

What's in a description?

Larry Zitnick

Microsoft Research

“Long, blue, spiky-edged shadows crept out across the snow-fields, while a rosy glow, at first scarce discernible, gradually deepened and suffused every mountain-top, flushing the glaciers and the harsh crags above them.” **John Muir**



Captions in the wild

flickrTM

Pigeon Point Lighthouse near Pescadero, California





I snapped this shot right before going to sleep in my car at the South Sister trailhead Thursday night. Nothing too spectacular by most standards, but it solidifies the memory in my brain, no doubt.

A year ago I was staring at my feet in a hospital bed after having a stroke. It was discovered that I had a hole in my heart that caused the stroke and it was quickly fixed. For my one year anniversary, I hiked Table Mountain. It's over 12 miles with an elevation gain of nearly 4000 feet. The elevation at the top is 11,300 feet. It's my favorite hike and I try and do it every year. Last year, sitting in the hospital, I only dreamed of doing this again. I can't tell you how grateful I am for my good health and a full recovery. Serendipity is the best word to describe my feelings at the moment this image was captured. Please check the comment box for an image from last year.





i tried to stop my 365 and found i missed is a lot, more than i really imagined i would. so i am back at it:
we all met at the Pancake house
so many delicious choices so we shared taste
getting up from the table full and happy
justin and kacy headed home
and brandon and courtney came back to the house for a bit
before leaving for Bellingham
he will leave is a bit too for the east side
leaving the dog and i here alone for a few days
to rest. . .

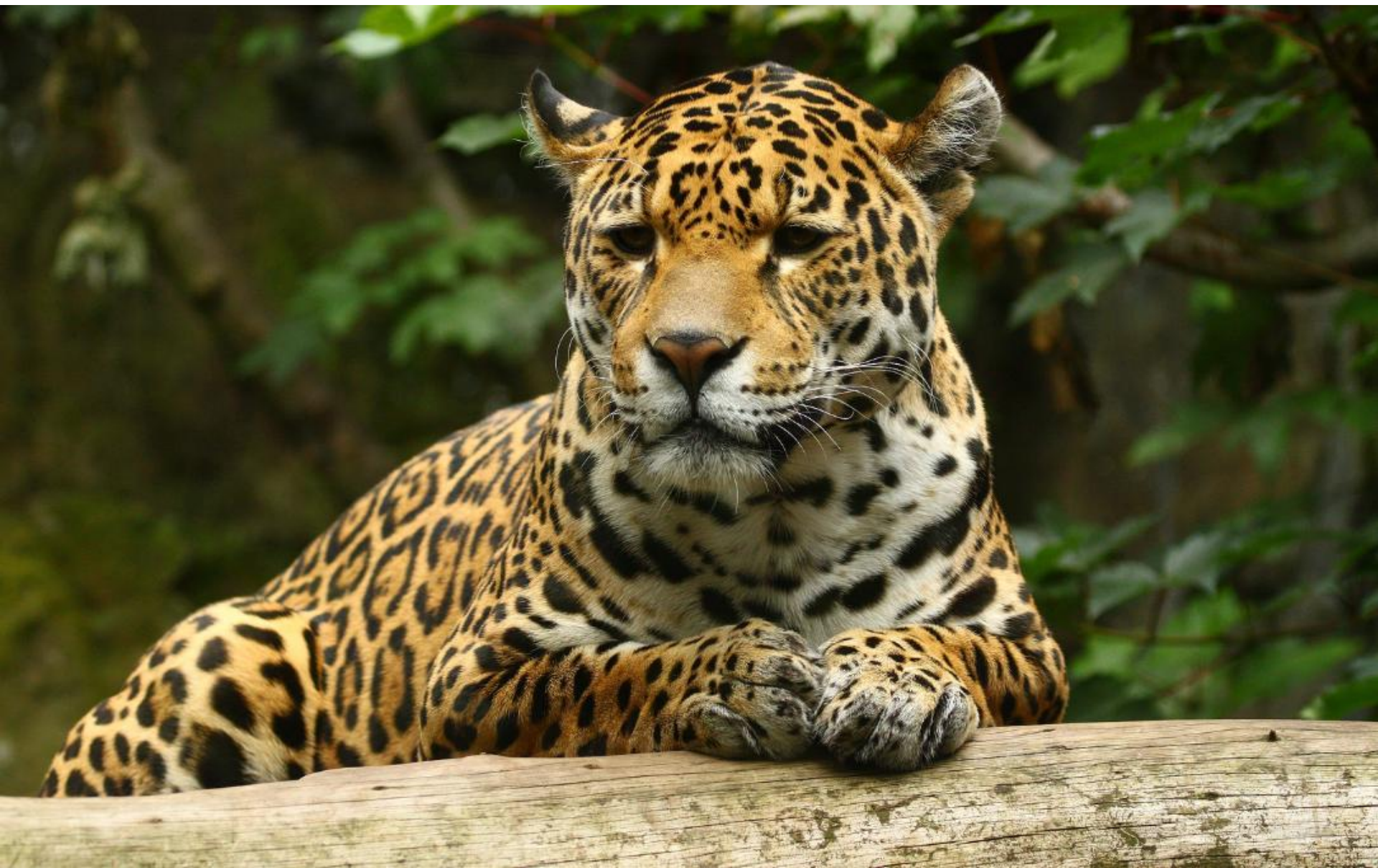
This is the private poker room, formally owned by the late George Whittle, located at the Thunderbird lodge. The lodge and property is now managed by the Thunderbird Lodge Preservation Society. George lost a large chunk of North Tahoe land to a card shark by the name of Joe King, who used the land to create the town of King's Beach, California nearly a century ago. This photo is the poker room when that famous card game was played.





