



# A **Poodle** or a **Dog** ?

## Evaluating Automatic Image Annotation Using Human Descriptions at Different Levels of Granularity

**Josiah Wang**    **Fei Yan**    **Ahmet Aker**    **Robert Gaizauskas**

Department of Computer Science



Centre for Vision, Speech and Signal Processing



# Main Idea



Label this picture with an object name

# Main Idea



Label this picture with an object name

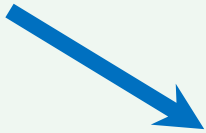
Dog Poodle Mammal  
Animal Pet Canine

# Main Idea

## System 1

1. Dog
2. Sheep
3. Sandal
4. Chihuahua
5. Footwear
6. Grass

...



1. {Dog, Chihuahua, Poodle}
2. {Sheep}
3. {Footwear, Sandal, Flip-flop}
4. {Grass, Pasture}
5. {Horse, Pony}

...



Annotation: Dog

## System 2

1. Poodle
2. Footwear
3. Grass
4. Horse
5. Dog
6. Cat

...



1. {Dog, Chihuahua, Poodle}
2. {Footwear, Sandal, Flip-flop}
3. {Grass, Pasture}
4. {Horse, Pony}
5. {Cat, Kitty}

...

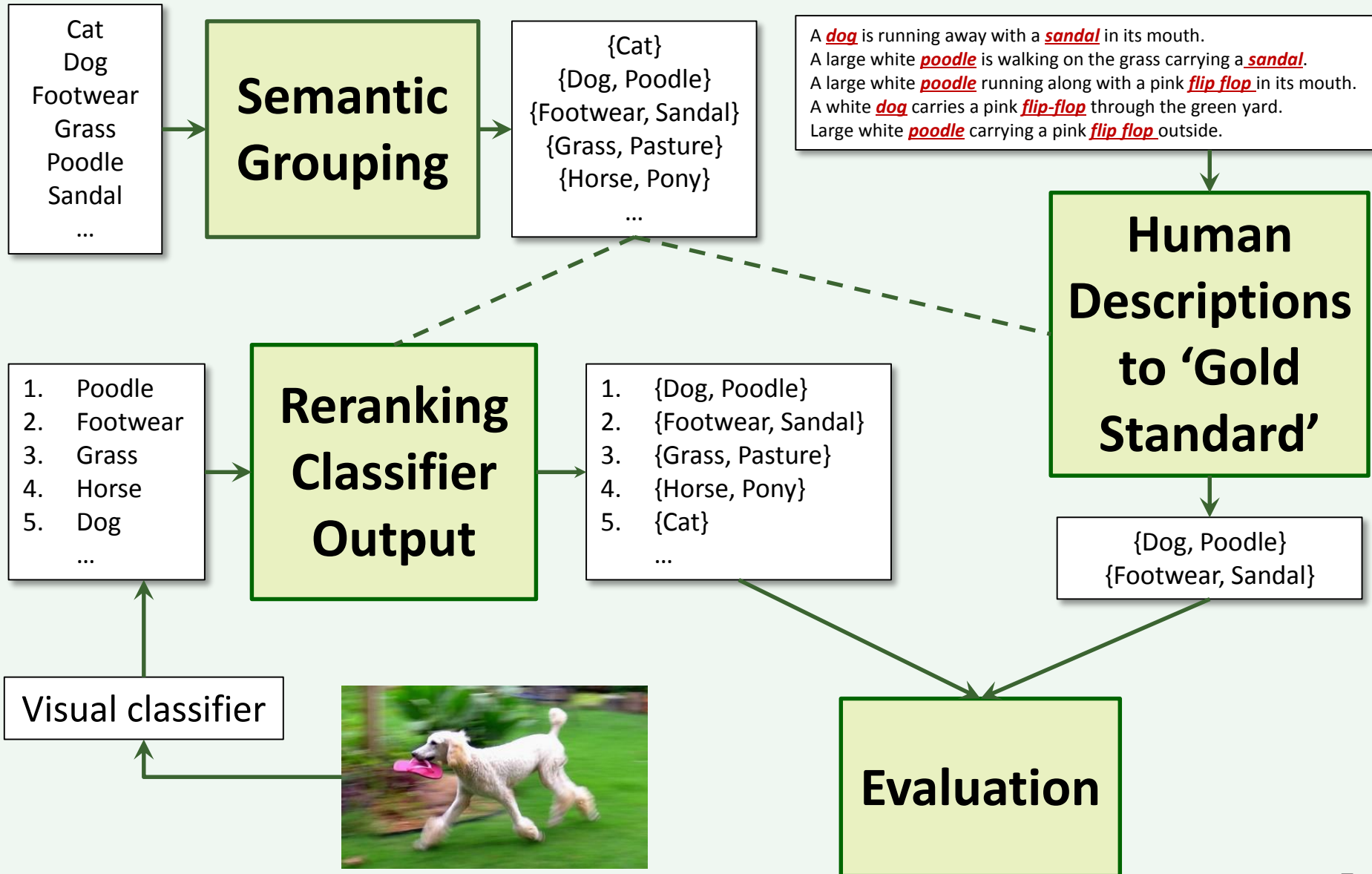
# 'Granularity aware' groupings

- Group semantically related concepts
- *Across* varying levels of granularity
- Used at **evaluation** time
  - Initial rankings are altered in a manner that gives different insights

