

# Computational Lexicology & Terminology Lab

VU University Amsterdam



# People

- Tomasso Caselli
- Agata Cybulska
- Marieke van Erp
- Antske Fokkens
- Paul Huygen
- Filip Ilievski
- Ruben Izquierdo
- Selene Kolman
- Alessandro Lopopolo
- Isa Maks
- Emiel van Miltenburg
- Roser Morante
- Minh Lê Ngọc
- Marten Postma
- Anneleen Schoen
- Roxane Segers
- Chantal van Son
- Hennie van der Vliet
- Piek Vossen

# Main research areas

- (Cross-lingual) language resources
- Semantic parsing
- Subjectivity and attribution modeling
- Text mining
- Entity detection
- Event modeling: event coreference, SRL, event relations
- Story lines, perspectives,

# Language resources

- Dutch Wordnet, Global WordNet Association, Global Wordnet Grid
- Combinatoric lexicon (Referentie Bestand Nederlands, RBN)
- Sentiment lexicons
- Lexicon-ontology mappings: DOLCE, SUMO, Predicate-matrix (SemLink)
- Terminology extraction, ontology learning

# Language resources

- DutchSemCor: 400K tokens, sense-annotated (Dutch)
- OpeNER: hotel & attraction reviews & news for 6 European languages
- ECB+: cross-document event coreference, English news
- WikiNews: semantically annotated news in 4 languages based on TimeML, FactBank, ACE, etc.
- BiographyNet: semantically annotated biographies

# Projects (past)

- 2006-2008: Cornetto, Stevin, Clarin-NL, combining Dutch wordnet & RBN
- 2009- 2011: KYOTO, FP7: event mining and ontology learning
- 2009-2014: SemanticsOfHistory, VU, historical event detection
- DutchSemCor, NWO, semantically annotated corpus & WSD
- 2013-2014: Open Source Dutch Wordnet, TaalUnie

# Projects (current)

- OpeNER: Opinion and entity mining for 6 languages, FP7
- NewsReader: event mining and modeling from large news streams for 4 languages, FP7
- BiographyNet: e-Science tools and methods for digital humanities
- Can we handle the news?: EYR, SurfSara, eScience, NWO
- Understanding Language By Machines: Spinoza, NWO
- Medical Trust Networks, Time will tell a different story, Political Discourse in the News: Academy Assistants, Network Institute, VU

# OpeNER

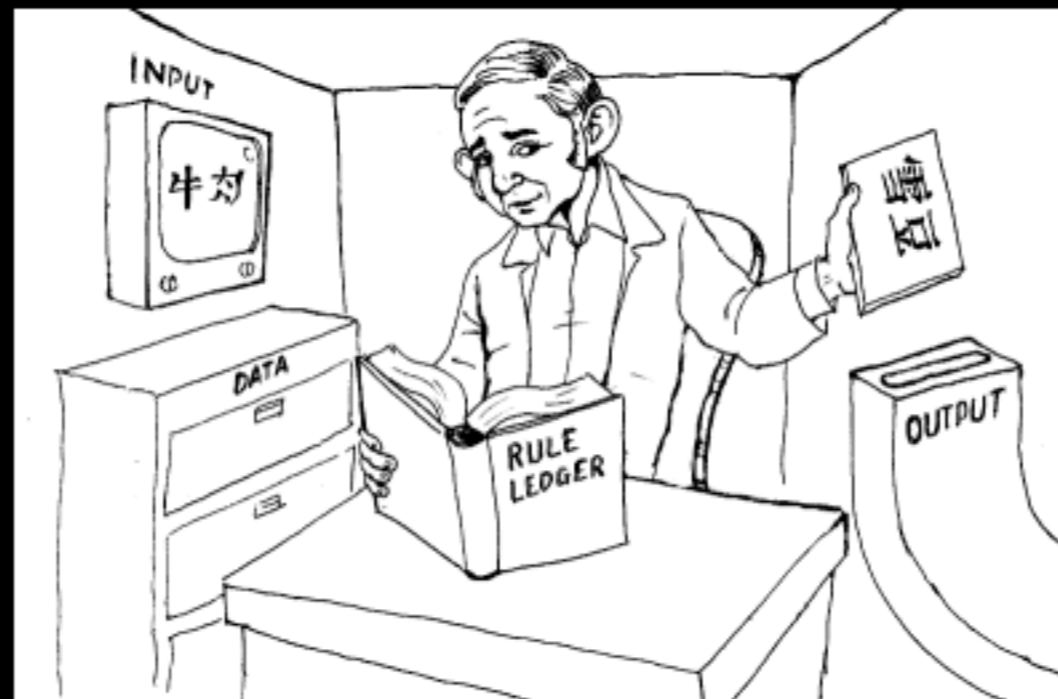
## Open Polarity Enhanced Named Entity Recognition

