

A note about NLP tools

- NLTK is primarily a teaching tool—its built-in taggers, parsers, etc. are not especially accurate.
- Open source software suites you may want to use for your projects:
 - ▶ Stanford CoreNLP (Java)
 - ▶ spaCy (Python)
 - ▶ Both support tokenization, NER, POS tagging, parsing in English and a few other languages. Stanford also supports coreference resolution, sentiment analysis.

Lecture 18

Semantic Role Labeling and Argument Structure

Nathan Schneider

ANLP | 13 November 2017

Language is flexible.

I'm thrilled to visit sunny California.

I'm thrilled to visit California, **where the weather is** sunny.

I'm thrilled to visit California, where **it's** sunny.

I'm **excited** to visit California, where it's sunny.

I'm excited to visit California, where it's sunny **out**.

I'm excited to **spend time in** California, where it's sunny out.

I'm **not** excited to visit sunny California.

I'm thrilled to visit sunny **Florida**.

I'm thrilled to visit sunny **Mountain View**.

I'm thrilled to visit California **because** it's sunny.

I'm **sort of happy about the** California visit.

אני נרגש לבקר קליפורניה שטופת שמש. 나는 맑은 캘리포니아를 방문 기뻐요.

Lexical Semantics

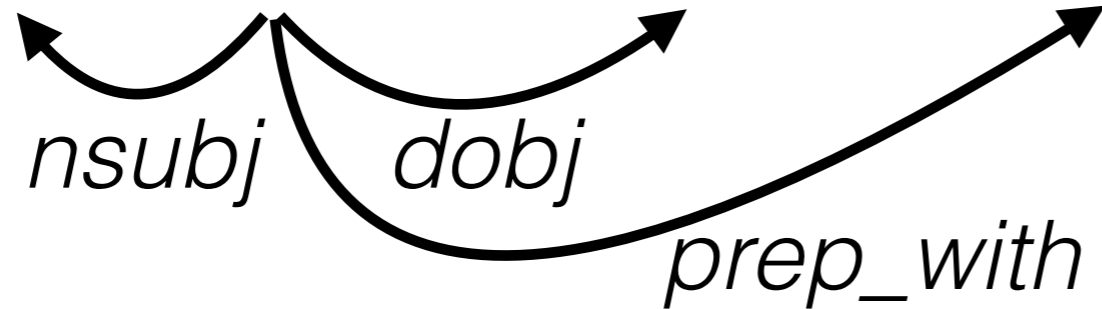
- So far, we've seen approaches that concern the **choice** of individual words:
 - sense disambiguation
 - semantic relations in a lexicon or similarity space
- Today: words that are fully understood by “**plugging in**” information from elsewhere in the sentence.
 - Specifically, understanding words that are (semantic) **predicates**, in relation to their **arguments**.
 - Especially **verbs**.
 - *Who did what to whom?*

Argument Structure Alternations

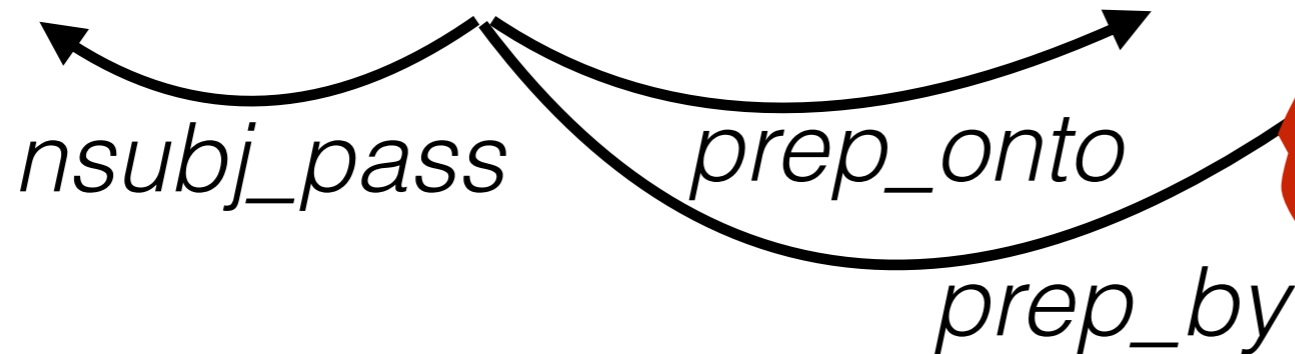
- Mary opened **the door**.
The door opened.
- John slices **the bread** with **a knife**.
The bread slices easily.
The knife slices easily.
- **Mary** loaded **the truck** with **hay**.
Mary loaded **hay** onto **the truck**.
The truck was loaded with **hay** (by **Mary**).
Hay was loaded onto **the truck** (by **Mary**).
- **John** got **Mary** a present.
John got a present for **Mary**.
Mary got a present from **John**.

