Distributional representations of concrete nouns for action verbs disambiguation

Irene Russo ILC CNR, Pisa

Linking Data, Ontologies and Distributional Models for the Representation of Lexical Meaning
Lablita, June 2017 Florence

Outline of the talk

- Introducing ModelAct (http://modelact.lablita.it/) and its dataset
- Reporting on two experiments with different source of information
  - first experiment: distributional semantic representations vs physical properties of objects
  - second experiment: distributional semantic representations vs visual features from pictures
- Final comments and conclusions

Word sense disambiguation in a nutshell

- In computational linguistics word sense disambiguation concerns the automatic identification of which sense of a word is used in a sentence when the word has more than one meaning

From WordNet

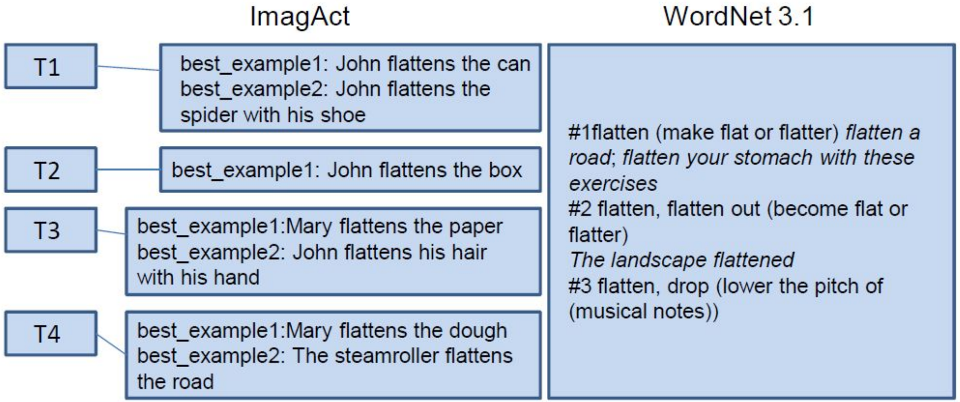
Cut the rope. →

- **cut (separate with or as if with an instrument)**
- reduce, cut down, cut back, trim, trim down, trim back, cut, bring down (cut down on; make a reduction in)
- swerve, sheer, curve, trend, veer, slue, slew, cut (turn sharply; change direction abruptly)
- cut (make an incision or separation)
- cut (discharge from a group)

...

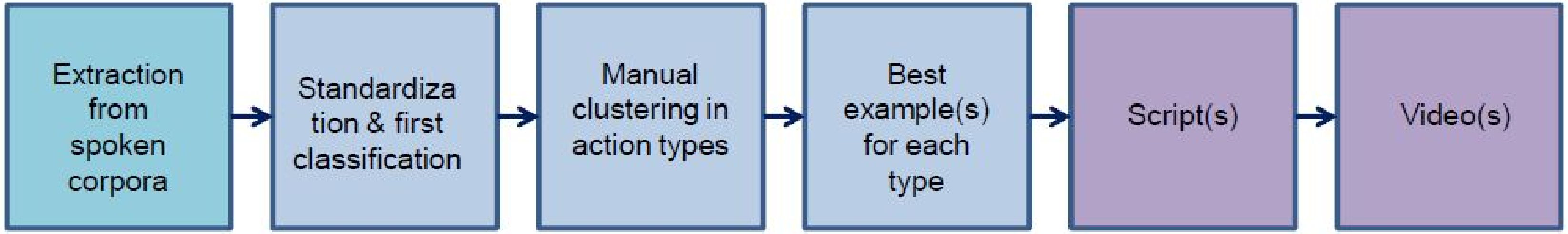
Word sense disambiguation in ModelAct

- One of the aim of ModelAct is the automatic disambiguation of action verbs
- Action verbs in ModelAct are not like in WordNet, senses' distinctions are derived from
  - the kind of (body) movement(s) involved is essential
  - different concrete objects in theme position



