Disambiguating action verbs

the role of objects

Irene Russo ILC-CNR Pisa

MODELACT Conference “ACTION, LANGUAGE and COGNITION“
June 2016 Rome

Outline of the talk

- Word sense disambiguation: some basic concepts
- Introducing ModelAct (http://modelact.lablita.it/) and its dataset
- Reporting on two experiments with different source of information
  - first experiment: encyclopedic knowledge vs physical properties of objects
  - second experiment: encyclopedic knowledge 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

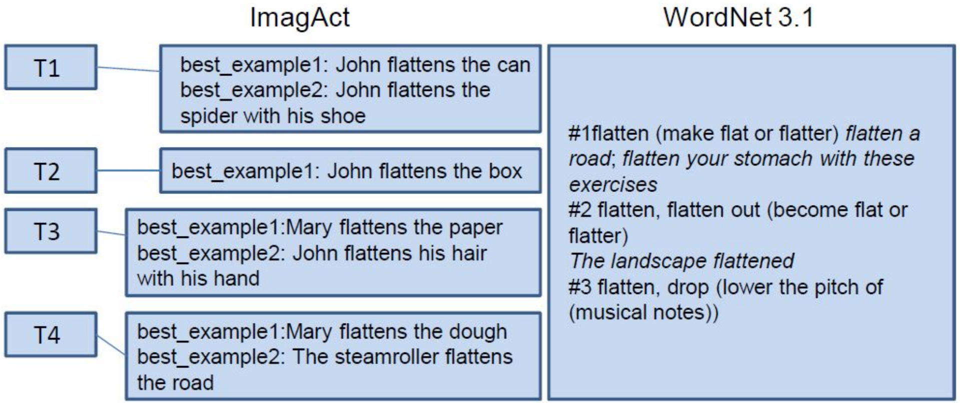
Cut the rope. →

- **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



T1: best_example1: John flattens the can
best_example2: John flattens the spider with his shoe

T2: best_example1: John flattens the box

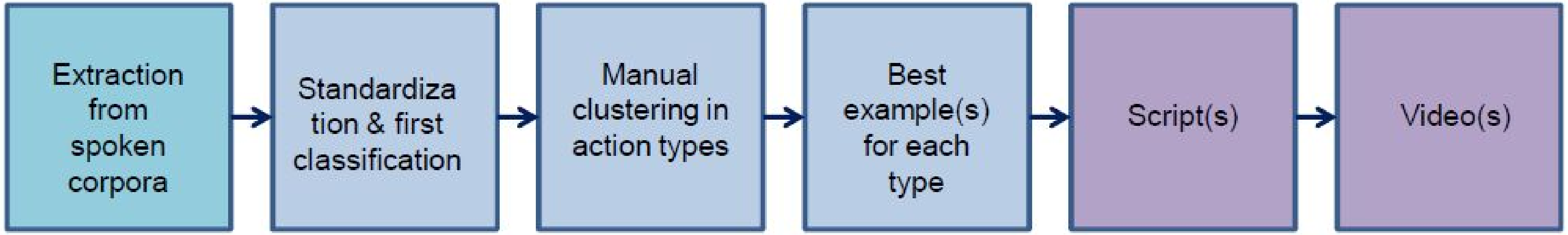
T3: best_example1:Mary flattens the paper
best_example2: John flattens his hair with his hand

T4: best_example1:Mary flattens the dough
best_example2: The steamroller flattens the road

WordNet 3.1:
#1flatten (make flat or flatter) *flatten a road; flatten your stomach with these exercises*
#2 flatten, flatten out (become flat or flatter)
The landscape flattened
#3 flatten, drop (lower the pitch of (musical notes))

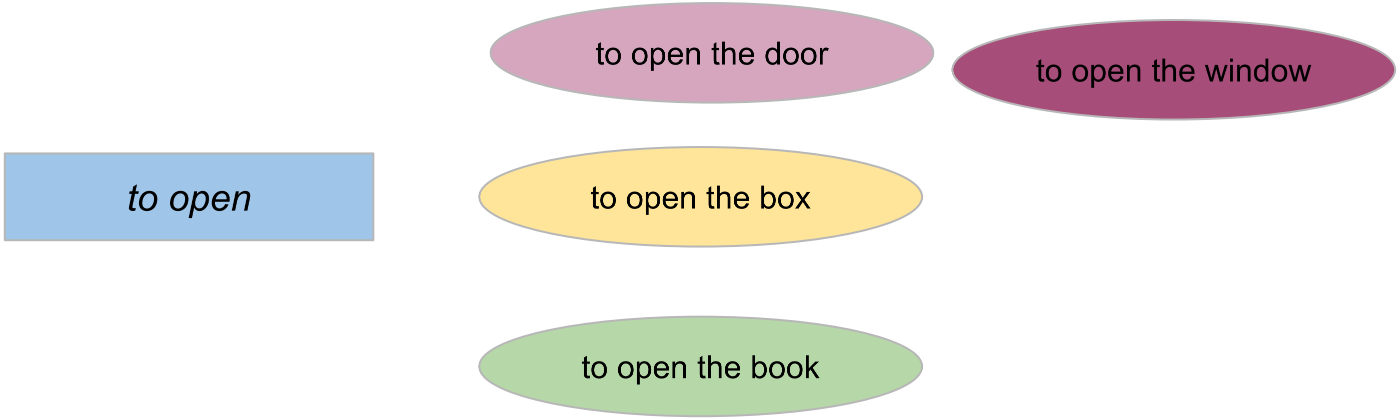
ImagAct dataset for word sense disambiguation