Late evening light lights up the famous pink rocks of Fionnphort and Kintra across the Sound of Iona.

Shot was taken thru a glass window. This was the best shot without glass reflection.



We did smile at each other after!




Walk up shot of Tower Bridge on an August evening



IanR

I know of a few people that could use this
reminder right about now. :)

A white rectangular card is centered in the frame, resting on a light-colored surface. The card has the words "You are Loved." written in a dark, cursive script. The card is surrounded by numerous small, delicate purple flowers, some of which are in sharp focus in the foreground and others are blurred in the background, creating a soft, romantic atmosphere.

You are
Loved.

Captions in the wild

- How the photo was taken
- Proper names
- Special meaning behind the photo
- Events surrounding the photo

Describe the image using a single sentence.

Describe the image using a single sentence.



~300,000 images
5 sentences per image

Talk on
Wec



Tsung-Yi Lin
Cornell Tech

Cyrus Rashtchian, Peter Young, Micah Hodosh, and Julia Hockenmaier.

[Collecting Image Annotations Using Amazon's Mechanical Turk.](#)

NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk.



A **man** with a **laptop** and a **cat** on a **table**.

A **cat** is sitting outside while a **man** works on his **laptop**.

A **hippy** sitting at a **table** with a **cat** on top of it working on a **laptop**.

A **man** using a **laptop** at a **table** with a **cat** on top of the table.

A **man** at a **table** with his **computer** and a **cat**.



A **dog** trots and pants in front of two **motorcycles** parked in front of a fence.
A brown, black and white **dog** walking on a **street**.
A close up of a **dog** walking in front of parked **motorcycles**
An image of a **dog** walking in the **street**
A **dog** crossing a **pavement path** near **motorcycles**.



A **man** sits on a motorcycle holding a **cat**.

A **motorcycle rider** putting a **cigarette** in his mouth.

A **person** sitting on a motorcycle with a **dog** in their jacket.

Man on motorcycle holding a black **cat** in his jacket.

A caucasian **female** motorcycle rider sitting astride her motorcycle with a **cat**.



A person wearing a helmet and riding a motorcycle.

A man riding a motorcycle with a **small dog** propped on the fuel tank.

A man riding a motor cycle down a street.

A person riding a motorcycle on a city street

A man who is riding on a motorcycle with a **cat**.



A man riding a large motorcycle near a sidewalk with bicycles.

A man on a black and red motorcycle riding down the street.

A man riding on the back of a motorcycle.

A man with a helmet on is riding a motorcycle in the street.

A man that is sitting on the back of a motorcycle.

The woman is working in the kitchen with the baby in the high chair next to her.

A living room with a couch a chair and a TV.



A living room and kitchen with various furniture.

The **woman** is working in the kitchen with the **baby** in the high chair next to her.

A view of living room with brown sofas, with the kitchen in the background.

A living room with a couch a chair and a tv.

A **woman** works in a kitchen that is joined to a living room.



A living room with a laundry basket on the floor.

A television set a gray chair a table lamp and some pictures.



A room with an older model television in it.
A living room with a laundry basket on the floor
A living room that has a television and a chair
A television set a gray chair a table lamp and some pictures
A picture of a old fashioned looking living room.



Small dog with both a cat and a mouse riding on its back.

