



SAPIENZA  
UNIVERSITÀ DI ROMA

# *Fatality* Killed the *Cat* or: BabelPic, a Multimodal Dataset for Non-Concrete Concepts

Agostina Calabrese, Michele Bevilacqua, Roberto Navigli  
{calabrese.a, bevilacqua, navigli}@di.uniroma1.it

Sapienza NLP Group @ Computer Science Department  
Sapienza University of Rome, Italy  
[nlp.uniroma1.it](http://nlp.uniroma1.it) |  @SapienzaNLP



Supported by the ERC Consolidator Grant MOUSSE No. 726487



# Motivations

Existing resources have a focus on concrete, tangible, things...

IMAGENET



flickr



# Motivations

Existing resources have a focus on concrete, tangible, things...

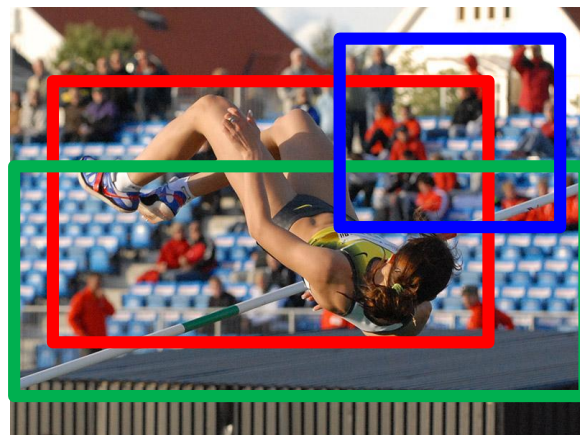
IMAGENET



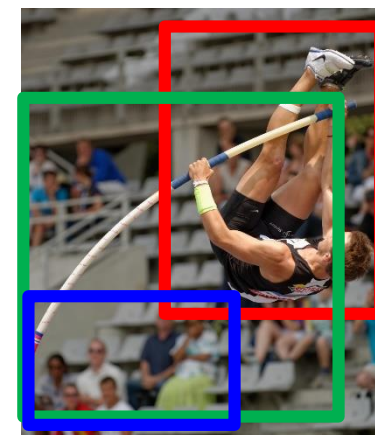
flickr



..but many non-concrete concepts are not well described by a bag of objects!



High jump



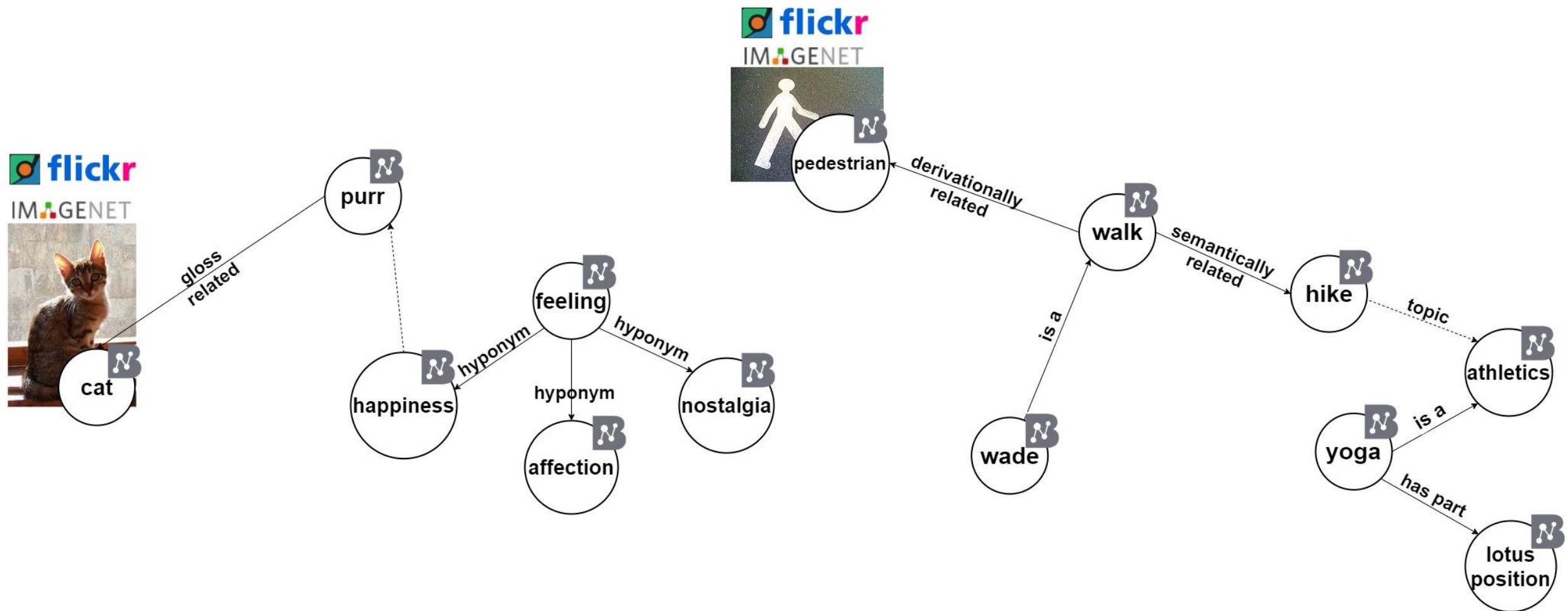
Pole vault



# BabelPic

New resource associating **concepts** with images:

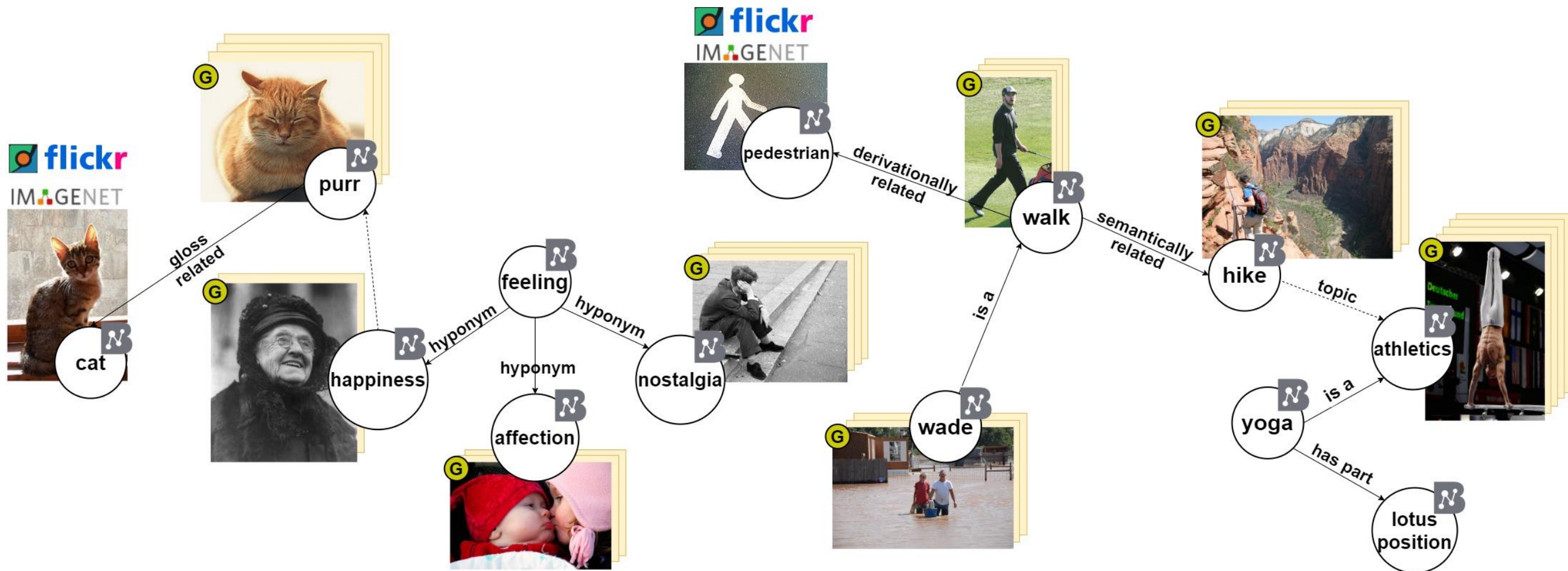
- Focus on **non-concrete** concepts
- Linked to WordNet and BabelNet



# BabelPic

New resource associating **concepts** with images:

- Focus on **non-concrete** concepts
- Linked to WordNet and BabelNet

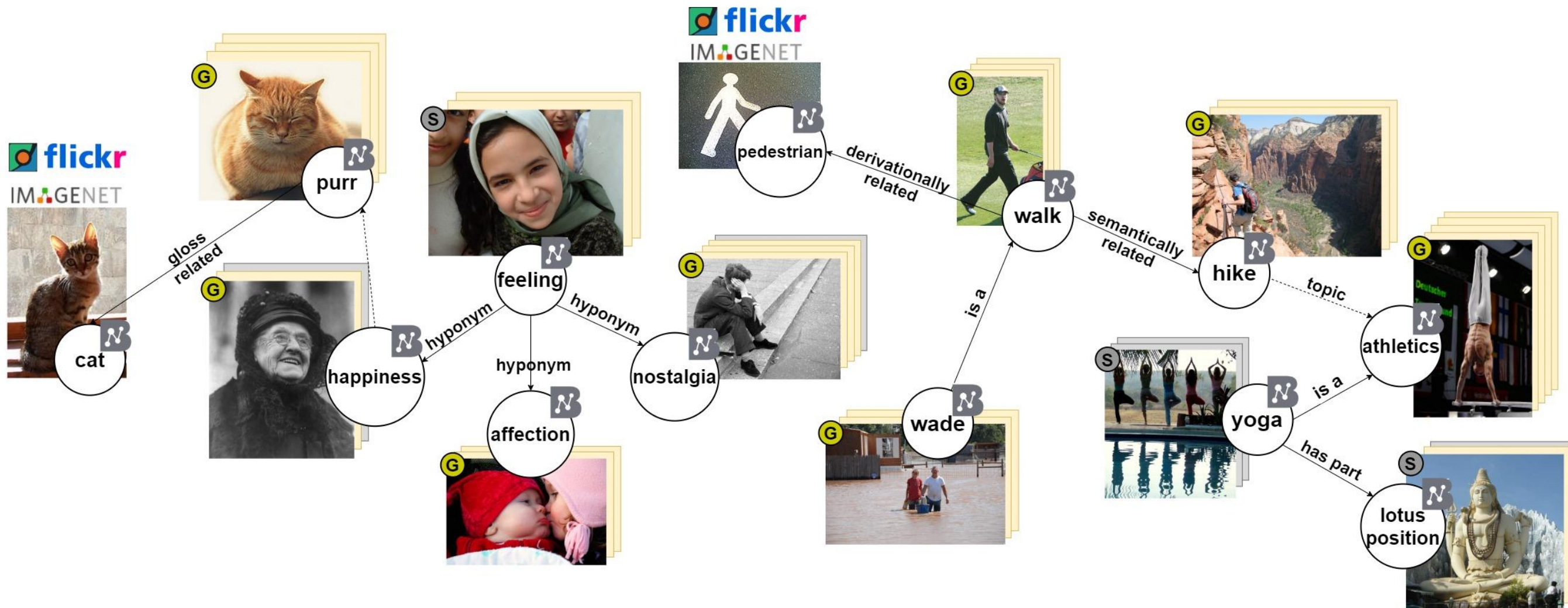




# BabelPic

New resource associating **concepts** with images:

- Focus on **non-concrete** concepts
- Linked to WordNet and BabelNet



+ methodology to extend the coverage to all BabelNet synsets!