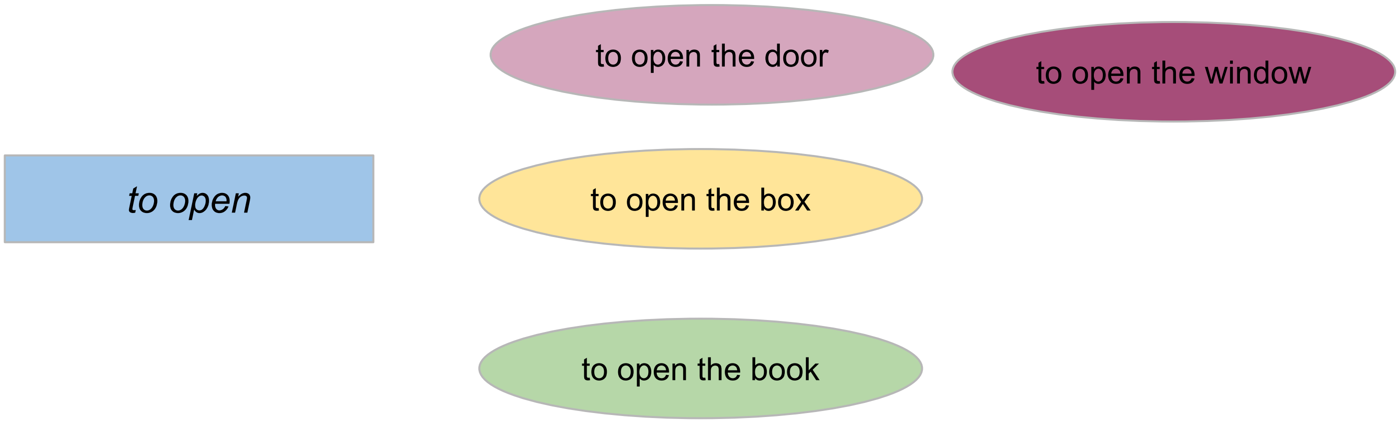
ImagAct dataset

- focus on very frequent action verbs (800 lexical entries in Italian and English and more than 3000 objects' mentions)
- manual annotation to induce basic action types from corpora
- have a look at www.imagact.it



Open issue: one verb, more action concepts

- a verb like *to open* can refer to different procedural sequences of sub-actions (different sequences of body movements performed by an agent) depending on the features of the objects
- *opening a box* is different with respect to *opening the book* or *the door*



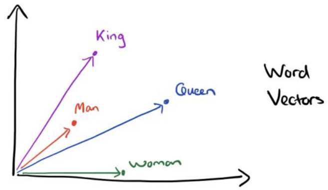
The meaning of the verb as a black box

- sense distinctions in ImagAct dataset could depend on the internal structure of verbs
- unfortunately, complete lexical resources reporting these distinctions are not available
- focus on objects in basic sentences: *John hammers the metal*, *John washes the bottle* etc.

Agent *Verb* ***Object***

Two psychological concepts in the toolbox

- artifactual categories as **situated conceptualization** where physical and situational properties meet (Barsalou 2002)
    - situational properties describe a physical setting or event in which the target object occurs (as *grocery store, fruit basket, slicing, picnic* for *apple*)
- Idea implemented in computational linguistics: similarity between concepts and conceptual categorizations (Erk, 2012; Turney and Pantel, 2010) reframed as similarity between vectors in distributional semantic models
    - two nominal concepts are similar and can be clustered in the same group if the corresponding lexemes occur in comparable linguistic contexts



Two psychological concepts in the toolbox

- notion of **affordance** as possibilities for actions that every environmental object offers (Gibson 1979)
  - affordance is quality of an object that enables an action: it concerns the relation between a perceptual property of the object and what an agent can do with it
  - humans can judge if something is do-able on the basis of perceptual information (Warren 1984)
  - Eleanor Gibson (2000, 2003): affordances are distinctive
- In a corpus study affordance verbs as verbs denoting the most distinctive actions performed on a specific object can be automatically extracted with measures of semantic associations (Russo et al. 2013)

What can be found in distributional semantic representations?

- (implicit) knowledge about concrete features of the noun?
- (implicit) knowledge about the events/actions in which the objects occurs?
- is it possible to combine distributional semantic representations with other kind of knowledge (i.e. visual features)?

The role of objects similarity

- nouns denoting concrete objects can be ordered according to several similarity measures
  - distributional semantics similarity
  - average weight and dimensions
  - visual similarity
- similarity between nouns denoting objects is central for disambiguation experiments

Experiment 1: classifying objects grasping possibilities

- Can grasping possibilities for concrete objects be automatically classified?
- Do we need encyclopedic knowledge (from distributional semantics models) or information about average dimensions/weights of artifacts?

- One-Hand_Grasp
- Two-Hands_Grasp
- Grasp_by_part
- Grasp_with_instrument_container

Experiment 1: classifying objects grasping possibilities

- For 168 nouns manually annotated by two annotators according to 4 categories (one hand grasp, two hands grasp, grasp by part, grasp with an instrument):
    - distributional semantics information from two corpora (GoogleNews and instructables.com) obtained with word2vec toolkit (Mikolov et al. 2013);
    - average dimensions (height, length and depth) for each object, obtained crawling at least 15 pages per object from amazon.com;
    - average weight for each object, obtained crawling at least 15 pages per object from amazon.com.

Distributional semantics vs physical properties

- Support Vector Multi-Classification is based on LibSVM software (Chang and Lin 2001) in WEKA with 10 fold cross-validation

features	precision	recall
instructable.com	0.113	0.336
GoogleNews	0.113	0.336
weight	0.364	0.406
dimensions	0.413	0.517
weight+dimensions	0.561	0.531
affording parts	0.25	0.399
instructables.com+all	0.443	0.552
GoogleNews+all	0.458	0.559

Results for 6 classes.