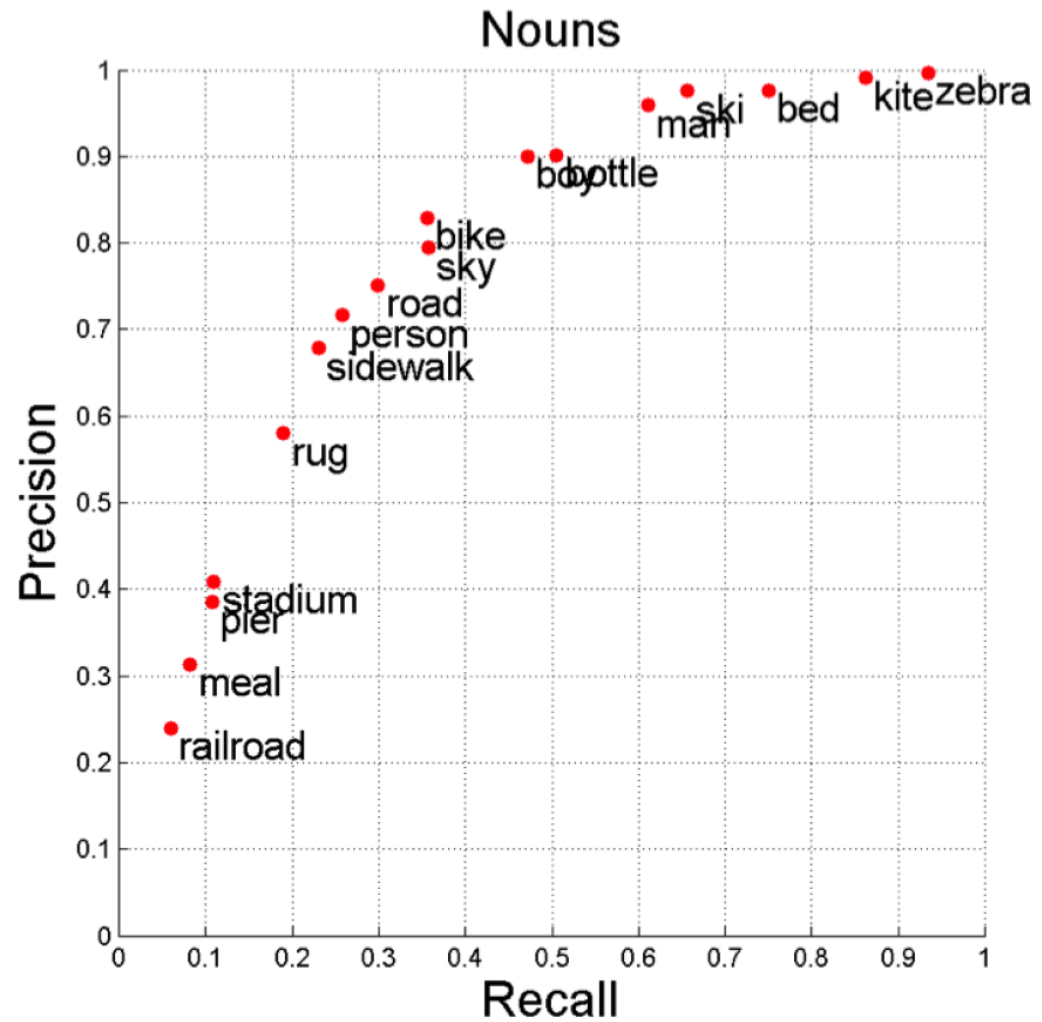
Cat and mouse riding on a dog on a busy city street.



