Language and Concepts

Ann Copestake

Computer Laboratory
University of Cambridge

January 2012

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?

An example: animals at night

An example: animals at night

Tapetum

Log in / create account

WIKIPEDIA
The Free Encyclopedia

Article | Talk | Read | Edit | View history | Search

Tapetum lucidum

From Wikipedia, the free encyclopedia

"Eyeshine" redirects here. For the Edge Rock band, see Eyeshine (band).

The ***tapetum lucidum*** (Latin: "bright tapestry", plural ***tapeta lucida***)[1] is a layer of tissue in the eye of many vertebrate animals.[2]

It lies immediately behind or sometimes within the retina. It reflects visible light back through the retina, increasing the light available to the photoreceptors. This improves vision in low-light conditions, but can cause the perceived image to be blurry from the interference of the reflected light.[*citation needed*] The tapetum lucidum contributes to the superior night vision of some animals. Many of these animals are nocturnal, especially carnivores that hunt their prey at night, while others are deep sea animals.

Reflection of camera flash from tapetum lucidum

In darkness, eyeshine reveals this raccoon

Contents [hide]

1 Eyeshine
1.1 Blue-eyed cats and dogs
2 Classification
3 Mechanism
4 Uses
5 Pathology
6 See also
7 References

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia
Toolbox
Print/export
Languages
Català
Česky
Deutsch
Español
Français

Eyeshine [edit]

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

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.

Course outline, continued

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

How the course is intended to work

- Session 2 - 6: seminar style — discussion centered around assigned readings (see course webpage for links).
- Two hour scheduled slots, start at 14:05, finish 15:55, short break mid-session.
- Session 7 and 8: open sessions. Presentations by students. Readings etc to be decided by end of week 4.
- Assessment: 3,000 word essay (based on presentation). Deadline: April 25.
- Room TBA
- Readings vary in complexity . . .
- Action items:
  1. Email aac@cl.cam.ac.uk if you want to do the module (or just attend the sessions).
  2. Do the reading for week 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)
- Other ideas . . .

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?

WordNet

Word to search for: feather Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **feather**, plume, plumage (the light horny waterproof structure forming the external covering of birds)
  - *direct hyponym* / *full hyponym*
  - ***part meronym***
    - S: (n) vane, web (the flattened weblike part of a feather consisting of a series of barbs on either side of the shaft)
    - S: (n) quill, calamus, shaft (the hollow spine of a feather)
  - *substance meronym*
  - *direct hypernym* / *inherited hypernym* / *sister term*
  - *part holonym*
  - *derivationally related form*
- S: (n) **feather**, feathering (turning an oar parallel to the water between pulls)

Verb

- S: (v) **feather** (join tongue and groove, in carpentry)
- S: (v) **feather** (cover or fit with feathers)
- S: (v) **feather**, square (turn the paddle; in canoeing)
- S: (v) **feather**, square (turn the oar, while rowing)
- S: (v) fledge, **feather** (grow feathers) *"The young sparrows are fledging already"*

FrameNet

Frame Index

A B C D E F G H I J K
L M N O P Q R S T U
V W X Y Z

Abandonment
Abounding_with
Absorb_heat
Abundance
Abusing
Access_scenario
Accompaniment
Accomplishment
Accoutrements
Accuracy
Achieving_first
Active_substance
Activity
Activity_abandoned_sta
Activity_done_state
Activity_finish
Activity_ongoing
Activity_pause
Activity_paused_state
Activity_prepare
Activity_ready_state
Activity_resume
Activity_start
Activity_stop

Travel

Lexical Unit Index

Definition:

In this frame a Traveler goes on a journey, an activity, generally planned in advance, in which the Traveler moves from a Source location to a Goal along a Path or within an Area. The journey can be accompanied by Co_participants and Baggage. The Duration or Distance of the journey, both generally long, may also be described as may be the Mode_of_transportation. Words in this frame emphasize the whole process of getting from one place to another, rather than profiling merely the beginning or the end of the journey.

Ellen JOURNEYED to Europe with five suitcases.

Samantha JOURNEYED 2500 miles with her family by sea to China.

The Osbournes *took* a TRIP from Beverly Hills to London on the Concorde.

FEs:

Core:

Area [Area]
Semantic Type: Location
This is the Area in which the traveling takes place. This frame element describes the enclosed area inside which travelling, of unspecified Source, Path or Goal takes place.
We TRAVELLED in Europe.

Direction [dir]
Excludes: Area
The direction in which the Traveler goes.
They began their ODYSSEY north.

Log in / create account

WIKIPEDIA
The Free Encyclopedia

Article | Talk | Read | Edit | View history | Search

Tapetum lucidum

From Wikipedia, the free encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox

Print/export

Languages
Català
Česky
Deutsch
Español
Français

"Eyeshine" redirects here. For the Edge Rock band, see Eyeshine (band).

The ***tapetum lucidum*** (Latin: "bright tapestry", plural ***tapeta lucida***)[1] is a layer of tissue in the eye of many vertebrate animals.[2]

It lies immediately behind or sometimes within the retina. It reflects visible light back through the retina, increasing the light available to the photoreceptors. This improves vision in low-light conditions, but can cause the perceived image to be blurry from the interference of the reflected light.[*citation needed*] The tapetum lucidum contributes to the superior night vision of some animals. Many of these animals are nocturnal, especially carnivores that hunt their prey at night, while others are deep sea animals.