- 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


Extraction from spoken corpora
Standardiza tion & first classification
Manual clustering in action types
Best example(s) for each type
Script(s)
Video(s)

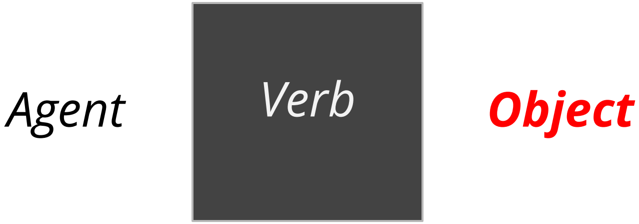
Open issue: one verb, more action concepts

- 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

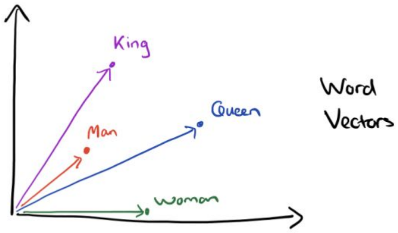
- sense distinctions in ImagAct dataset could depend on the internal structure of verbs
- unfortunately, complete lexical resources reporting these distinctions are not available
- focus on objects in basic sentences: *John hammers the metal, John washes the bottle* etc.



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: concepts' similarity 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)

The role of objects' similarity

- nouns denoting concrete objects can be ordered according to several similarity measures
    - distibutional 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?

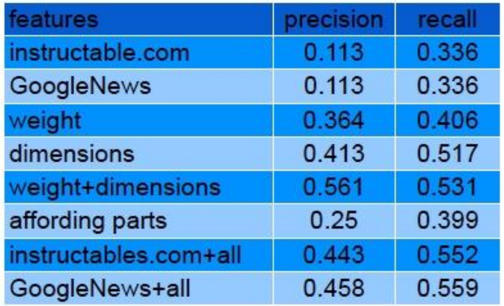
- One-Hand_Grasp
- Two-Hands_Grasp
- Grasp_by_part
- Grasp_with_instrument_container

Experiment 1: classifying objects' grasping possibilities

- For 168 nouns manually annotated by two annotators according to 4 categories (one handed grasp, two handed grasp, grasp by part, grasp with instrument):
  - distributional semantics information from two corpora (Google News 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.

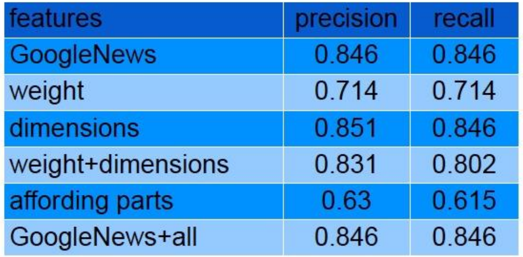
Encyclopedic knowledge vs physical properties

- 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.



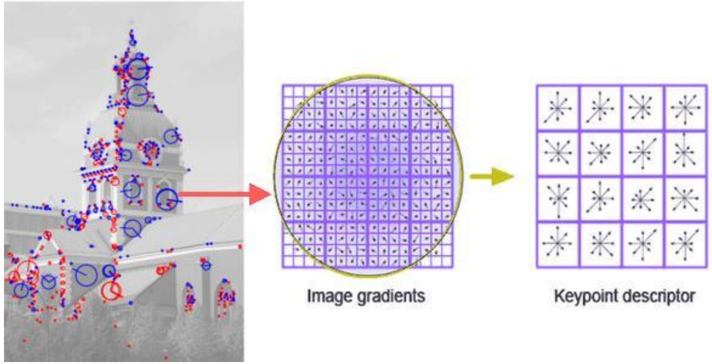
features	precision	recall
GoogleNews	0.846	0.846
weight	0.714	0.714
dimensions	0.851	0.846
weight+dimensions	0.831	0.802
affording parts	0.63	0.615
GoogleNews+all	0.846	0.846

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

- Best results with encyclopedic knowledge

Conclusions

- We can find encyclopedic 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
- 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 :)