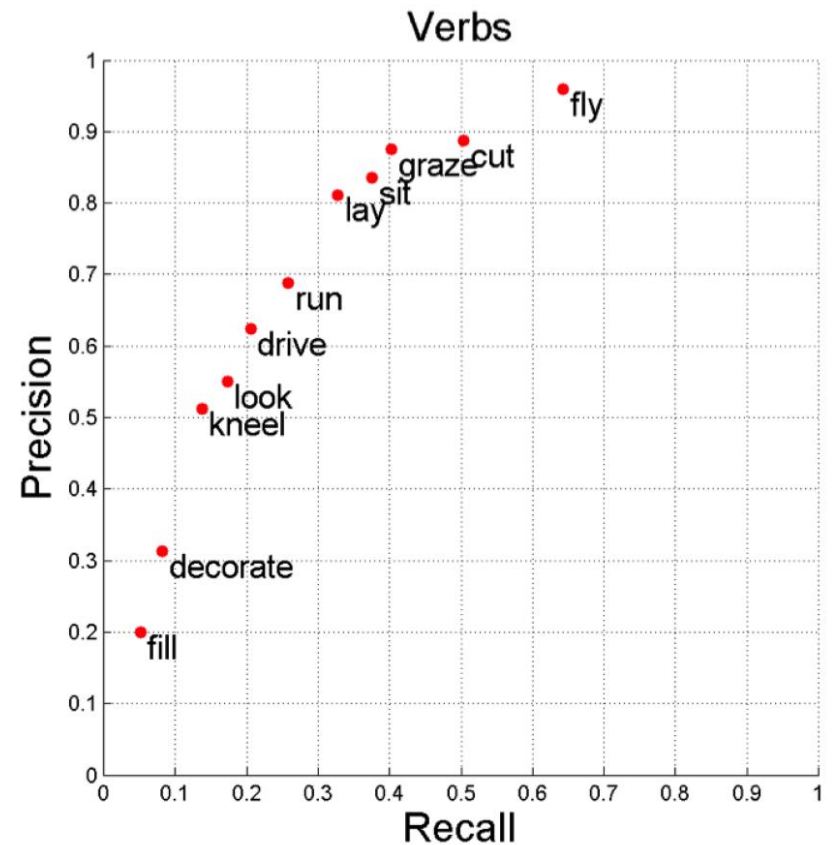
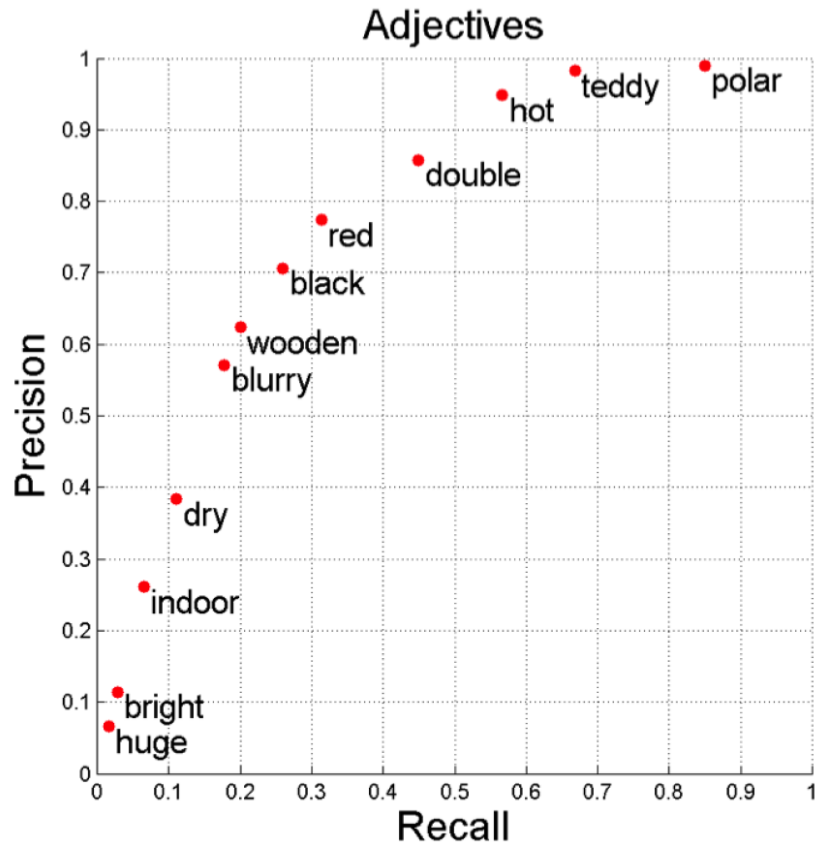
Human agreement



Saurabh Gupta,
Berkeley



Human agreement



How do we evaluate?



A photo taken indoors.

A dark haired man dressed in a black suit and blue paisley tie eating a hotdog in one hand and holding a beer and plate with a hotdog in the other.

Man suit hotdog tie eat.

✓ Like Share You, Nanxi Che and 56 others like this.