- Ready to use tools to perform natural language processing tasks in 6 European languages
- Free and easy to adapt for Academia, Research and Small and Medium Enterprise to integrate them in their workflow.
- Detect and disambiguate entity mentions and perform sentiment analysis and opinion detection on the texts,
- Extract the sentiment and the opinion of customers about a certain resource (e.g. hotels and accommodations) in Web reviews.
- <http://www.opener-project.eu>



# Spinoza

- Understanding Language by Machines
- Chinese room dilemma (Searl 1980)



# Spinoza

- Borders of ambiguity: Marten Postma (PhD), Ruben Izquierdo (PostDoc)
- Word, concept, perception and brain: Emiel van Miltenburg (PHD) and Alessandro Lopopolo (PhD)
- Stories and world-views (interpretation as a function of the writer) Tomasso (PostDoc) and Roser Morantes (PostDoc)
- A quantum model for meaning (interpretation as a function of the reader) Minh Lê Ngọc (PhD) and Filip Ilievski (PhD)
- Support: Selene Kolman (project manager), Paul Huygen (scientific programmer)

# (1) Borders of Ambiguity

- Relation: word  $X$  – concept  $Y$  – context  $Z$
- Qualify and quantify contexts of words & concepts → variation of  $Z$  in relation to  $X$  and  $Y$
- Define types of polysemy → variation of  $Y$  on  $X$
- Distributional properties in text: → variation on  $Z$
- Unsupervised acquisition of text that fits  $X - Y - Z$

# (2) Word, concept, perception & brain

- How language (vocabulary) cuts up the perceptual world?
- Images: bouwwerk (500), bouwsel, huis, kerk, hut, skihut, berghut, trekkershut, plaggenhut, schuilhut, vanghut, herdershut
- Movements (5,000):  
kwakken;knikkeren;jenzen;bliksemen;lazeren;mikken;kegelen;  
kieperen;flikkeren;kogelen;keilen;plompen;donderen;gooien;  
zwiepen;kukelen
- Sounds (3,000): gesjirp;geblèr;gekerm;gekrijt;gegil;gekrijs;geraas;gejubel;  
gejuich;hoerageroep;gekwek;gekwaak;gekraak;geknars;  
geklater;gespetter;gekletter
- Smell & taste (50): wildsmaak, ijzersmaak, bijsmaak, zuur, zoet, bitterzoet,  
vies, bitter, zoetzuur, chutney, mierzoet, scherp, pikant, goor

# hut or construction?



# Building, edifice

A structure that has a roof and walls and stands more or less permanently in one place; "there was a three-story building on the corner"; "it was an imposing edifice"

1421 pictures

96.88% Popularity Percentile



**building, edifice (267)**

- abattoir, butchery, shambles, slaughterhouse (0)
- apartment building, apartment house (2)
- architecture (0)
- aviary, bird sanctuary, volary (0)
- bathhouse, bathing machine (0)
- bowling alley (0)
- center, centre (9)
- chapterhouse (0)
- clubhouse, club (0)
- dormitory, dorm, residence hall, hall, student residence
- farm building (6)
- feedlot (0)
- firetrap (0)
- gambling house, gambling den, gambling hell, gaming h
- gazebo, summerhouse (1)
- government building (9)
- greenhouse, nursery, glasshouse (2)
- hall (4)
- hall (0)
- Hall of Fame (0)
- hotel (15)
- hotel-casino, casino-hotel (0)
- house (59)
  - beach house (0)
  - boarding house, boardinghouse (1)
  - bungalow, cottage (0)
  - cabin (1)
    - log cabin (0)
  - chalet (0)
  - chapterhouse, fraternity house, frat house (0)
  - country house (6)