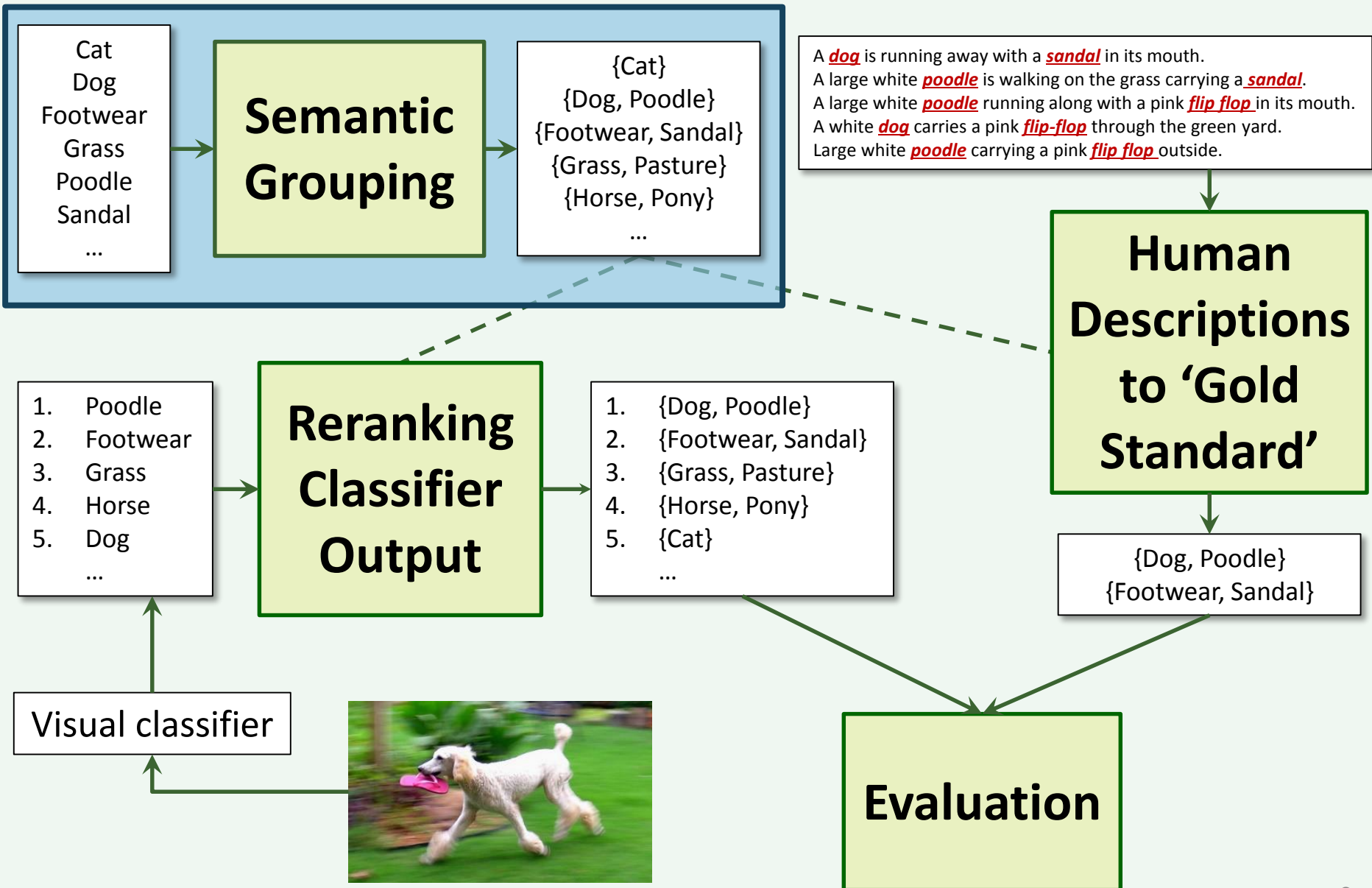
# Related Work

- Basic-level categories (Biederman 1995)
- Differences in abstraction level when labelling groups of images vs. describing individual images (Rorissa 2008)
- Classifiers decide the best level of abstraction (Deng et al 2012)
- Learn most ‘natural’ basic-level category from text corpora (Ordonez et al 2013)
- We focus on *evaluating* systems across *multiple* levels of granularity

