The concept of concept

- humans have concepts such as **thumb**, **walk**,**water**, **blue**, **pencil**, **two**, **employee**, **democracy**, **DNA** . . .
- humans have mental representations of these concepts (but note terminology issue: for many psychologists, the concept IS the mental representation)
- humans acquire concepts
- we want computers to (at least partially) represent (some) concepts

Some questions about concepts

- How do concepts relate to words / senses?
- Can anything be a concept? Is there a limit to the complexity of concepts?
- What does it mean to know a concept? What should a theory of concepts account for?
- How do humans learn concepts? How do concepts relate to perceptions?
- Are there different categories of concept? (**dog**, **two**, **democracy**?)
- How can we represent concepts (in computational systems)?
- Are concepts atomic? Are relations concepts?

Why should we care about concepts?

- background for research in semantics in computational linguistics and AI
- modelling semantics for psychology
- databases, semantic web etc (and their limitations)
- it's interesting . . .

Course outline 2012

1. Introduction. Informal concept representation: dictionaries, encyclopedias and folksonomies. Computational exploitation of these resources.
2. Concepts in computer science. Description logics and their use in the semantic web. Terminology databases, taxonomies and ontologies in eScience.
3. Concepts in logic and linguistics. Concepts and compositional semantics. Quantification and number in natural languages.
4. Concepts in computational linguistics. Inference and concepts. Distributional semantics and its relationship to symbolic approaches to concepts.

5 Concepts in cognitive science and philosophy. Grounding. Human concept acquisition and the innateness debate.

6 Concepts in neuroscience. Experimental evidence concerning the brain's encoding of word meaning.

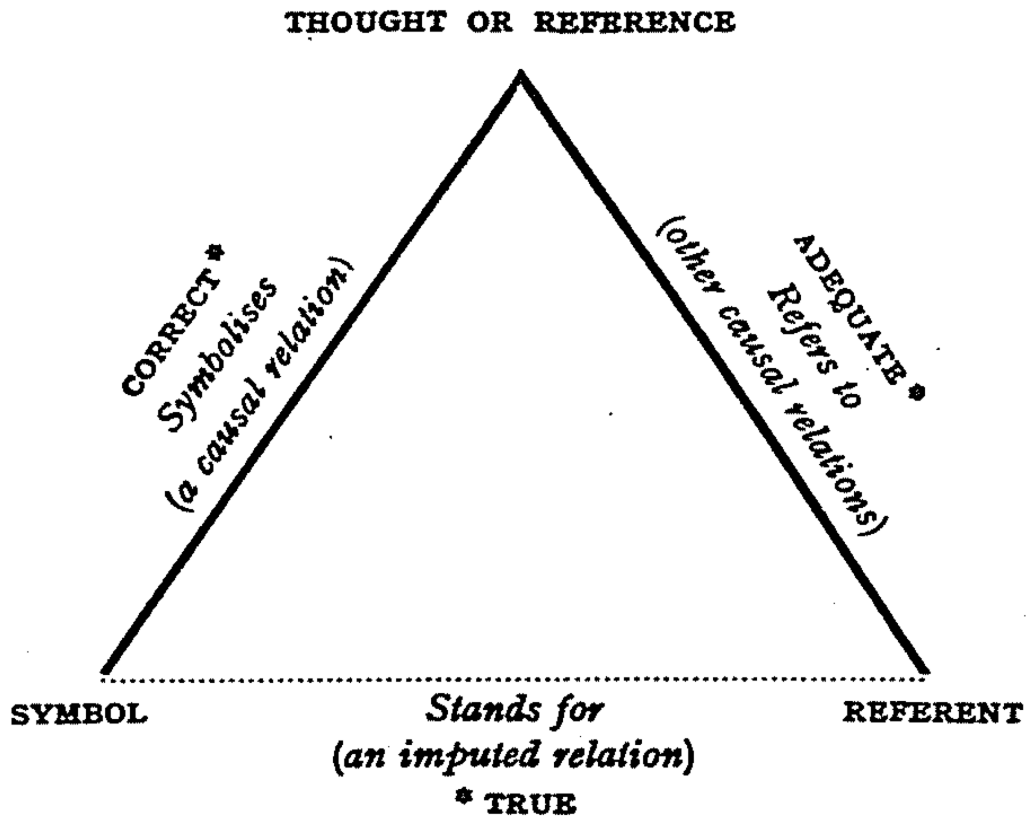
7 Open session 1

8 Open session 2

Some possible topics for open sessions

- More about folksonomy, description logics, generics, quantifiers, distributional semantics, concept acquisition (humans and computers), neural modelling etc
- Colour terms and colour categorization (linguistics/psychology/neuroscience)
- Metaphor (linguistics/psychology/neuroscience)
- Prototype theory
- Other ideas . . .