# Starting Point

EN **graduation** • commencement • commencement ceremony • commencement exercise • graduation exercise



An **academic** exercise in which **diplomas** are conferred *WordNet* [More definitions](#)

IS A exercise • Social procedure • ceremony  
HAS PART baccalaureate  
DESCRIBED BY SOURCE Otto's encyclopedia



**BabelNet:** large multilingual resource linking WordNet to Wikipedia and other resources



# Starting Point (2)

EN hurdling · hurdle race · **hurdles**



BabelNet

A **footrace** in which contestants must negotiate a series of **hurdles** WordNet [+ Più definizioni](#)

IS A  
PART OF  
HAS KIND  
SPORT

track event · event in sports · track and field

decathlon

60 metres hurdles · 110 metres hurdles · 100 metres hurdles [+](#)

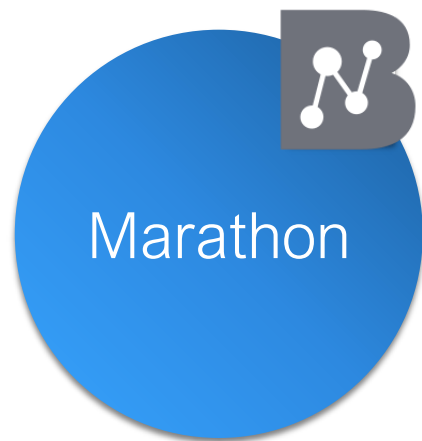
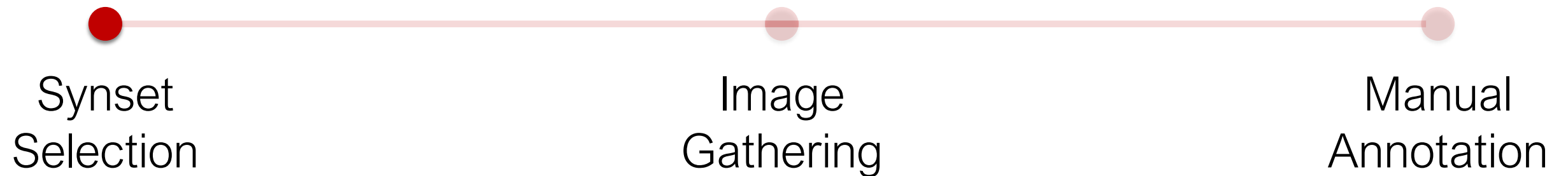
athletics



**BabelNet:** large multilingual resource linking WordNet to Wikipedia and other resources



# Gold Dataset: Building Process

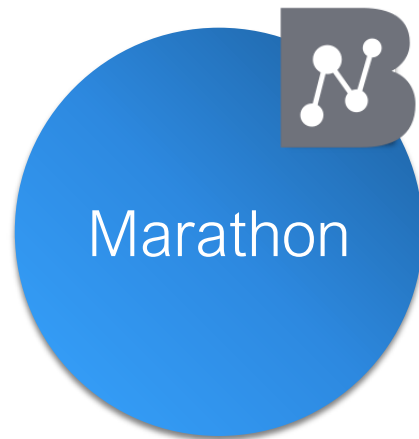


# Gold Dataset: Building Process

Synset  
Selection

Image  
Gathering

Manual  
Annotation

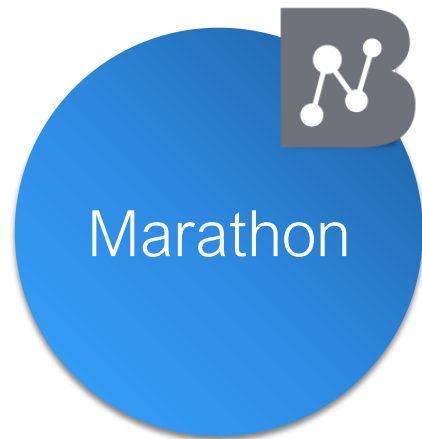


# Gold Dataset: Building Process

Synset  
Selection

Image  
Gathering

Manual  
Annotation



is the image  
pertinent to

A footrace of 26  
miles 385 yards?



