind of entity denoted

TOME/TEXT: the book is on the table vs. the book is complex

2 Linguistic tests for inherent polysemy

- **Copredication** tests show that there is a single referent or ontological correlate despite the predicates' contradictory selectional restrictions

*This thick **book** is incomprehensible:* physical object and information object

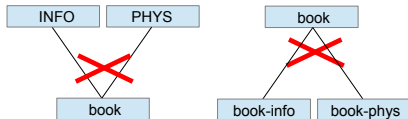
*The **inflammation** is acute and visible to the naked eye:* process and physical object

*The **university** in the city center specializes in humanities:* building and institution (and staff)

*Brazil is a large two-century-old portuguese-speaking **country**:* land, institution, people

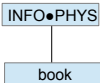
3 What inherent polysemy isn't

- It is **not conjunction**: categories of “aspects” are most often disjoint, conjunction would yield inconsistency
In a taxonomy of classes representing nouns, multiple inheritance cannot do
- It is **not disjunction**, not a simple polysemy: aspects are present “together” and selecting one aspect with a predication doesn't rule out the other aspect
In a taxonomy of classes representing nouns, subsumption cannot do



- Pustejovsky (1994) and Asher (2011) argue that “dual-aspect” nouns denote entities having **complex types** (or dot types) called **dot objects**

book is of type $INFO \bullet PHYS$



4 Is this a purely linguistic phenomenon?

- Concepts are often accessed through language: polysemy is known to affect ontology construction, even when the ontology is not built from texts.
- Ill-defined classes associated to nouns presenting standard inherent polysemy found in many ontologies
 - ▶ *Places* in DBpedia: covers buildings, countries, lakes...
 - ▶ *Diabetic Cataract* in UMLS: both disease (process) and anatomical abnormality (physical object)
- Is the issue more serious with systematic polysemy? Can we eliminate it by a careful identification of well-defined “simple” categories?
 - ▶ What to do with **individuals**? 0 “*Brazil*” and 3 *Brazil-territory*, *Brazil-state*, and *Brazil-nation* or 1 “*Brazil*” arbitrarily picked among these?
 - ▶ What to do with **properties that apply to several aspects** at once?
What do you read: the information or the physical artifact?
What do you diagnose and treat: the physical symptom or the disease?

5 Ontological grounds for dot objects and complex categories

- Language is often arbitrary. But no arbitrariness for inherent polysemy like for homonymy (*bank* – financial institution / *bank* – river side).
No two senses can be glued up to refer to dot-objects at will
- (Arapinis 2013) argues that inherent polysemy arises only when there are **dependence** relations involved, and that the dot-object is related to its aspect components through a kind of **constitution** relation (Fine, Baker)
 - ▶ Rigid Existential Dependence
 $x \text{ RED } y =_{df}$ Necessarily, x exists only if y exists
This book RED its information content
 - ▶ Generic Existential Dependence
 $x \text{ GED } F =_{df}$ Necessarily, x exists only if some F exists
The University of Rio GED professors, students, etc.
The information content of this book GED its copies

6 Beyond material constitution

- Traditionally, constitution requires vertical material coincidence between the constituting and constituted entity
- The constitution of a dot entity places a horizontal requirement of coincidence between aspects (the vertical is derivatively obtained)
Concrete categories glued only when spatio-temporal **coincidence** occur
 - ▶ *country*: people (usually) located on land
 - ▶ *university, newspaper*: staff (usually) working in building
 - ▶ *newspaper*: institutional building and paper copy **cannot** be glued together
- Coincidence can be extended to cover abstract categories as well
 - ▶ *book, newspaper*: text physically realized on paper
 - ▶ *university*: staff acting on behalf of the institution

7 Clean ontologies need complex categories

- The question is **not how to avoid** stumbling on the linguistic phenomenon of inherent polysemy by properly identifying classes of homogeneous individuals belonging to a single top-level category
- The question rather is **how to deal with the conceptual and ontological issues** underlying this linguistic phenomena and inevitably appearing within our ontologies
 - ▶ We need to accommodate complex categories in ontologies
 - ▶ We need to accommodate complex individuals, i.e., dot objects, in ontologies
- How to do so?