29892 Games Played

Goal: 100,000

man in red shirt on horse

Time Elapsed

19

Score

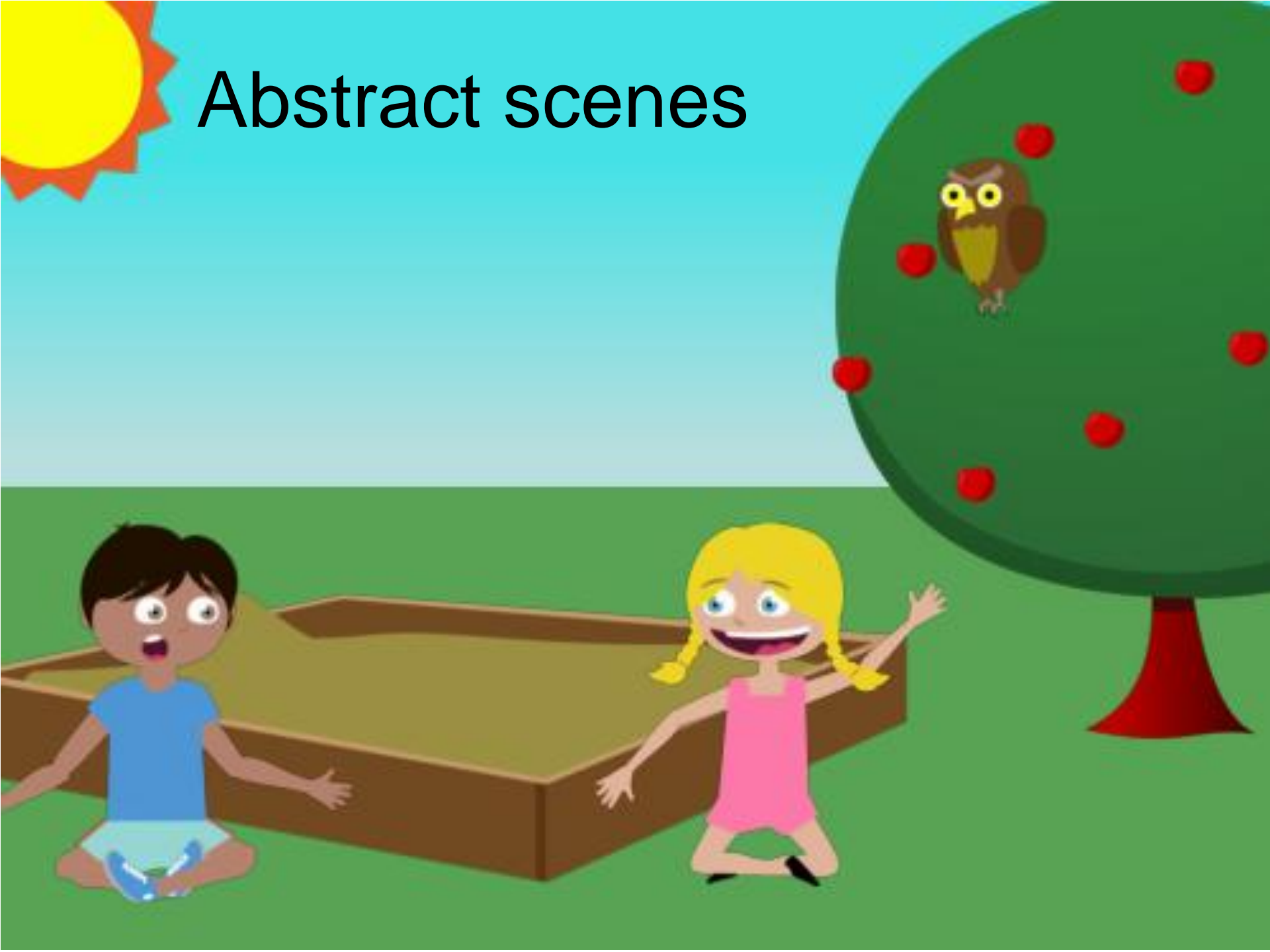
38



man in red shirt on horse

ReferItGame: Referring to Objects in Photographs of Natural Scenes
Sahar Yazemzadeh, Vicente Ordonez, Mark Matten, Tamara L. Berg,
Empirical Methods in Natural Language Processing (EMNLP) 2014.

Abstract scenes



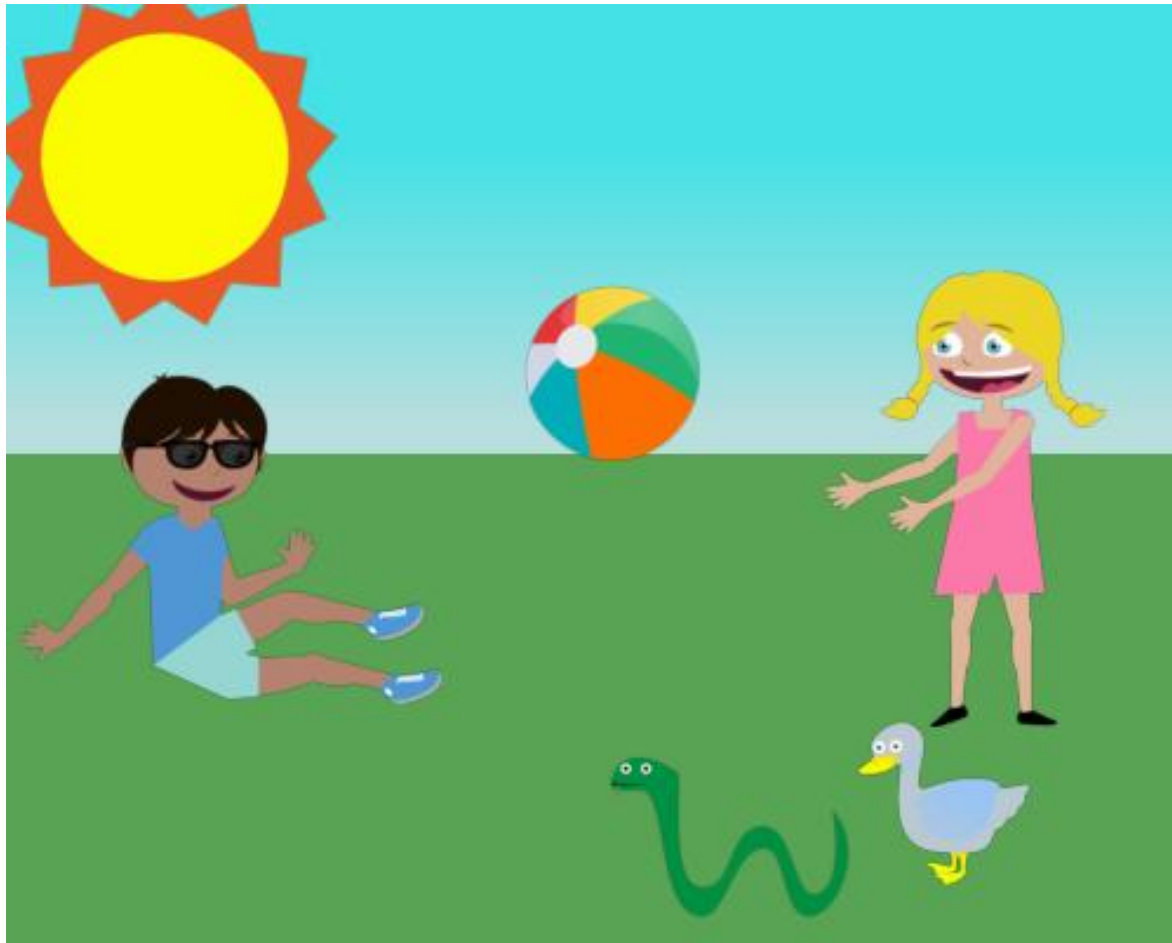
Write **three** simple sentences
that describe the scene.