Reflection of camera flash from tapetum lucidum

In darkness, eyeshine reveals this raccoon

Contents [hide]

1 Eyeshine
1.1 Blue-eyed cats and dogs
2 Classification
3 Mechanism
4 Uses
5 Pathology
6 See also
7 References

Eyeshine

[edit]

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: see Carlson reading in session 3)
- Recall? Sense disambiguation?
- Moving beyond IS-A? (e.g., session 4, Harrington and Clark (2007))
- 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

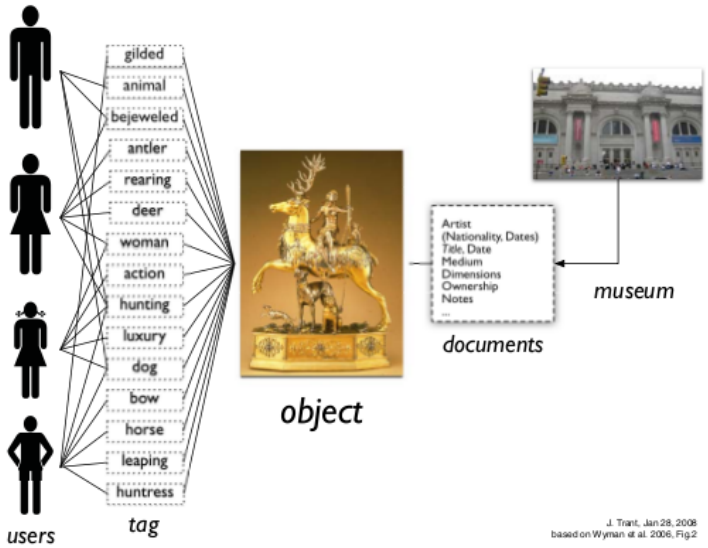

Figure 4-1. Differing perspectives / differing vocabularies: while users tag from multiple perspectives, the museum documents from a single, institutional point of view

© J. Trant / Archives & Museum Informatics | (cc) BY-NC-SA

Existing documentation

Museum Data	Definition	Req.	Example
Institution	The display version of your museum's name	y	The Metropolitan Museum of Art
Image filename	The name of the image file that matches this object.	y	ape854.jpg
Institution ObjectID	An identifying number assigned to the object in your institution, other than an accession number (e.g. the record number from your collections management system).	n	11479
Title/Object Name	The display for of the work's preferred Title or Name.	y	The Gulf Stream
Date	The display form of the work's date.	y	1899
Maker/Culture	The display form of the creator's [or creators'] name[s] (and dates if desired)	y	Winslow Homer, 1836-1910
Medium	The display form of the material, support and/or techniques used to make the work	y	oil on canvas
Dimensions	The display form of the work's dimensions	y	28 1/8 x 49 1/8 in. (71.4 x 124.8 cm)
Object Type	The type of work, e.g. painting, sculpture, installation	y	painting
Credit Line	The display form of any credit to be shown with the work of art.	y	Catharine Lorillard Wolfe Collection, Wolfe Fund, 1906
Copyright	The Display form of any copyright statement for the work of art.	n	© Jasper Johns/Licensed by VAGA, New York, NY
Accession Number	The accession number assigned to the work of art in your museum	y	06.1234
Notes	curatorial or other notes about the work.	n	*The Gulf Stream* was based upon studies made during Homer's two winter trips to the Bahamas in 1884-85 and 1898–99. First exhibited at the Pennsylvania Academy of the Fine Arts in Philadelphia in 1900, the picture was subsequently reworked and "improved" [...]

Table 6-2: steve.museum Data Guidelines. Documentation for works of art submitted to be tagged was requested from participating museums in a format comparable to that of CDWA lite. A simple CSV format was used for data transfer, and files were accompanied by a digital image file (1024 x 768 pixels) for each work.

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

Useful misperceptions

ID	Work Details	*Useful Misperception* Terms	Total
929	Arthur Garfield Dove, American, 1880-1946, That Red One, 1944, Oil and wax on canvas, 68.58 x 91.44 cm (27 x 36 in.), Gift of the William H. Lane Foundation, Boston Museum of Fine Arts	absract expressionism abstract expressionism black olive on chop sticks chopsticks donut Expressionism Eye Moon	1 1 1 2 2 1 3 2
Total Taggers: 25	Total Tags: 111 Distinct Tags: 74	Useful Misperceptions:	19
880	the Dokimasia Painter, Mixing bowl (calyx krater) with the killing of Agamemnon, about 460 B.C., Ceramic, Red Figure, Height: 51 cm (20 1/16 in.); diameter: 51 cm (20 1/16 in.), William Francis Warden Fund, Boston Museum of Fine Arts	Aegisthus amphora men and women terracotta Theater vase	1 3 1 1 2 4
Total Taggers: 11	Total Tags: 45 Distinct Tags: 32	Useful Misperception	12

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.)

A reminder: action items

1. Email aac@cl.cam.ac.uk if you want to do the module (or just attend the sessions.
2. Do the reading for week 2.