8 Dot objects

- Two formal approaches to dot objects
- Asher's proposal (2011):
 - ▶ dot objects are **primary entities**, aspects are derived through a kind of *qua*-construction
 - ▶ dot objects are **not** built upon simpler pre-existing objects denoted by their separate aspects
 - ▶ Problem: book_1 *qua* Info and book_2 *qua* Info are different entities, even when there is in fact **one** information content
- Mereological account (suggested but not developed by Cooper 2006, 2007):
 - ▶ dot objects are **mereological sums** of their aspect components
 - ▶ dot-objects (of complex categories) and their aspect components (of simple categories) are citizens of the ontology on an equal footing, linked by a kind of mereological composition

9 Asher's objections to a mereological account

- Aspects are not parts, i.e., no parthood expression is able to pick them up
“Normal parts of objects have names and can be referred to. This isn't true of the inhabitants of ●-types like lunches. This should lead us to be suspicious of this view.”
- Mereology gets identity criteria wrong
More precisely, it gets counting books wrong, as it posits objects (sums) that we never count when we count books

10 Are aspects parts?

- Assumption that parts have names
 - ▶ Not generally the case. The left half of your body is certainly a part of your body but it has no dedicated name nonetheless.
- Assumption that aspects do not have names
 - ▶ The aspects of dot objects do have names in many cases e.g. *book*: *tome* (*volume, paper copy*) and *text* (*content*) (Cruse)
 - ▶ Dual aspect nouns often have specific senses dedicated to one aspect in *the novel Pride and prejudice has been translated in many languages*, *novel* doesn't refer to the info aspect of any particular dot-object novel.
- True enough, no parthood expression on dot objects picks their aspects
 - ▶ Parthood expressions do not just denote the P relation, they are much more constrained. Reciprocally, there is no reason that all occurrences of $P(x, y)$ in the world should be describable through some linguistic parthood expression.

11 Does mereology get identity criteria wrong?

- Two many sums, classical argument against fusion in **GEM**
 - ▶ A problem for the ontologist assuming Mereology strictly captures the structure of reality, not a problem for those who take it as a formal tool: No need to assume that “all that there is” is part of the domain of discourse and can be named in language
Just as we do not assume that all complex properties whose existence is posited by logic are named universals
- Asher’s argument focuses on the sums of aspects that are supposed to count as books.
 - ▶ shelf with 1 copy of Austen’s collected works, and 3 copies of the bible
 - ▶ 7 books with one copy of Austen’s collected works: 7 *INFO* objects printed on 1 *PHYS* object, would yield 7 sums of type *PHYS*•*INFO*.
 - ▶ with the 3 copies of the bible would yield a count of 10 *PHYS*•*INFO* objects on the shelf
 - ▶ but commonsense admits only either 4 or 8 books there!

12 Does mereology get identity criteria wrong?

- Asher **doesn't** count sums of Austen's novels with the physical artefacts on which the bible is printed
 - ▶ **Implicitly** applies a **principle of coincidence**
 - ▶ Not all sums of *book-PHYS* and *book-INFO* are instances of *book-PHYS●INFO*
 - ▶ **Restriction** to those entities that have been glued together by the printing process, the physical realization of the text on paper
- With this “glue” constraint made explicit, right sums
 - ▶ Each of Austen's works (*INFO* object) here **coincides** only with a **part** of the *PHYS* object (a sub-collection of its pages) on which it is printed
 - ▶ Reciprocally, the whole *PHYS* object **coincides** with only **one** *INFO* object, the collected works
 - ▶ Assuming this *INFO* object counts as an *book-INFO*, the coincidence restriction yields only 1 admissible *PHYS●INFO* sum as *book-PHYS●INFO*
 - ▶ With the 3 bibles (*book-PHYS●INFO*), this yields 4

13 Towards a mereological account of ●

- The general mereological **sum operator**: +
to be clarified if **GEM** is adequate
- **Together with** a **coincidence** relation C , and the **constitution** relation $Const$
 - ▶ various sorts of coincidence depending on the types of entities involved
 - ▶ coincidence implies a functional relationship, a sort of dependence, between (at least) one aspect component and the other(s)
lunch: participation (both ways dependence), *book*: dependence of the phys-book on the info-book
- let A and B be two types, dot-objects of the complex type $A \bullet B$:
 $\forall x (A \bullet B(x) \leftrightarrow \exists yz (A(y) \wedge B(z) \wedge Const(y+z, x) \wedge (C(y, z) \vee C(z, y))))$
- $\forall xyz ((Const(y + z, x) \wedge C(y, z)) \rightarrow (C(x, y) \wedge C(x, z)))$

14 Summing up

- Inherent polysemy is a linguistic phenomenon emerging from conceptual and ontological grounds
- This “conceptual” inherent polysemy is pervasive in existing ontologies, and even appears in those relying on foundational ontologies
- To build clean, coherent, ontologies as well as well-founded lexicons, we need to recognize the phenomenon and introduce complex categories and dot-objects in ontologies
- Mereological sum together with constitution and coincidence, two dependence relations, are “standard” formal ontology tools that appear adequate to do this
- Future work
 - ▶ Clarify the mereology needed, the constitution and coincidence relations
 - ▶ Clarify subsumption relations between dot categories