Jenny is upset because Mike isn't sharing the soccer ball.
Mike is wearing sunglasses.
Jenny is wearing a silly hat.



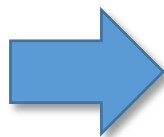
There is a hot dog on the table.
The cat looks at the hot dog.
There's food on the grill.



Mike and Jenny are playing with the beach ball.
Mike is wearing sunglasses.
The snake and the duck are watching.

Goal

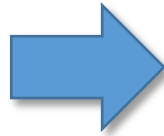
Jenny is catching the ball.
Mike is kicking the ball.
The table is next to the tree.



Semantics

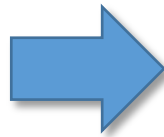
<primary object> <relation> <secondary object>

Jenny is catching the ball.
Mike is kicking the ball.
The table is next to the tree.



<Jenny> <catch> <ball>
<Mike> <kick> <ball>
<table> <next to> <tree>

Jenny and Mike are running
from the snake.



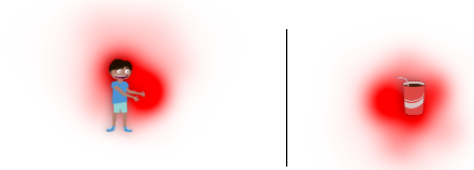
<Jenny> <run from> <snake>
<Mike> <run from> <snake>