Treemap Visualization

Images of the Synset

Downloads

ImageNet 2011 Fall Release > Structure, construction > Building, edifice

<p><b>House</b></p>	<p><b>Packinghouse</b></p>	<p><b>Hall</b></p>	<p><b>Student</b></p>	<p><b>Whorehouse</b></p>	<p><b>Feedlot</b></p>
<p><b>Hall</b></p>	<p><b>Ministry</b></p>	<p><b>Aviary</b></p>	<p><b>Rest</b></p>	<p><b>Signal</b></p>	
<p><b>Hotel-casino</b></p>	<p><b>Office</b></p>	<p><b>Architecture</b></p>	<p><b>Gazebo</b></p>	<p><b>Roman</b></p>	
<p><b>Restaurant</b></p>	<p><b>Abattoir</b></p>	<p><b>Rotunda</b></p>	<p><b>Planetarium</b></p>	<p><b>Library</b></p>	<p><b>Skyscraper</b></p>
<p><b>Place</b></p>	<p><b>Chapterhouse</b></p>	<p><b>Bowling</b></p>	<p><b>Ruin</b></p>	<p><b>Apartment</b></p>	<p><b>Tavern</b></p>
<p><b>Bathhouse</b></p>	<p><b>Observatory</b></p>	<p><b>Dormitory</b></p>	<p><b>Gambling</b></p>	<p><b>Greenhouse</b></p>	<p><b>Medical</b></p>
				<p><b>House</b></p>	<p><b>Go</b></p>

Antonio  
Torralba, Rob  
Fergus and  
William T.  
Freeman (2008)

[http://  
groups.csail.mit.  
edu/vision/  
TinyImages/](http://groups.csail.mit.edu/vision/TinyImages/)

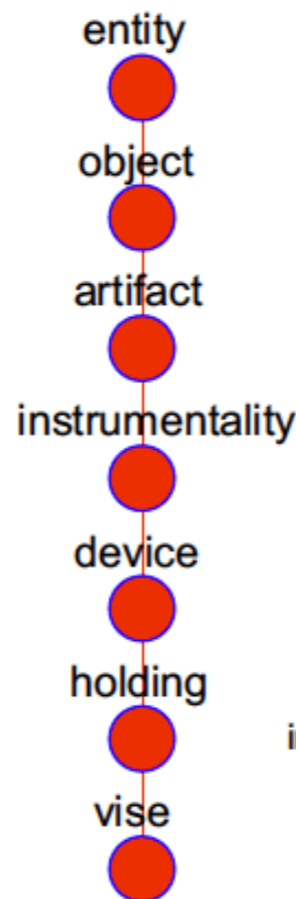
Visual words:  
pixel location  
pixel color  
linked to WordNet