# Method Outline



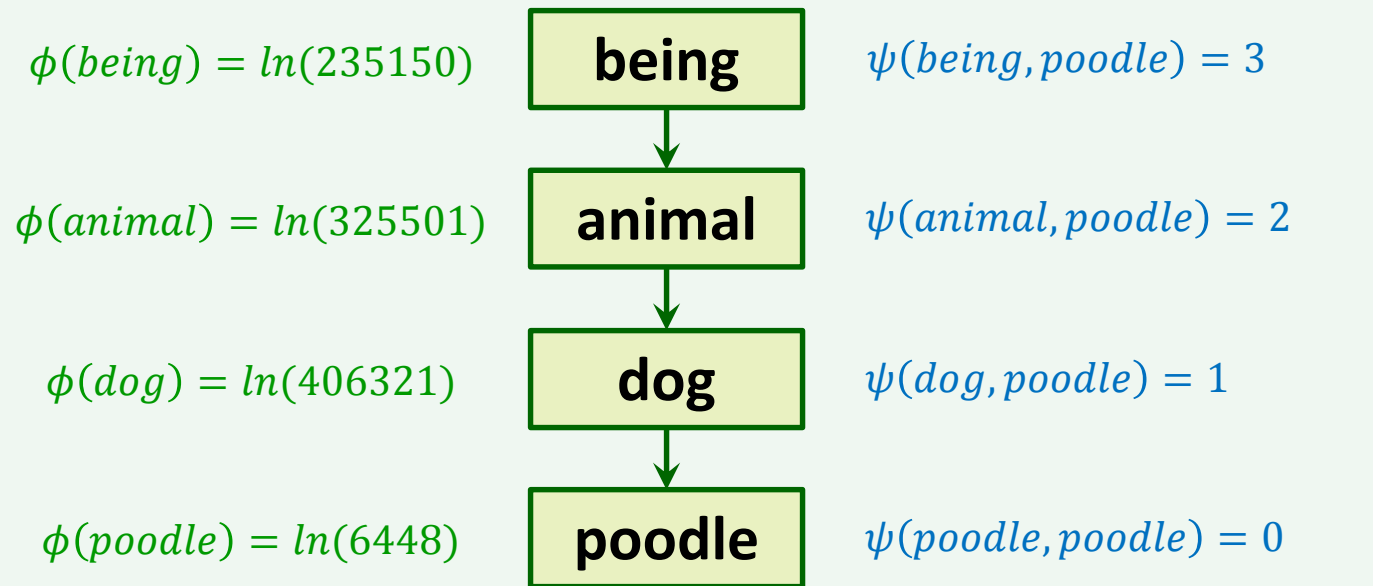
# Semantic Grouping





# Semantic Grouping

Translation function from concept  $v$  to basic-level concepts (Ordonez et al. 2013)



$$\tau(v, \lambda) = \arg \max_{w \in \Pi(v)} [ \lambda \phi(w) - (1 - \lambda) \psi(w, v) ]$$