# Gold Dataset: Building Process (2)

**Synset Selection:** focus on events and properties

- Select all nominal synsets inheriting from *feeling.n.01* and *event.n.01* in WordNet
- Select all synsets belonging to the *verb.competition*, *verb.motion* and *verb.social* WordNet lexicographer files

# Gold Dataset: Building Process (2)

**Synset Selection:** focus on events and properties

- Select all nominal synsets inheriting from *feeling.n.01* and *event.n.01* in WordNet
- Select all synsets belonging to the *verb.competition*, *verb.motion* and *verb.social* WordNet lexicographer files

**Image Gathering:**

- Associate each synset with the first 15 images in BabelNet 4.0

# Gold Dataset: Building Process (2)

**Synset Selection:** focus on events and properties

- Select all nominal synsets inheriting from *feeling.n.01* and *event.n.01* in WordNet
- Select all synsets belonging to the *verb.competition*, *verb.motion* and *verb.social* WordNet lexicographer files

**Image Gathering:**

- Associate each synset with the first 15 images in BabelNet 4.0

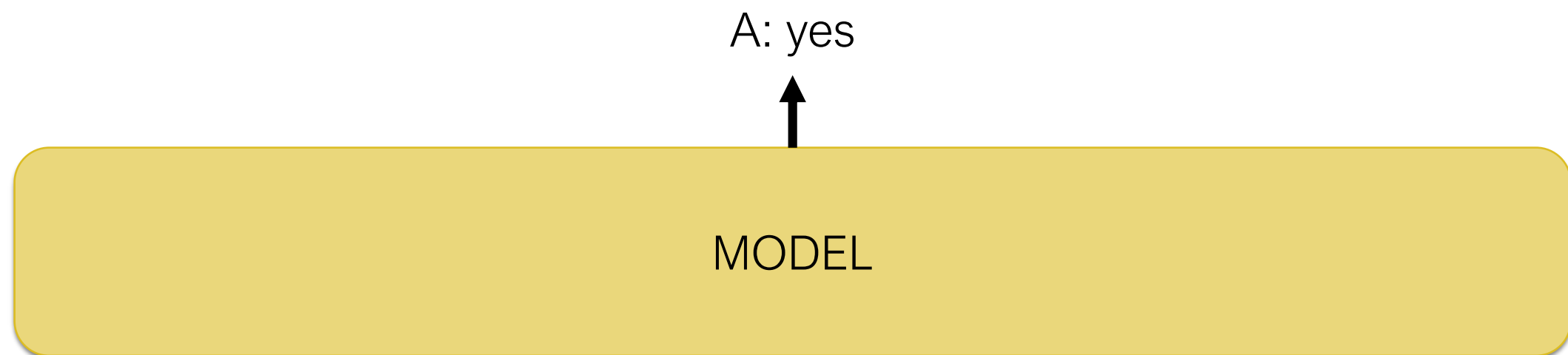
**Manual Annotation:**

- Two expert annotators: each one is assigned ~ half of the concept-image association candidates
- Agreement on sample of 100 instances: 80.39% ( $\kappa$  value of 0.6078)



# Model

Manual validation is time consuming, but wide coverage is needed...



tea party: A party at which tea is served

# Model

Manual validation is time consuming, but wide coverage is needed...

A: yes



VLP (Zhou et al., 2020)



tea party: A party at which tea is served

# Model

Manual validation is time consuming, but wide coverage is needed...

A: yes



VLP (Zhou et al., 2020)