Semantic Triangle (Ogden and Richards)



Defining: traditional idea

definiendum: genus and differentia
Examples from Cambridge International Dictionary of English (CIDE):
fondant: a soft sweet made from sugar that seems to melt in the mouth
little: small in size or amount
But:
class: a group of students who are taught together . . .
mechanize: to use a machine for something that used to be done by hand.
in: (caused to be) positioned inside something, or contained, surrounded or enclosed by something.
hot: (of a person's mood) easily made worse

Lexicographic defining practice: Landau (1984)

Principles of defining:

- Avoid circularity
- Define every word used in a definition
- Define the entry word

Good practice:

- Priority of essense
- Substitutability (not a universal rule)
- Reflection of grammatical function
- Simplicity
- Brevity
- Avoidance of ambiguity

Definitions of *feather*

any of the light horny epidermal outgrowths that form the external covering of the body of birds and that consist of a shaft bearing on each side a series of barbs which bear barbules which in turn bear barbicels commonly ending in hooked hamuli and interlocking with the barbules of an adjacent barb to link the barbs into a continuous vane (Merriam Webster)

one of the very many light objects with hair-like material along each side of a long thin central part which cover a bird's body (CIDE)

Dictionaries and notions of concept

- The ‘genus and differentia’ idea is central to description logics / ontologies.
  BUT: what about terms that don’t fit into this pattern?
- The practice of enumerating word senses seems to have come from dictionaries.
- Lexicographer vs (analytic) philosopher: lexicographers are not attempting to completely describe a word’s meaning.

Extraction of ontologies from MRDs

- MRD: machine-readable version of a conventional printed dictionary.
- Most work in 1980s.
  1. Start from type-setting tape . . .
  2. Analyse meaning of font changes etc, build a database of entries.
  3. Parse definitions to extract genus term (mostly just nouns).
  4. Disambiguate genus term with respect to dictionary senses (e.g., using Lesk's method).
  5. Use links to build a taxonomy.
- and then?

Extraction of ontologies from text

- Relies on patterns in text (words, syntax or semantics, possibly combined with context). Wikipedia is a suitable source.
- Filtering necessary to get high precision: is the text really about a concept? (generics)
- Recall? Sense disambiguation?
- Moving beyond IS-A?
- Probably most useful to extend existing ontologies.

Folksonomy and museum objects

Tagging, Folksonomy and Art Museums: Results of steve.museum's research, J. Trant

4.2 A Model of Tagging Works of Art

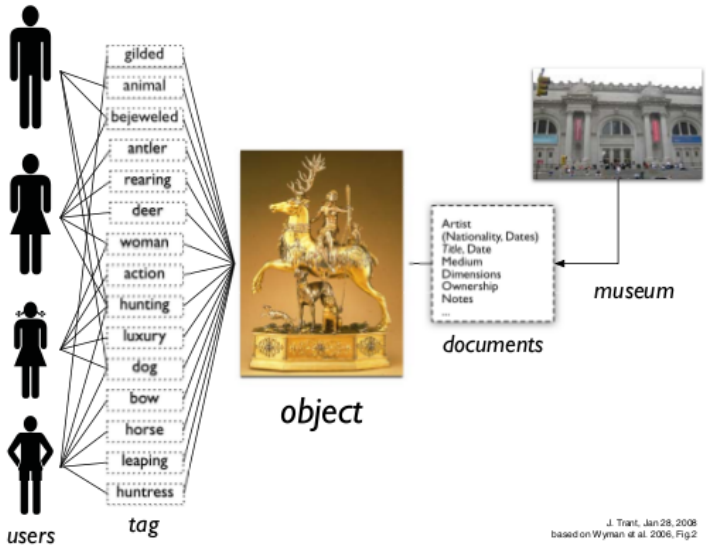


Figure 4-1. Differing perspectives / differing vocabularies: while users tag from multiple perspectives, the museum documents from a single, institutional point of view

Tags

work #	Image	# unique terms	# extended docs	possible doc hits	# docs applicable	# docs string	# docs n/a	# null doc hits	# tags assigned	# hits applicable	# hits n/a or string
148		19	11	209	23	5	16	165	21	51	87
160		46	14	644	100	6	44	494	64	303	168
670		34	6	204	43	0	3	158	54	164	3
993		16	13	208	13	0	4	191	16	14	6

Informal concept representation

- Dictionaries, WordNet, encyclopedias, folksonomy: all use words, with some amount of additional structure.
- The only available source of information for most concepts.
- Extraction for computational purposes may be complex, but fundamental problems are more about representation than e.g., syntactic parsing. (Efforts to build large-scale formal ontologies manually have not met with much success.)