**hypernyms of  $v$**  ↗    **'naturalness'** (YFCC100M dataset)    **semantic distance** (WordNet)

# Semantic Grouping

Group all concepts  $v$  translating to the same hypernym  $w$

dog

$\tau(\text{chihuahua}, \lambda) = \text{dog}$

$\tau(\text{dog}, \lambda) = \text{dog}$

$\tau(\text{poodle}, \lambda) = \text{dog}$

$\tau(\text{terrier}, \lambda) = \text{dog}$

footwear

$\tau(\text{footwear}, \lambda) = \text{footwear}$

$\tau(\text{sandal}, \lambda) = \text{footwear}$

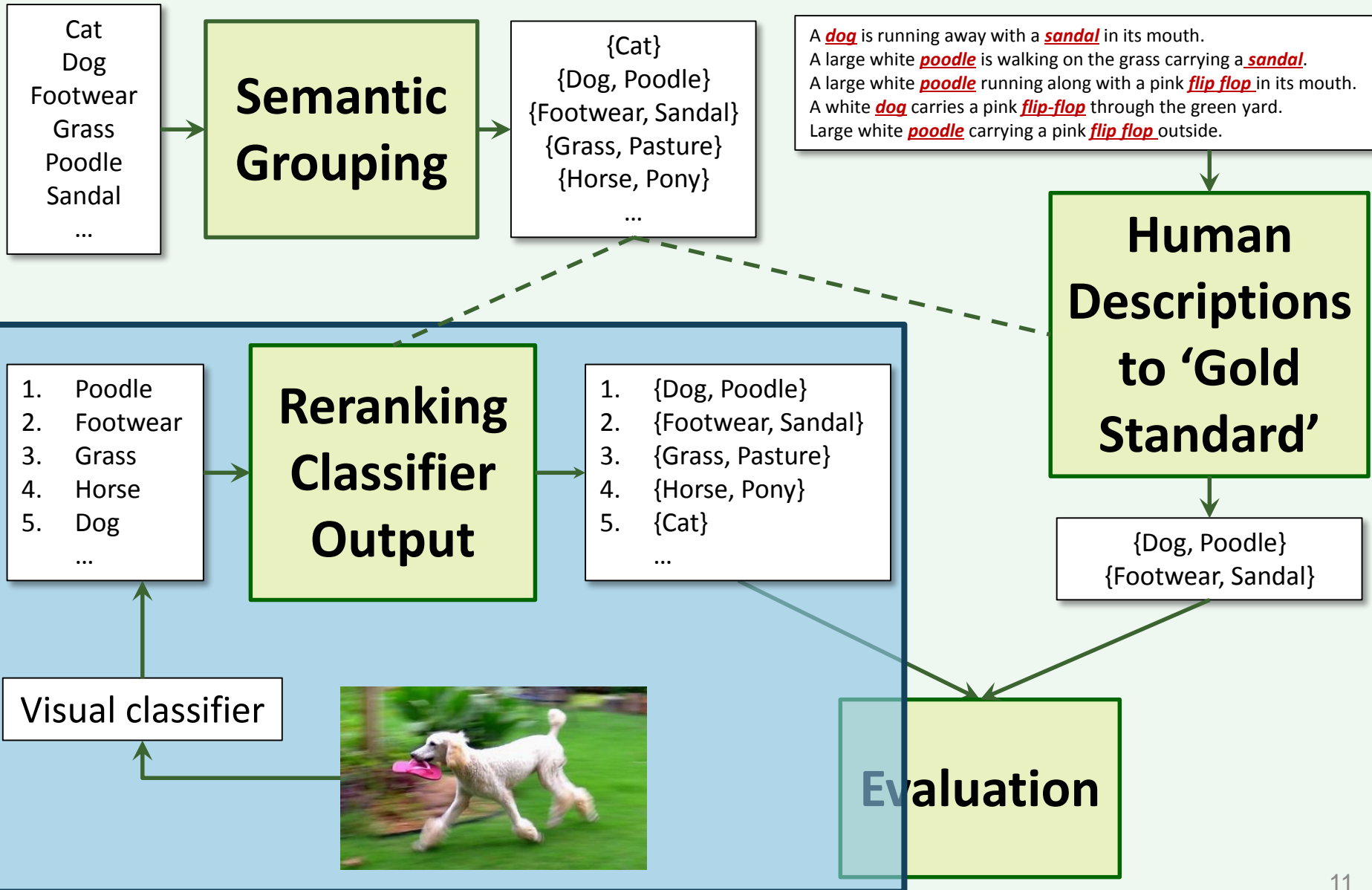
animal

$\tau(\text{animal}, \lambda) = \text{animal}$

$\tau(\text{creature}, \lambda) = \text{animal}$

$\tau(\text{mussel}, \lambda) = \text{animal}$

# Reranking Classifier Output



# Reranking Classifier Output

$$\lambda = 0.5$$



Visual  
classifiers

1. poodle (0.92)
2. footwear (0.81)
3. grass (0.80)
4. horse (0.75)
5. dog (0.74)
- ...

**dog:** {chihuahua, dog, poodle}  
**footwear:** {footwear, sandal}  
**frisbee:** {frisbee}  
**horse:** {cob, horse, pony}  
**pasture:** {pasture, grass}  
**animal:** {animal, creature, mussel}  
...

1. **dog:** {chihuahua, dog, poodle} (0.92)
2. **footwear:** {footwear, sandal} (0.81)
3. **pasture:** {pasture, grass} (0.80)
4. **horse:** {cob, horse, pony} (0.75)
5. **frisbee:** {frisbee} (0.69)
- ...

# Human Descriptions to 'Gold Standard'

Cat  
Dog  
Footwear  
Grass  
Poodle  
Sandal  
...

**Semantic Grouping**

{Cat}  
{Dog, Poodle}  
{Footwear, Sandal}  
{Grass, Pasture}  
{Horse, Pony}  
...

A dog is running away with a sandal in its mouth.  
A large white poodle is walking on the grass carrying a sandal.  
A large white poodle running along with a pink flip flop in its mouth.  
A white dog carries a pink flip-flop through the green yard.  
Large white poodle carrying a pink flip flop outside.

**Human Descriptions to 'Gold Standard'**

1. Poodle  
2. Footwear  
3. Grass  
4. Horse  
5. Dog  
...

**Reranking Classifier Output**

1. {Dog, Poodle}  
2. {Footwear, Sandal}  
3. {Grass, Pasture}  
4. {Horse, Pony}  
5. {Cat}  
...

{Dog, Poodle}  
{Footwear, Sandal}

Visual classifier



**Evaluation**

# Human Descriptions to ‘Gold Standard’

- Flickr8k Dataset (Hodosh et al. 2013)
  - 8,091 images, 5 descriptions each



A dog is running away with a sandal in its mouth.

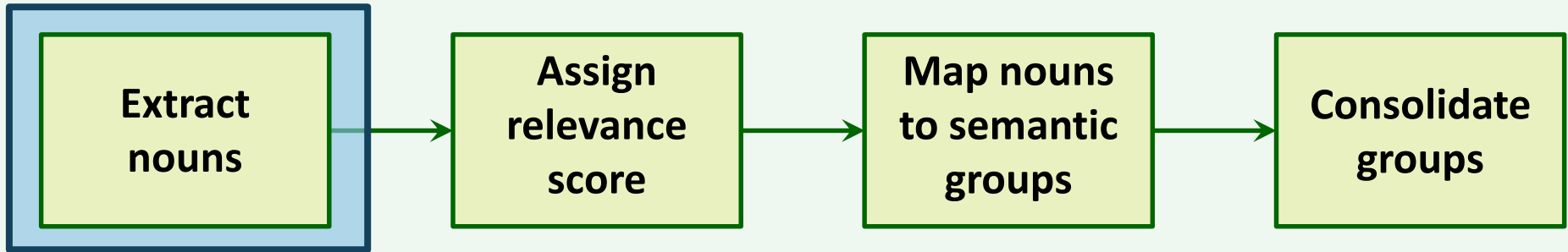
A large white poodle is walking on the grass carrying a sandal.

A large white poodle running along with a pink flip flop in its mouth.

A white dog carries a pink flip-flop through the green yard.

Large white poodle carrying a pink flip flop outside.

# Human Descriptions to 'Gold Standard'



A dog is running away with a sandal in its mouth.

A large white poodle is walking on the grass carrying a sandal.