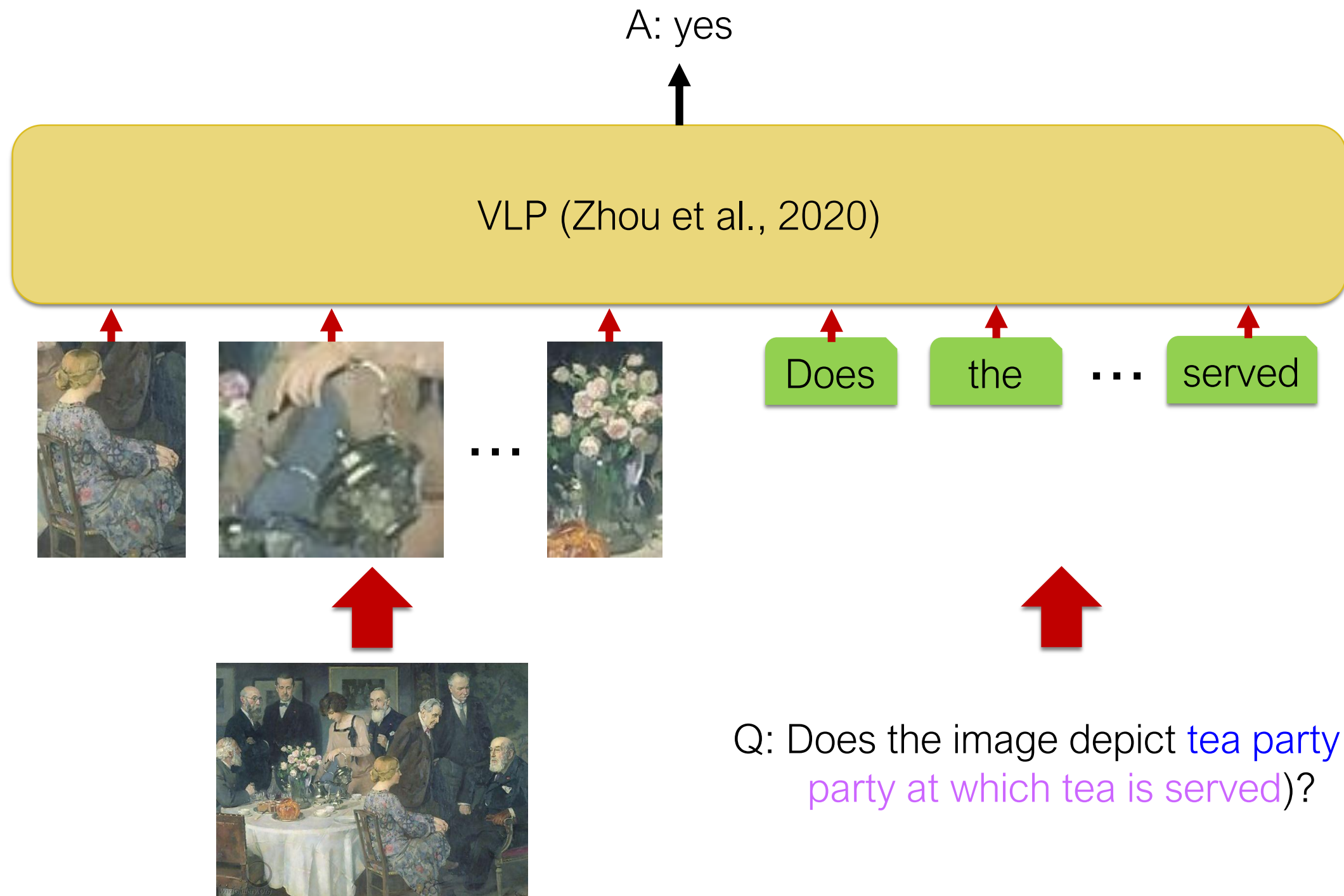


Q: Does the image depict **tea party** (**A party at which tea is served**)?



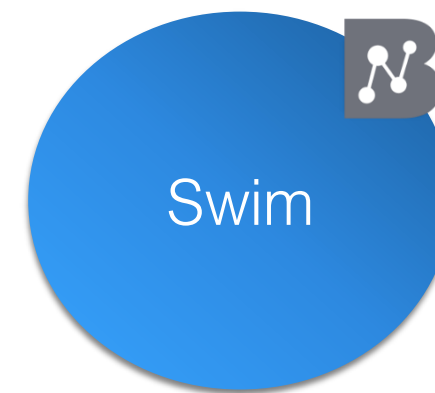
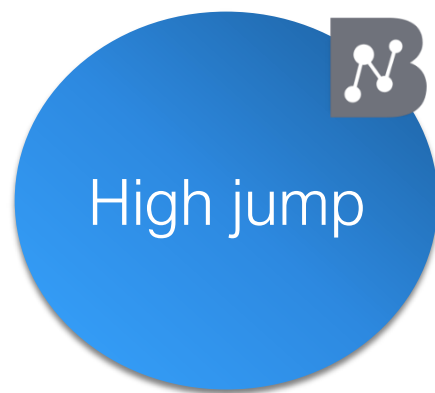
# Model

Manual validation is time consuming, but wide coverage is needed...



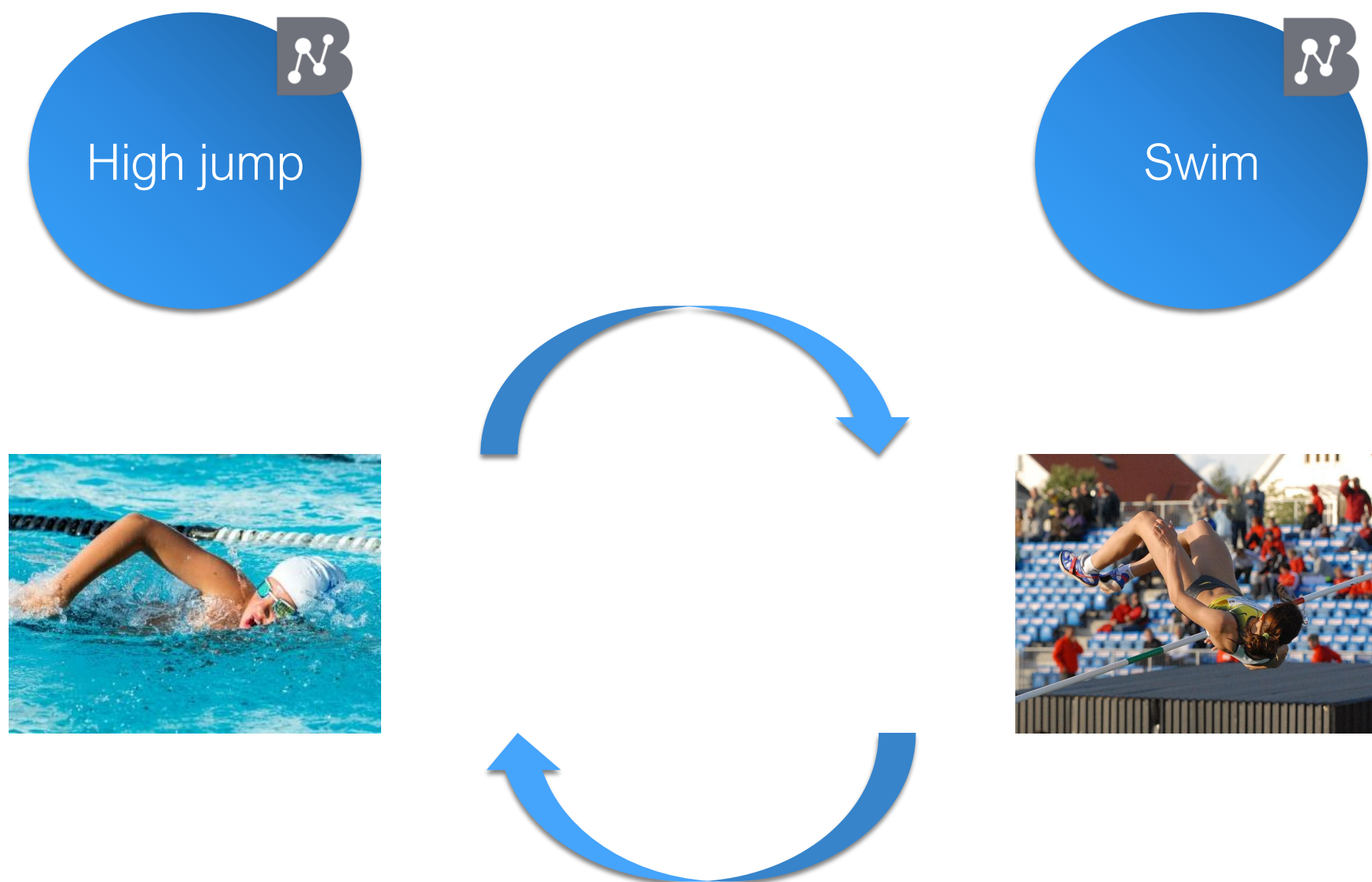
# How do we evaluate a system on BabelPic?