Relations

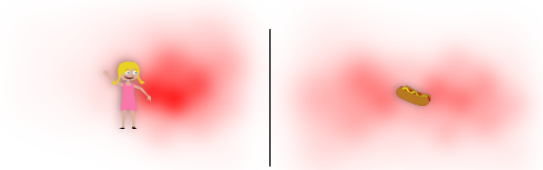
run away from



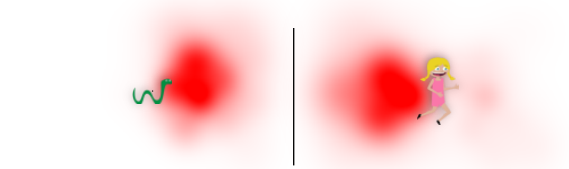
hold



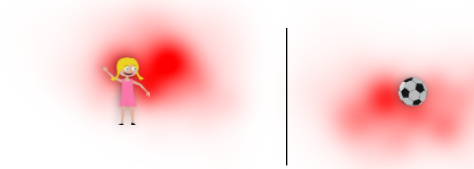
want



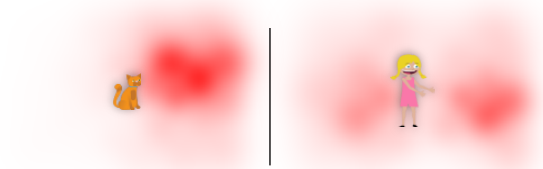
chase



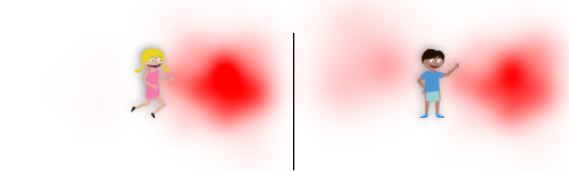
throw



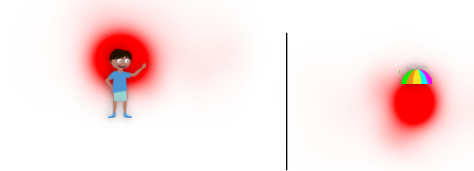
watch



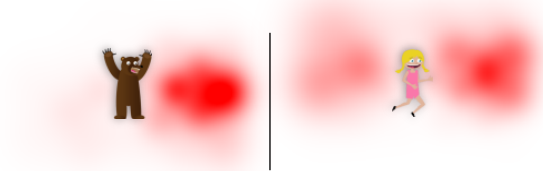
run towards



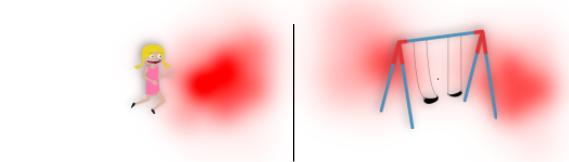
wear



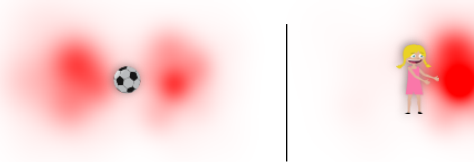
scare



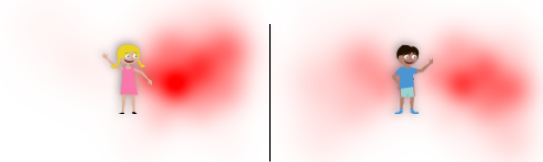
run to



to



laugh



Qualitative results

Input Description

Jenny is catching the ball.
Mike is kicking the ball.
The table is next to the tree.

Mike is sitting next to Jenny.
The cat is sitting next to the tree.
Jenny is throwing the ball.

Tuples

<<Jenny>, <catch>, <ball>>
<<Mike>, <kick>, <ball>>
<<table>, <be>, <>>

<<Mike>, <sit next to>, <Jenny>>
<<cat>, <sit next to>, <tree>>
<<Jenny>, <throw>, <ball>>

GT



Full-CRF



Simple sentence = easy to parse

More sentences = more details

Describe a scene for a children's story book.



Jenny loves to play soccer but she is worried that Mike will kick the ball too hard.



A cat anxiously sits in the park and stares at a unattended hot dog that someone left on a yellow bench.



Mike trips on his way to catch a frisbee. Jenny is frustrated with his incompetence.

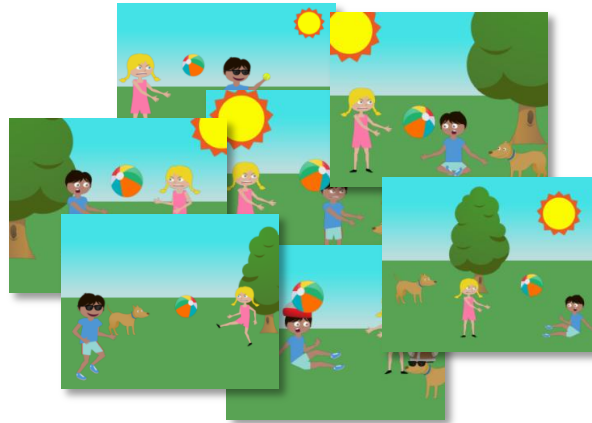


Jenny's pie got wet. Mike was happy it was ruined since she'd refused to share it with him.

Generating data



“Jenny just threw the beach ball angrily at Mike while the dog watches them both.”



Bringing Semantics Into Focus Using Visual Abstraction,
Zitnick and Parikh, CVPR 2013

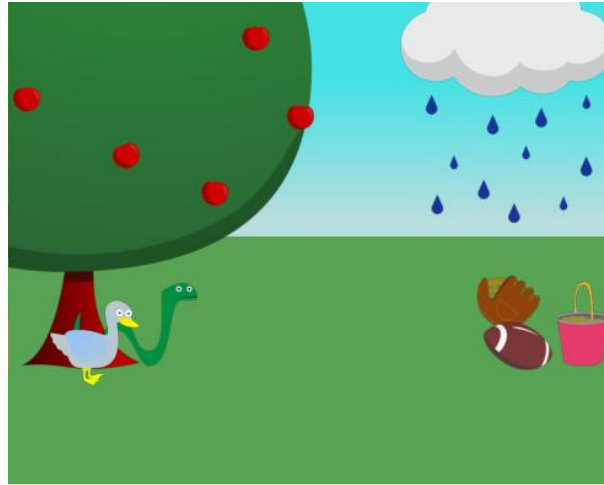
Mike fights off a bear by giving him a hotdog while jenny runs away.



Jenny and Mike are both playing dangerously in the park.