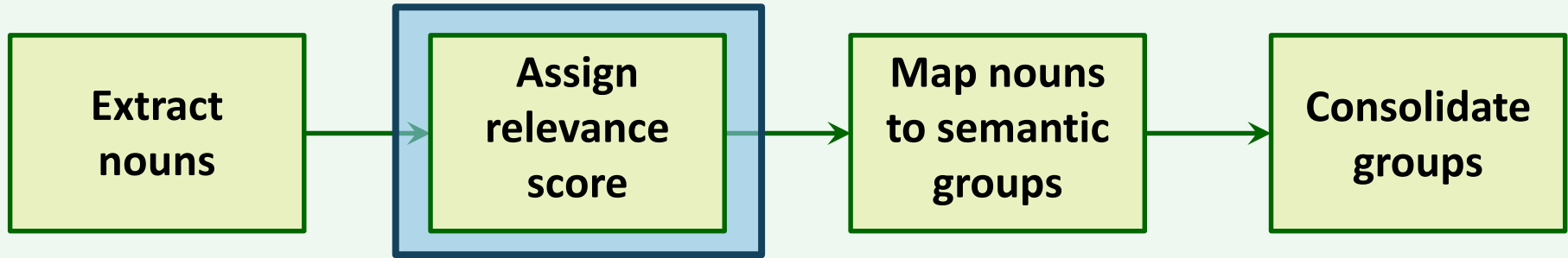
A large white poodle running along with a pink flip flop in its mouth.

A white dog carries a pink flip-flop through the green yard.

Large white poodle carrying a pink flip flop outside.

**dog**   **poodle**   **sandal**   **flip-flop**   **grass**   **yard**   **mouth**

# Human Descriptions to 'Gold Standard'



A dog is running away with a sandal in its mouth.

A large white poodle is walking on the grass carrying a sandal.

A large white poodle running along with a pink flip flop in its mouth.

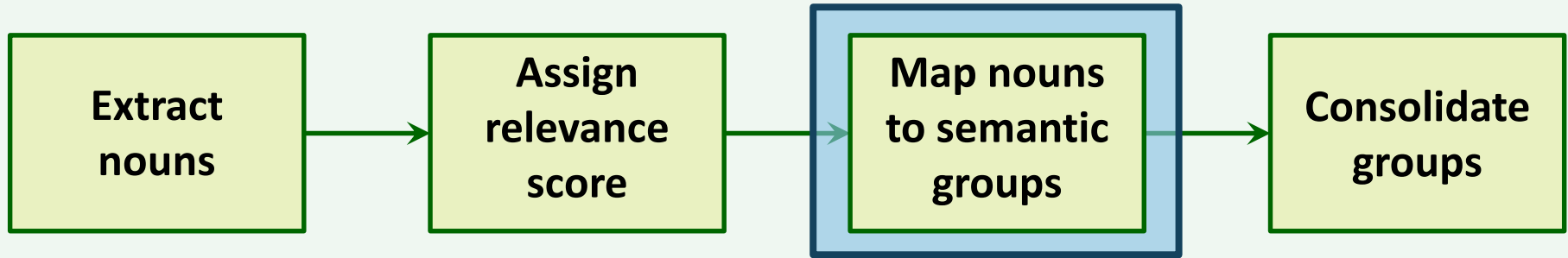
A white dog carries a pink flip-flop through the green yard.

Large white poodle carrying a pink flip flop outside.

<b>dog</b>	<b>poodle</b>	<b>sandal</b>	<b>flip-flop</b>	<b>grass</b>	<b>yard</b>	<b>mouth</b>
<b>2</b>	<b>3</b>	<b>2</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>2</b>



# Human Descriptions to 'Gold Standard'



A dog is running away with a sandal in its mouth.

A large white poodle is walking on the grass carrying a sandal.

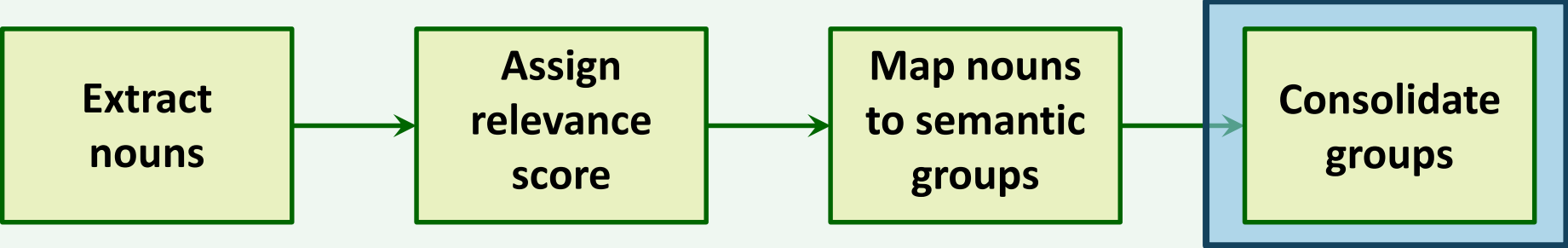
A large white poodle running along with a pink flip flop in its mouth.

A white dog carries a pink flip-flop through the green yard.