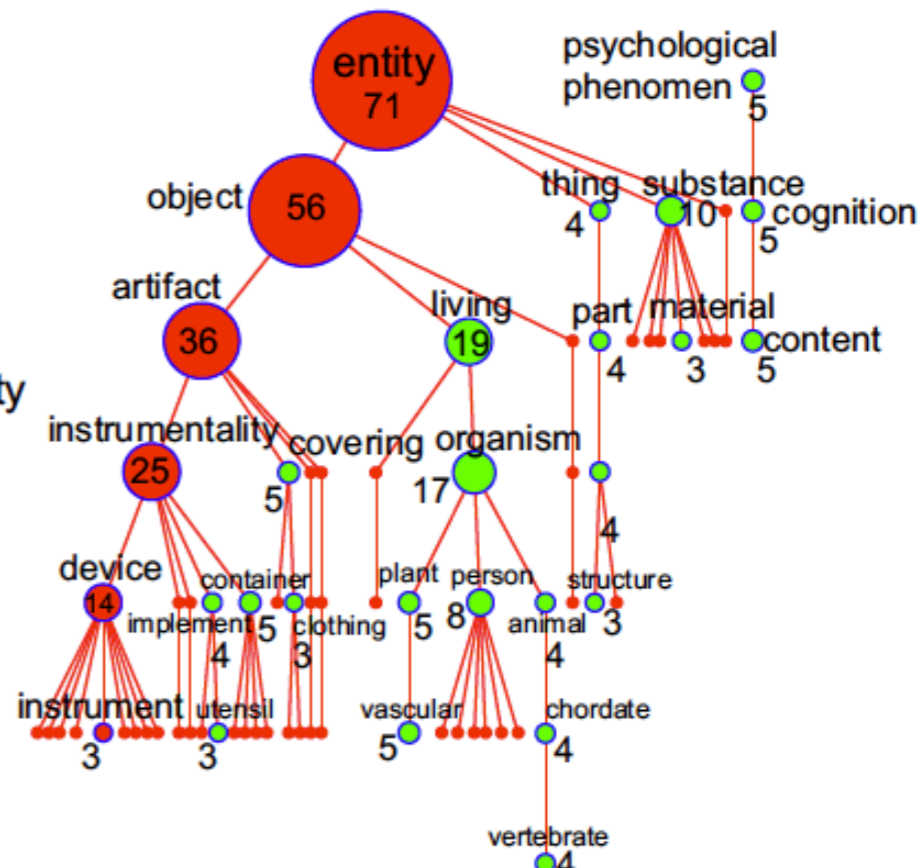
a) Input image



b) Neighbors



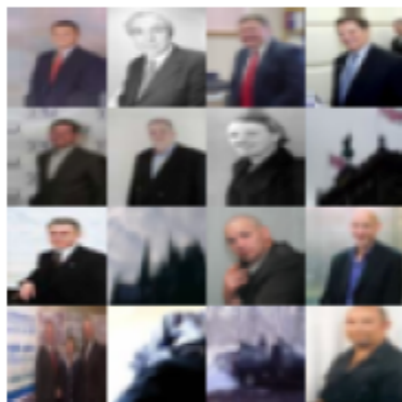
c) Ground truth



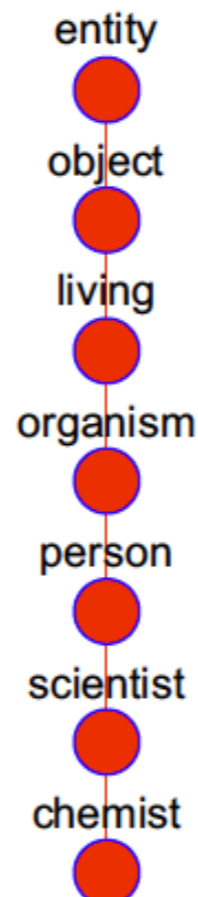
d) Wordnet voted branches



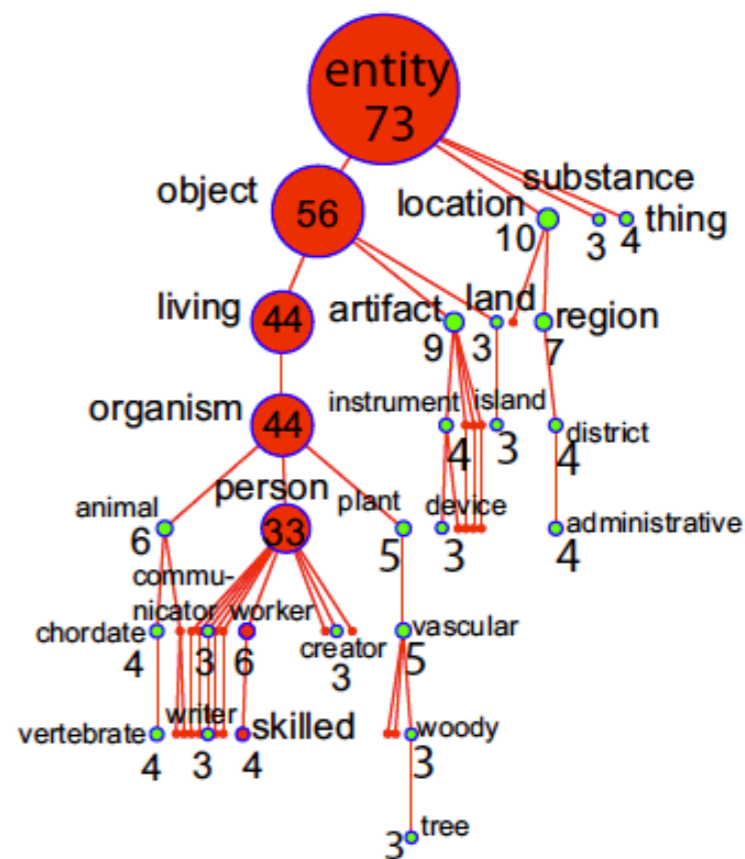
a) Input image



b) Neighbors



c) Ground truth



d) Wordnet voted branches

# parts and animals

head  
beak



tail

paw  
leg



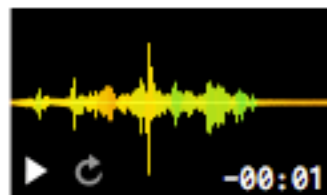
# [www.freesound.org/browse/](http://www.freesound.org/browse/)