Generation of **negative instances**:



# How do we evaluate a system on BabelPic?

Generation of **negative instances**:

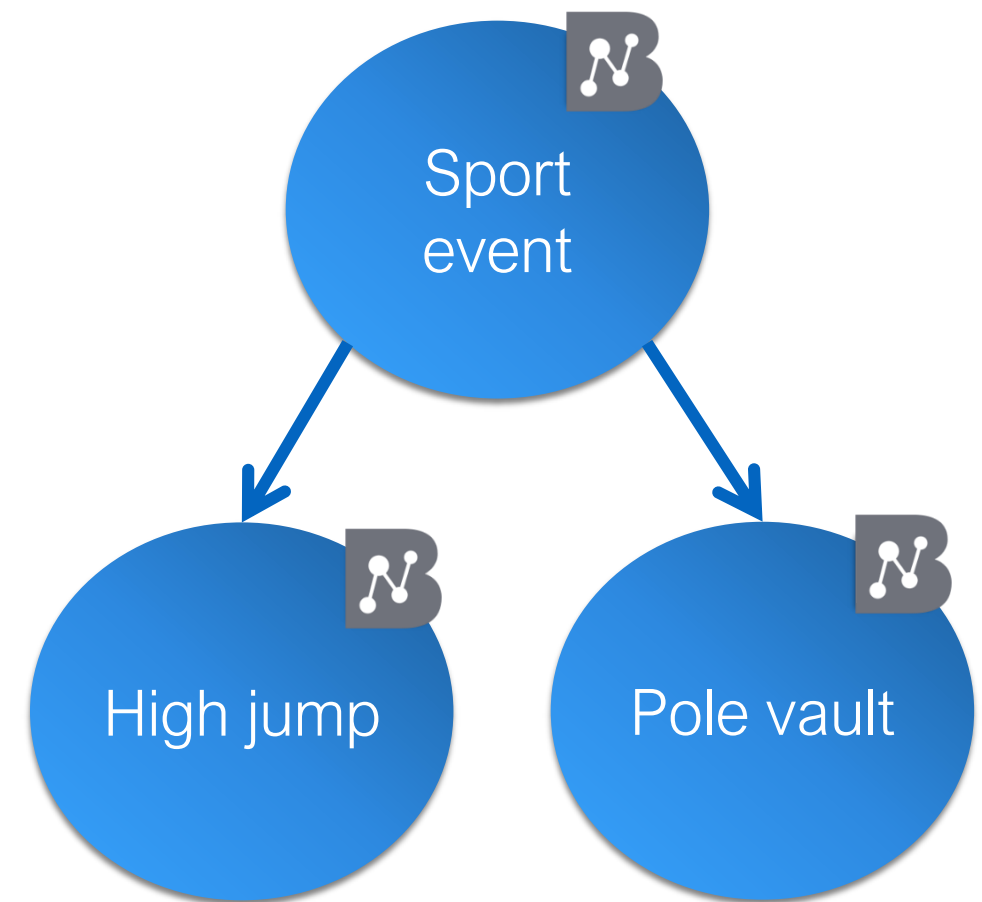




# How do we evaluate a system on BabelPic?

Generation of **negative instances**:

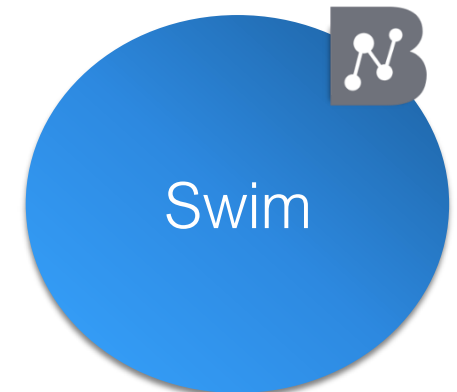
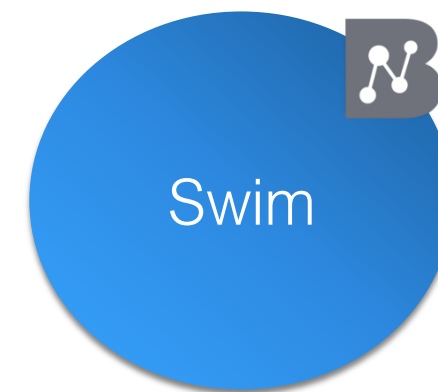
- Sibling



# How do we evaluate a system on BabelPic?

Generation of **negative instances**:

- Sibling
- Polysemy



# How do we evaluate a system on BabelPic?

Generation of **negative instances**:

- Sibling
- Polysemy
- Unrelated



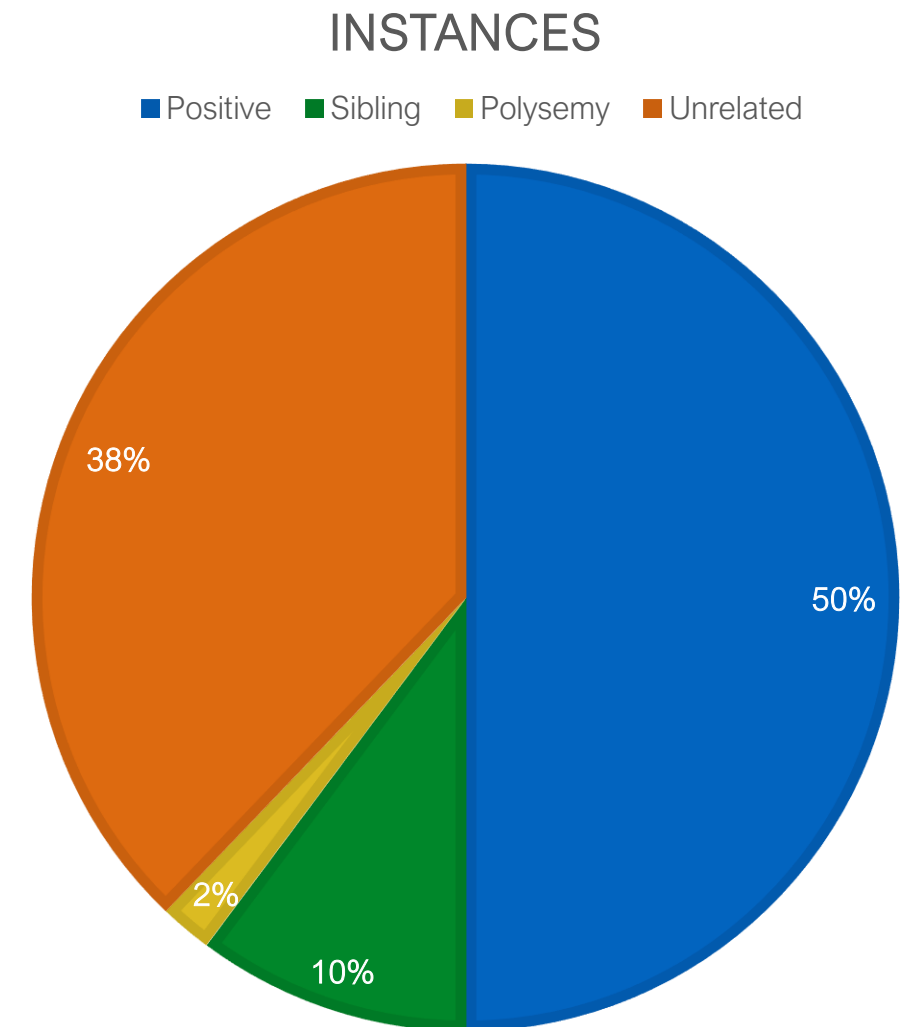
# How do we evaluate a system on BabelPic?

Generation of **negative instances**:

- Sibling
- Polysemy
- Unrelated

Split into training, validation and test set:

- 80% / 10% / 10% rule





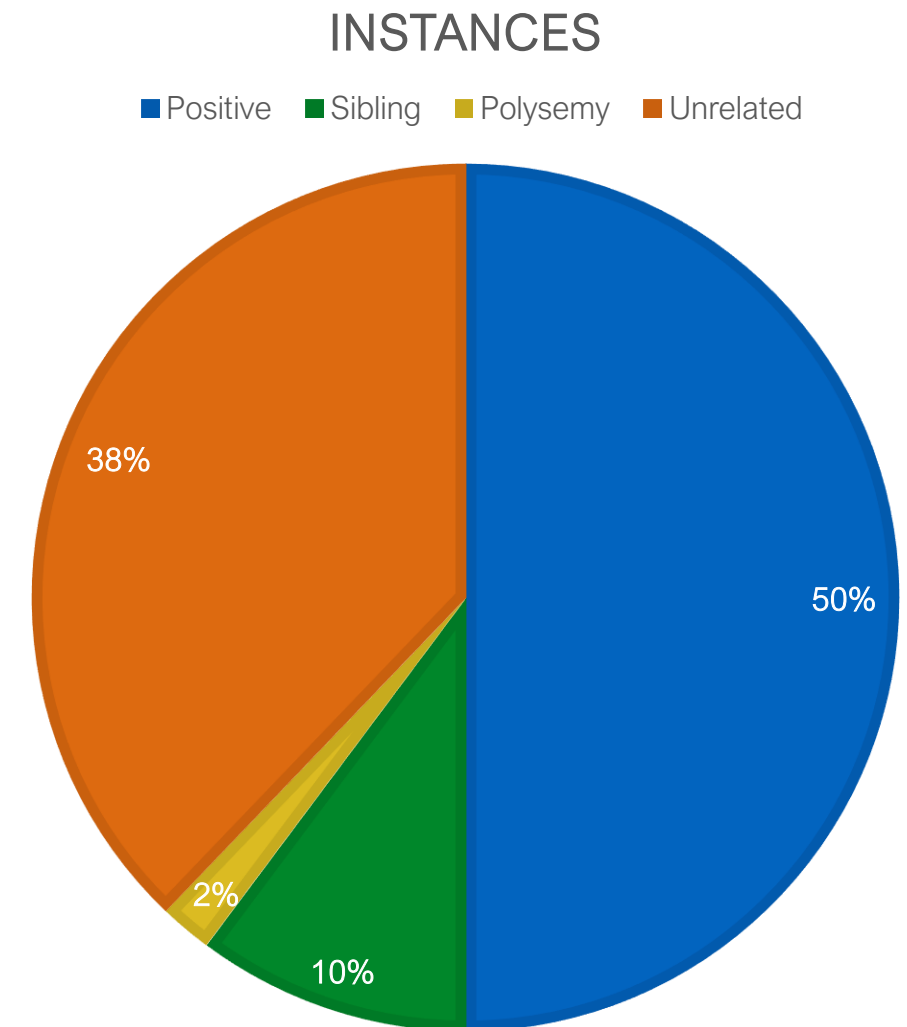
# How do we evaluate a system on BabelPic?