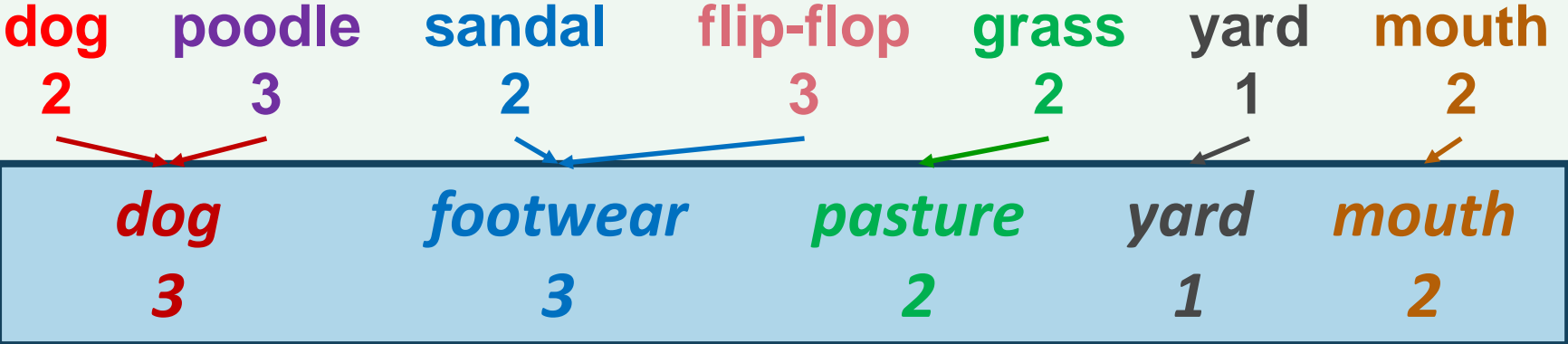
Large white poodle carrying a pink flip flop outside.

<b>dog</b>	<b>poodle</b>	<b>sandal</b>	<b>flip-flop</b>	<b>grass</b>	<b>yard</b>	<b>mouth</b>
<b>2</b>	<b>3</b>	<b>2</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>2</b>
<b>dog</b>	<b>dog</b>	<b>footwear</b>	<b>footwear</b>	<b>pasture</b>	<b>yard</b>	<b>mouth</b>

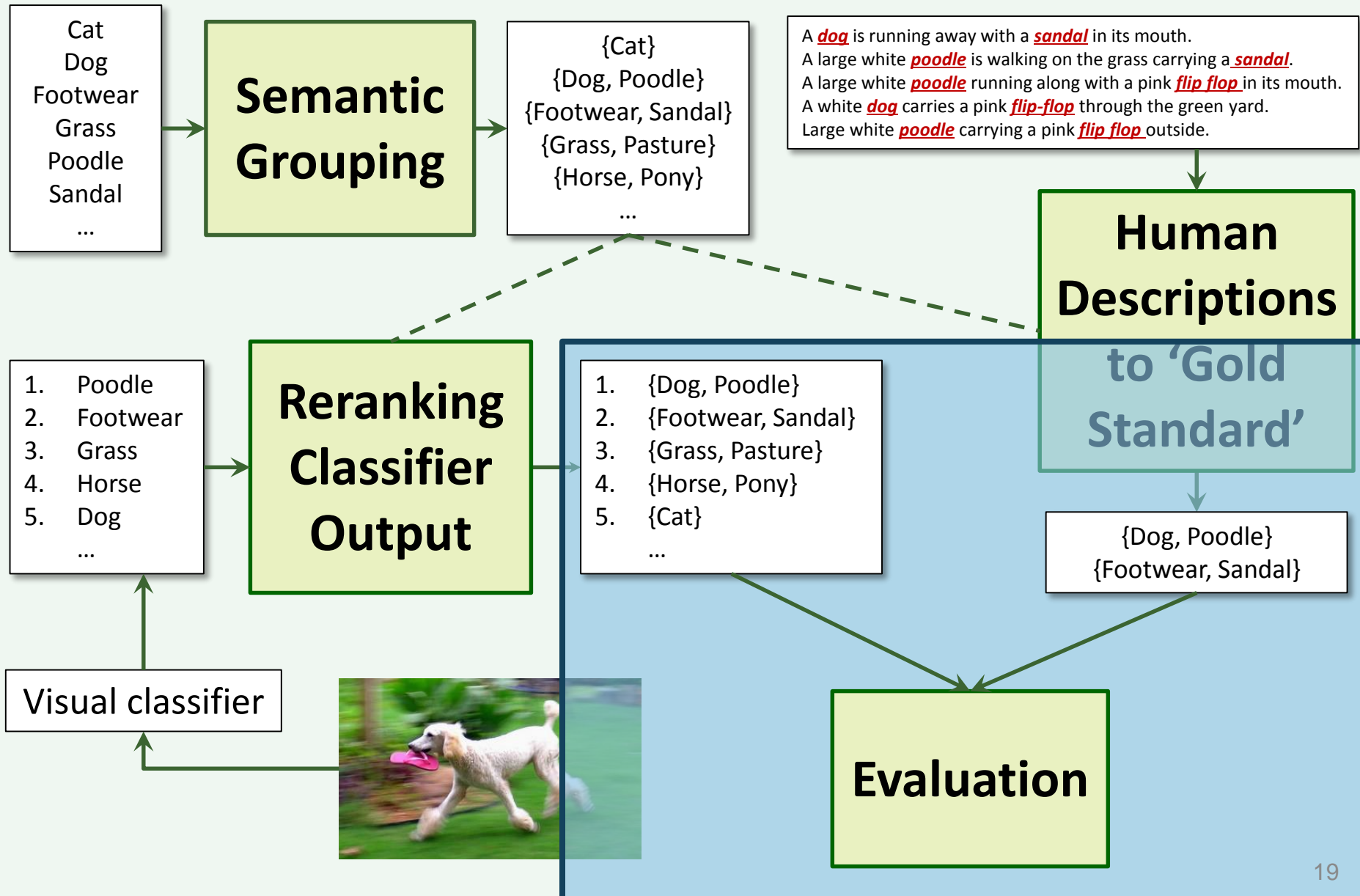
# Human Descriptions to 'Gold Standard'