- car: 4386
  - dog: 1389
  - cat: 790
  - water: 7846
  - paper: 1606
  - chair: 972
- eat: 518
  - sleep: 400
  - drink: 667
  - sick: 421
  - run: 2972
  - rubbing: 700

rubbing Automatic by relevance search

Show advanced search options

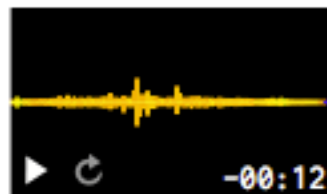
previous next 1 2 3 4 5 6 7 ... 21 | 700 sounds



**rubbing\_glass\_6.wav** ★★★★★  
1 of 5 sounds recording glass being rubbed together. 5 minute effort, will probably follow this up with different selections ...  
glass harsh rubbing rubbing-glass

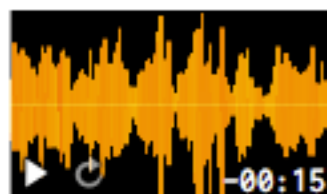
+ 5 more results in the same pack "rubbing glass"

**harri**  
October 31st, 2005  
410 downloads  
1 comment  
@ 🎧 📄 ⓘ



**Rub Hands** ★★★★★  
I rub my hands together.  
hands rubbing MTC500-M002-s13

**retroptam**  
January 27th, 2013  
216 downloads  
0 comments  
@ 🎧 📄 ⓘ



**rubbing paper 1.wav** ★★★★★  
rubbing paper together with hands.  
hands paper rubbing

**thecityrings**  
November 16th, 2013  
37 downloads  
0 comments  
@ 🎧 📄 ⓘ

+ 1 more result in the same pack "BMS"