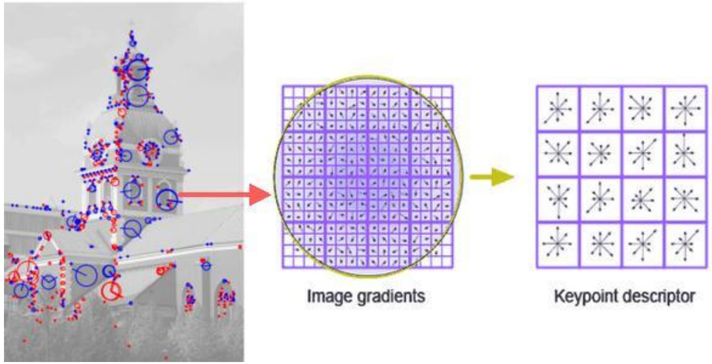
features	precision	recall
GoogleNews	0.846	0.846
weight	0.714	0.714
dimensions	0.851	0.846
weight+dimensions	0.831	0.802
affording parts	0.63	0.615
GoogleNews+all	0.846	0.846

Results for 2 classes (one hand, two hands).

- Best precision/recall from information about average weight/dimensions
- Mixed features (situational and physical properties) don't improve the classifier’s performance when combined

Second experiment: clustering the objects you can open

- testing the role of encyclopedic knowledge and visual features
- encyclopedic knowledge: context-predictive semantic vectors (word2vec) trained on GoogleNews
- visual features: Bag-of-Visual-Words (BoVW) from SIFT (Bruni et al. 2012)
  - labeled images from ImageNet
  - low-level features from SIFT that capture part of objects
  - each concept has a vector of 200k dimensions



Second experiment: clustering the objects you can open

- manual clusters for *to open* from ImagAct:

  open_1564, door, gate, car, window, building, shutter, gown

  open_1567, pack, pit, bottle, lid, letter, mail, packet, desk, cheque, hole, present

  open_1568, bandage, scroll, book, card

  open_1668, pin, binder, lock

  open_3616 ,nut

- 106 manually annotated nouns that are theme from parsed instructables.com corpus (*biscuit, fridge, tube,* etc...)

Distributional semantics vs SIFT

- context-predictive semantic vectors (word2vec) trained on GoogleNews, SIFT from Bruni et al. 2012, mix of GoogleNews and SIFT (with SVD), clustering with Cluto (k-1 repeated bisections, cosine similarity)

  homogeneity: all the clusters contain only data points which are members of a single class

  completeness: all the data points that are members of a given class are elements of the same cluster

	homogeneity	completeness
GoogleNews	0.57	0.59
SIFT BoVW	0.33	0.26
GoogleNews+SIFT BoVW	0.36	0.32

Similarity between nouns in a basic DSM

- instructables.com corpus (20 million tokens), basic DSM, most relevant nouns in obj position from English TenTen corpus (19 billion words)
- cosine distance with the noun that is in the best example in ImagAct
- 5 verbs with one basic action type, 5 verbs with two basic action types
- One type: to braise (steak), to measure (length), to scrub (floor), to sprinkle (powder), to stir (soup)
- Two types: to crush (insect, ginger), to dust (cake, furniture), to mop (floor, counter), to mould (clay, candle), paint (picture, wall)

Similarity between nouns in a basic DSM

verb	**object**	**instructables.com**	**word2vec**
to braise	steak	rib 0.9077 cabbage 0.8715 pork 0.9320 oxtail not found beef 0.8929 **fennel 0.5531**	rib 0.4441 cabbage 0.412 pork 0.4350 oxtail 0.582 beef 0.512 fennel 0.4211
to sprinkle	powder	salt 0.890 cinnamon 0.8765 cheese 0.7600 flour 0.9021 **glitter 0.5414** **dust 0.577** sugar 0.8683	salt 0.3304 cinnamon 0.3488 cheese 0.3087 flour 0.4283 glitter 0.3121 dust 0.3068 sugar 0.3242

Similarity between nouns in a basic DSM

to paint	picture portrait 0.8727 mural 0.6586 nail 0.9045 ceiling 0.8845 landscape 0.8801 face 0.9494 scene 0.9218	wall portrait 0.8617 mural 0.669 nail 0.9412 ceiling 0.9451 landscape 0.8885 face 0.9538 scene 0.9179
to crush	insect skull 0.9879 garlic 0.9819 grape 0.9691 ice 0.9912 clove not found stone 0.9912 tablet 0.9977	ginger skull 0.9981 garlic 0.9996 grape 0.9994 ice 0.9965 clove not found stone 0.9965 tablet 0.9890

Conclusions

- We can find conceptual knowledge concerning objects in language looking at nouns in distributional semantics models
- Mixed models (information about physical/visual properties of objects + encyclopedic knowledge) don't improve the results
  - Is it a problem of integration?
- The experiments so far are about objects similarity out-of-context
- More complex models are possible: DSMs that combine vectors in a compositional way (Baroni and Zamparelli 2010)
- Future work: what is in the black box of verbal semantics?
  - action primitives as action terminals combined hierarchically into a temporal sequences of actions of increasing complexity (Pastra & Aloimonos 2012)

    Cut the bread: extend hand1 - grasp with hand1 knife - cut with knife bread

Thank you :)