Generation of **negative instances**:

- Sibling
- Polysemy
- Unrelated

Split into training, validation and test set:

- 80% / 10% / 10% rule
- **Zero-shot**: ~ 500 instances in validation and test sets refer to synsets that are not present in the training set



# Experiments

- Automatic verification of synset-image associations:

Model	Test		Zero-Shot	
	P	F1	P	F1
F-VLP	75.94	75.99	77.67	71.67

- By exploiting the natural language definitions in WordNet, F-VLP shows strong results on **zero-shot** classification!

# Experiments

- Automatic verification of synset-image associations:

Model	Test		Zero-Shot	
	P	F1	P	F1
F-VLP	75.94	75.99	77.67	71.67

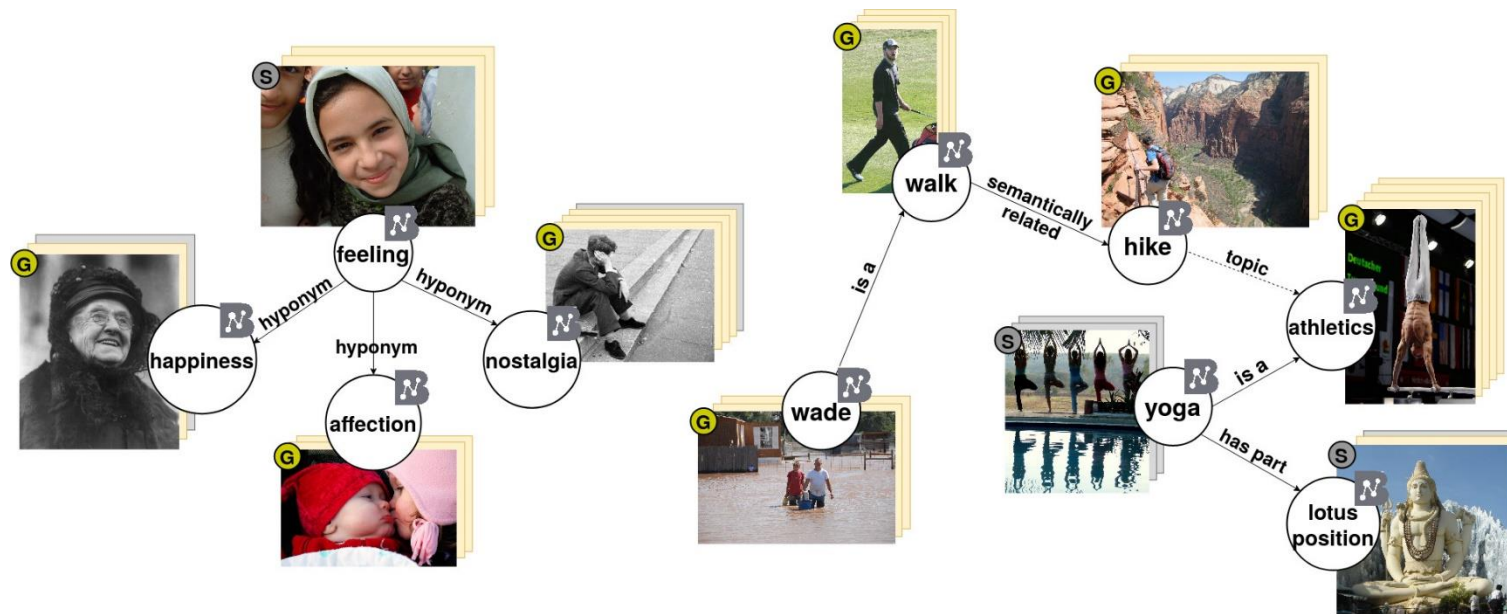
- By exploiting the natural language definitions in WordNet, F-VLP shows strong results on **zero-shot** classification!

Relation	Test	Zero-Shot
Unrelated	83.63	89.01
Sibling	53.11	62.07
Polysemy	44.83	45.45

# Conclusions

## BabelPic:

- First multimodal dataset with a focus on verbal and nominal non-concrete concepts
- Linked to the WordNet and BabelNet knowledge bases
- Gold data include **2,733 synsets and 14,931 images**
- Equipped with a methodology for the automatic extension of its coverage!  
→ We release a silver dataset containing **10,013 synsets**



Download BabelPic (Gold and Silver!) at <http://babelpic.org>







SAPIENZA  
UNIVERSITÀ DI ROMA

# Thank you!

## *Fatality Killed the Cat* or: BabelPic, a Multimodal Dataset for Non-Concrete Concepts

Agostina Calabrese, Michele Bevilacqua, Roberto Navigli  
{calabrese.a, bevilacqua, navigli}@di.uniroma1.it

Sapienza NLP Group @ Computer Science Department  
Sapienza University of Rome, Italy  
[nlp.uniroma1.it](http://nlp.uniroma1.it) | [@SapienzaNLP](https://twitter.com/SapienzaNLP)



Supported by the ERC Consolidator Grant MOUSSE No. 726487 under the European Union's Horizon 2020 research and innovation programme.