## licenses

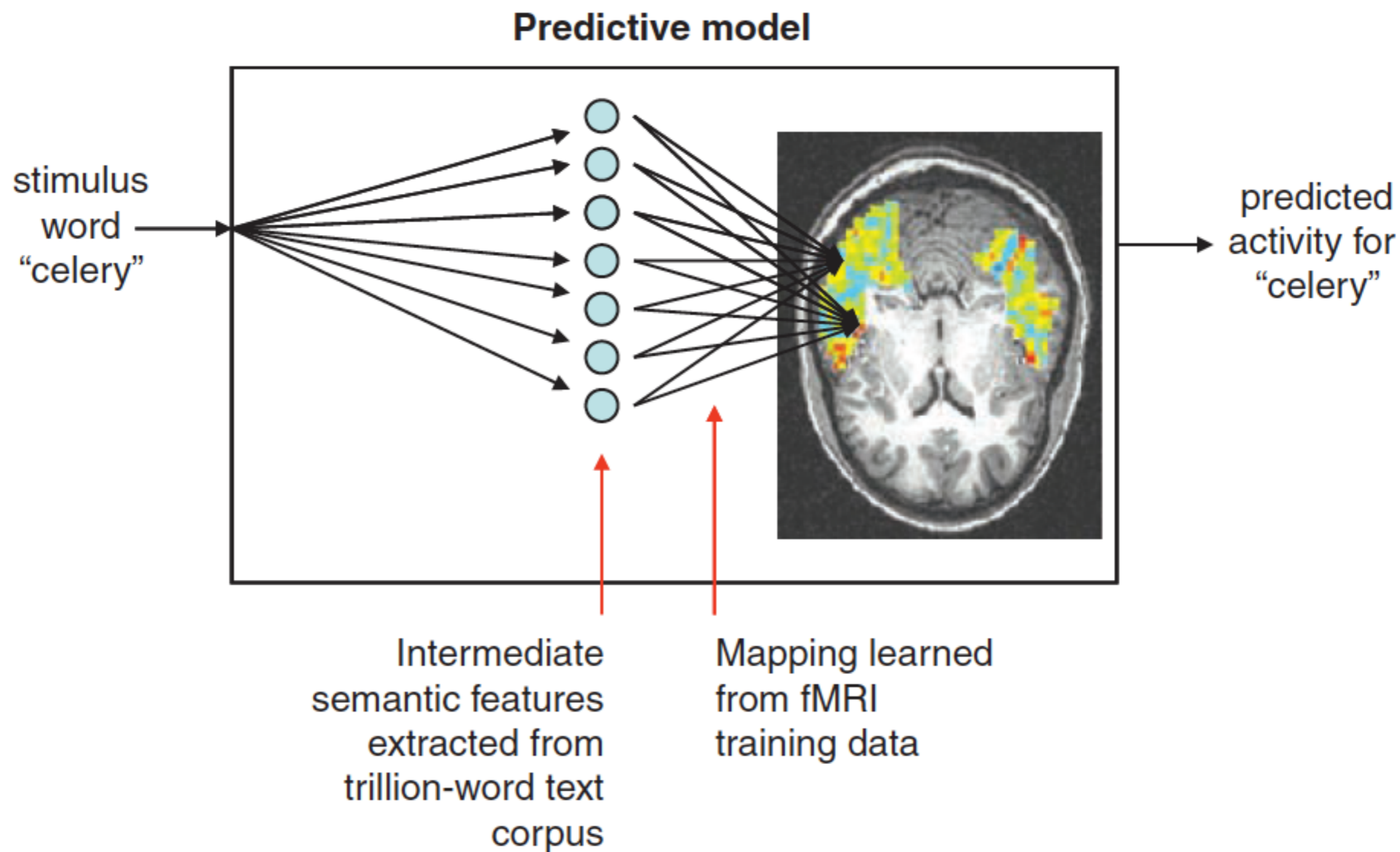
- Attribution (356)
- Attribution Noncommercial (44)
- Creative Commons 0 (261)
- Sampling+ (39)

## tags

alert audio cd contact cup dark disc dream group hand hands impact industrial knife metal mic microphone monster phone ping plastic pocket reverb ring rub rubbed rubbing vibrate vibrating weird

## type

- aif (58)
- aiff (128)
- flac (30)
- mp3 (36)
- wav (441)



**Fig. 1.** Form of the model for predicting fMRI activation for arbitrary noun stimuli. fMRI activation is predicted in a two-step process. The first step encodes the meaning of the input stimulus word in terms of intermediate semantic features whose values are extracted from a large corpus of text exhibiting typical word use. The second step predicts the fMRI image as a linear combination of the fMRI signatures associated with each of these intermediate semantic features.

# (3) Stories & world views

- NewsReader builds a history recorder:
  - computer 'reads' daily news streams
  - What happened, where, when and who is involved
  - Compares this the news of yesterday, last week, month, year
- Reconstruction of histories told by thousands of sources over longer periods of time

# (3) Stories & world views

- We create a narrative on the changes in the world:
  - colored by our view on the world
  - with a plot that explains these changes: causes, intentions, motivations, guilty, victim, 'the good & bad'
- Perspective of the writer or source:
  - automatic mining of world views
  - relate to the way the story is told

# Perspectives

Presupposition

Europe's dependency on single energy supplier (Russia)

reuters.com  
Merkel says U.S. shale gas might help Europe diversify energy

21-March-2014



Angela Merkel

-

+

energy

energy diverse



non-diverse

2006: Governments in Europe are also

concerned about increasing reliance on imported energy, especially Russian gas....

Ms Merkel said energy should play a greater role in defining the EU's international

www.ft.com

# (4) Quantum model of meaning

- Language technology uses a 'pipeline' architecture:
  - words (5%error) → lemmas (10%e) → grammar (25%e) → meaning (40%e) → relations (40%e)
- Stacking of errors and no back tracking

# (4) Quantum model of meaning

- Quantum meaning is the results of the way of reading:
  - All possible interpretations are left open
  - Determine the fit and coherence of each interpretation from a world view and storyline
  - How does a doctor and a patient read the same brochure?
- Complexity of information layers asks for new forms of representation, reasoning and decision system.
- Linked-Open-Data cloud as a background model