A dog is running away with a sandal in its mouth.  
A large white poodle is walking on the grass carrying a sandal.  
A large white poodle running along with a pink flip flop in its mouth.  
A white dog carries a pink flip-flop through the green yard.  
Large white poodle carrying a pink flip flop outside.



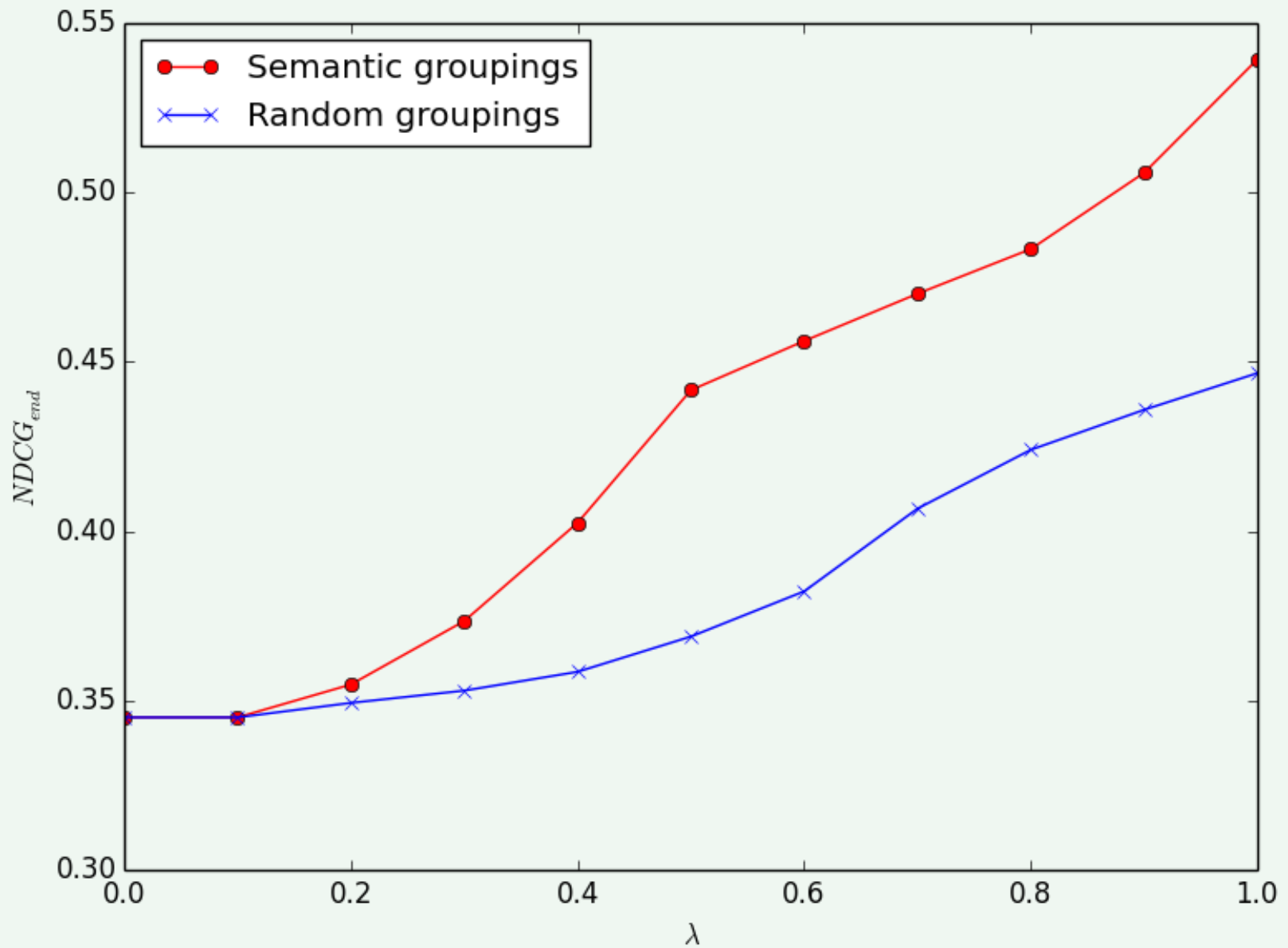
# Evaluation



# Evaluation Measure

- Normalized Discounted Cumulative Gain (NCDG)
  - Favours most relevant items first
  - Does not penalise irrelevant items
  - Normalised: between 0.0 to 1.0
    - Comparable across rankings with different no. of items
- Baseline: Concepts are grouped randomly

# Results



# Results



***dog (5)***  
***road (2)***  
***sidewalk\* (1)***  
***street (1)***

***sidewalk: {pavement, sidewalk}***

$\lambda = 0.0$

beagle  
boston\_terrier  
corgi  
basset  
hound  
spaniel  
border\_collie  
terrier  
dachshund  
pup  
st\_benard  
bulldog  
springer\_spaniel  
leash  
kitten  
pet  
**dog**  
sheepdog  
penguin  
**sidewalk**