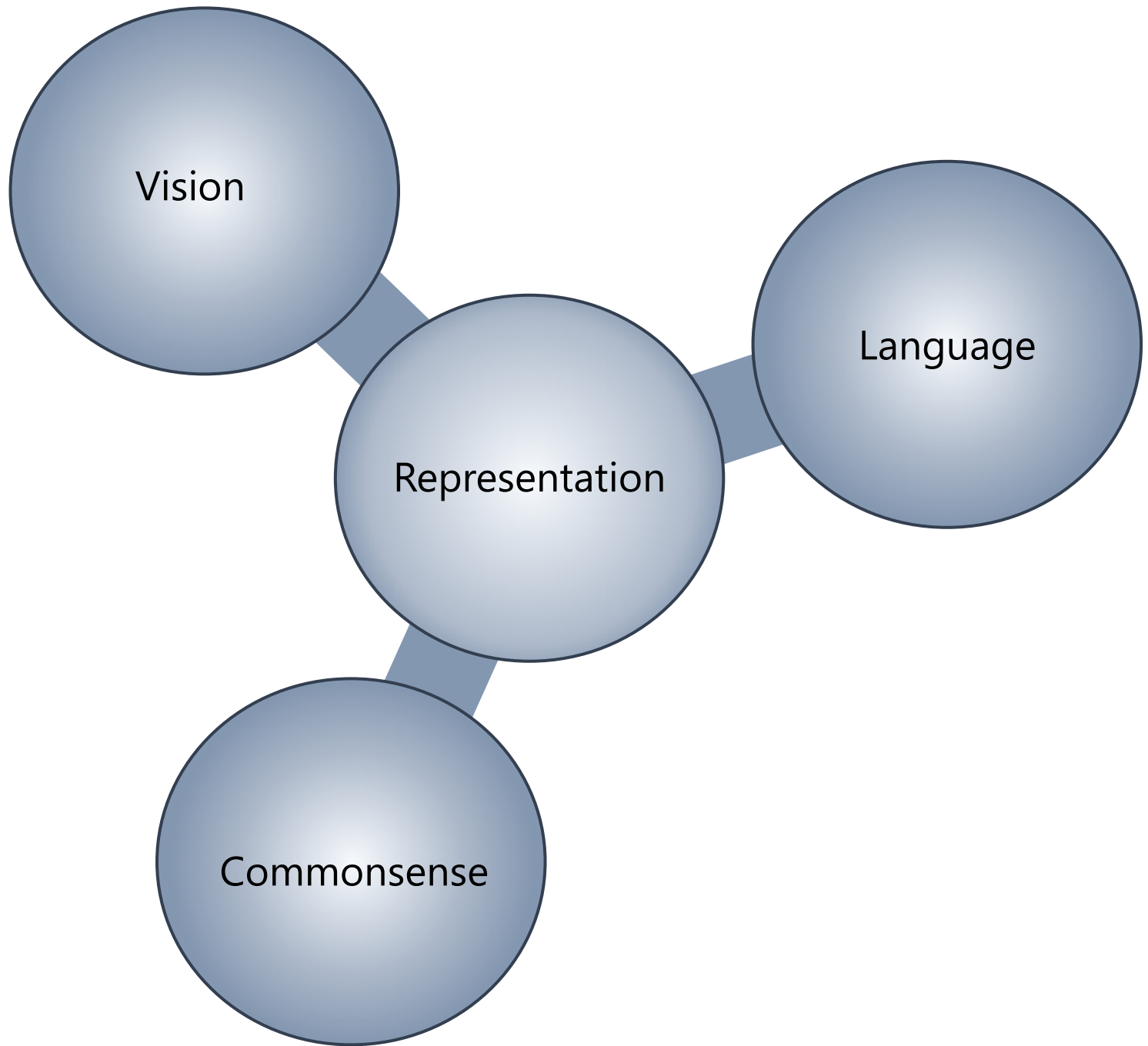


It was raining in the park and a duck and a snake were trying to take shelter.



Jenny and Mike are familiar = more stories
Simple visuals \neq simple description

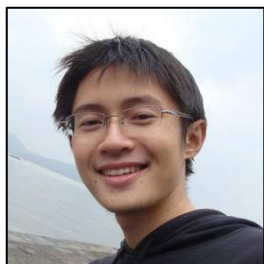






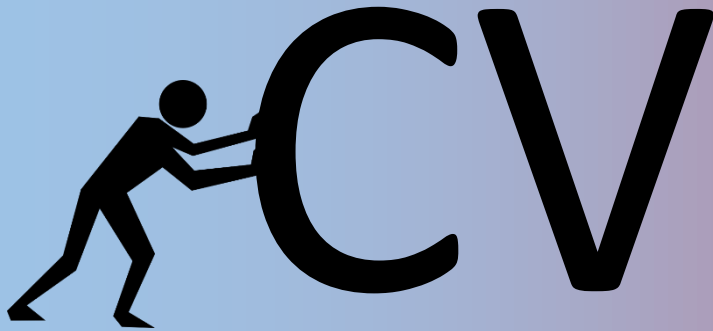






Xinlei Chen, CMU

Semantics



AI