# Results



***dog (5)***  
***road (2)***  
***sidewalk\* (1)***  
***street (1)***

***sidewalk: {pavement, sidewalk}***

$\lambda = 0.0$

beagle  
boston\_terrier  
corgi  
basset  
hound  
spaniel  
border\_collie  
terrier  
dachshund  
pup  
st\_benard  
bulldog  
springer\_spaniel  
leash  
kitten  
pet  
**dog**  
sheepdog  
penguin  
**sidewalk**

$\lambda = 0.3$

beagle  
boston\_terrier  
**dog**  
basset  
hound  
spaniel  
border\_collie  
terrier  
dachshund  
pup  
bulldog  
springer\_spaniel  
leash  
kitten  
pet  
penguin  
**sidewalk**  
doberman  
collie  
cat

# Results



***dog (5)***  
***road (2)***  
***sidewalk\* (1)***  
***street (1)***

***sidewalk: {pavement, sidewalk}***

$\lambda = 0.3$

beagle  
boston\_terrier  
**dog**  
basset  
hound  
spaniel  
border\_collie  
terrier  
dachshund  
pup  
bulldog  
springer\_spaniel  
leash  
kitten  
pet  
penguin  
**sidewalk**  
doberman  
collie  
cat

$\lambda = 0.5$

**dog**  
animal  
leash  
kitten  
pet  
penguin  
**sidewalk**  
cat  
artifact  
person  
student  
goat  
livestock  
rabbit  
duck  
baseball  
chair  
child  
frisbee  
spectator



# Results



***dog (5)***  
***road (2)***  
***sidewalk\* (1)***  
***street (1)***

***sidewalk: {pavement, sidewalk}***

$\lambda = 0.5$

**dog**

animal

leash

kitten

pet

penguin

**sidewalk**

cat

artifact

person

student

goat

livestock

rabbit

duck

baseball

chair

child

frisbee

spectator

$\lambda = 0.8$

**dog**

animal

leash

being

bird

**sidewalk**

cat

artifact

student

baseball

chair

child

frisbee

ball

slope

equipment

fabric

rug

seat

support

# Results



***boat (4)***  
***graham (1)***  
***raceway\* (1)***  
***vessel\* (1)***

***raceway: {race, raceway}***  
***vessel: {vessel, watercraft}***

$\lambda = 0.0$

motorboat  
speedboat  
lifeguard  
lifeboat  
racer

**vessel**

car  
barge  
sidecar  
kayak  
**boat**

tugboat  
paddle  
dinghy  
motor  
ferry  
bumper  
preserver  
**raceway**  
oar

$\lambda = 0.3$

**boat**

lifeguard  
lifeboat  
car

**vessel**

sidecar  
kayak  
tugboat  
paddle  
dinghy  
motor  
ferry  
glass  
preserver  
**raceway**  
oar  
airplane  
vehicle  
raft  
hatchback

# Results



***boat (4)***  
***graham (1)***  
***raceway\* (1)***  
***vessel\* (1)***

***raceway: {race, raceway}***  
***vessel: {vessel, watercraft}***

$\lambda = 0.3$

**boat**  
lifeguard  
lifeboat  
car  
**vessel**  
sidecar  
kayak  
tugboat  
paddle  
dinghy  
motor  
ferry  
glass  
preserver  
**raceway**  
oar  
airplane  
vehicle  
raft  
hatchback

$\lambda = 0.5$

**boat**  
lifeguard  
car  
**vessel**  
sidecar  
kayak  
tugboat  
paddle  
motor  
glass  
equipment  
**raceway**  
oar  
airplane  
vehicle  
raft  
hatchback  
device  
screen  
field

# Results



***boat (4)***  
***graham (1)***  
***raceway (1)***  
***vehicle (1)***

***vehicle: {buggy, bulldozer,  
camper, carriage, cart, ...,  
sailboat, ..., vehicle, vessel,  
wagon, watercraft,  
wheelbarrow, yacht}***

$\lambda = 0.5$

***boat***  
lifeguard  
car  
***vessel***  
sidecar  
kayak  
tugboat  
paddle  
motor  
glass  
equipment  
***raceway***  
oar  
airplane  
vehicle  
raft  
hatchback  
device  
screen  
field

$\lambda = 0.8$

***boat***  
being  
car  
***vehicle***  
artifact  
tugboat  
paddle  
motor  
glass  
equipment  
***raceway***  
airplane  
raft  
structure  
device  
screen  
field  
wrapping  
brush  
pop

# Discussion

- We proposed grouping semantically related concepts *across* different levels of granularity
- Semantic groups are used to alter visual classifier rankings, provides different insights
- Human-centric evaluation with descriptions
- Trade-off between flexibility vs. informativeness
- Future work:
  - Different methods of grouping concepts
  - Incorporate visual classifiers for grouping & reranking



# A **Poodle** or a **Dog** ?

## Evaluating Automatic Image Annotation Using Human Descriptions at Different Levels of Granularity

**Josiah Wang**    **Fei Yan**    **Ahmet Aker**    **Robert Gaizauskas**

Department of Computer Science



Centre for Vision, Speech and Signal Processing