Stanford Dependencies

- **Mary** loaded **the truck** with **hay**.



- **Hay** was loaded onto **the truck** by **Mary**.



Syntax is not enough!

Syntax-Semantics Relationship

Add another family member

Relationship Status:

Interested in:

Looking for:

- Single
- In a Relationship
- Engaged
- Married
- It's Complicated**
- In an Open Relationship
- Widowed

Networking

Political Views:

Religious Views:

Outline

- Syntax \neq semantics
 - The **semantic roles** played by different participants in the sentence are not trivially inferable from syntactic relations
 - ...though there are patterns!
- Two computational datasets/approaches that describe sentences in terms of semantic roles:
 - PropBank — simpler, more data
 - FrameNet — richer, less data
- The idea of semantic roles can be combined with other aspects of meaning. Glimpse of **AMR**, which is one way to do this.

PropBank

- Abstracts away from syntax to predicate-argument structures

PropBank

Mary loaded the truck with hay at the depot on Friday.

- *load*: load.01 ‘cause to be burdened’

Roles:

Arg0-PAG: loader, agent

Arg1-GOL: beast of burden

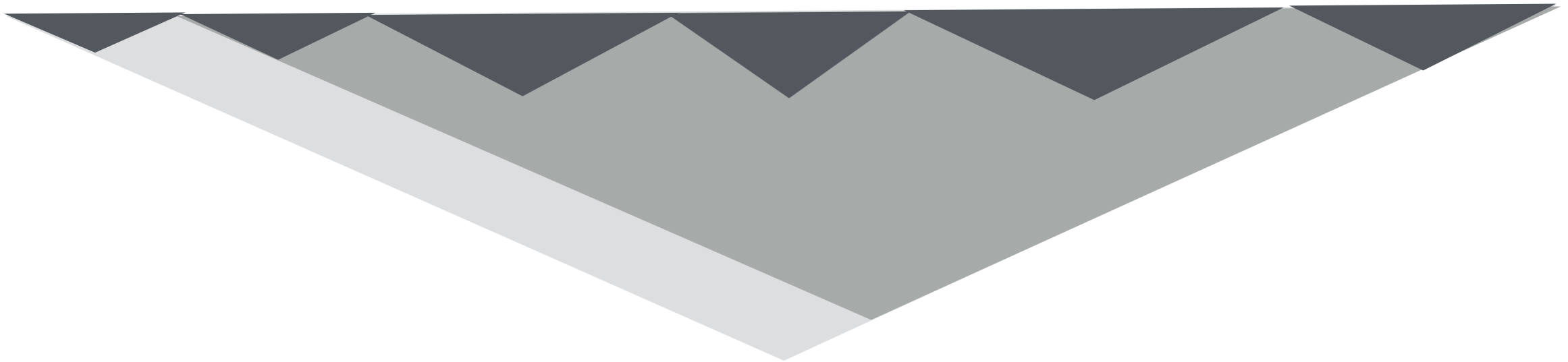
Arg2-PPT: cargo

Arg3-MNR: instrument

- *load_up*: load.02 ‘phrasal cause to be burdened’
- *load*: load.03 ‘fix, set up to cheat’

PropBank

Mary loaded the truck with hay at the depot on Friday.



PropBank

Mary **loaded** the truck with hay at the depot on Friday.



load.01

A0 loader

A1 bearer

A2 cargo

A3 instrument

AM-LOC

AM-TMP

AM-PRP

AM-MNR

...

PropBank

Mary **loaded** the truck with hay at the depot on Friday.



load.01

A0 loader

A1 bearer

A2 cargo

A3 instrument

AM-LOC

AM-TMP

AM-PRP

AM-MNR

...

PropBank

Mary **loaded** the truck with hay at the depot on Friday.



load.01

A0 loader

A1 bearer

A2 cargo

A3 instrument

AM-LOC

AM-TMP

AM-PRP

AM-MNR

...

PropBank

Mary **loaded** the truck with hay at the depot on Friday.



load.01

A0 loader

A1 bearer

A2 cargo

A3 instrument

AM-LOC

AM-TMP

AM-PRP

AM-MNR

...

PropBank

Mary **loaded** the truck with hay at the depot on Friday.



load.01

A0 loader

A1 bearer

A2 cargo

A3 instrument

AM-LOC

AM-TMP

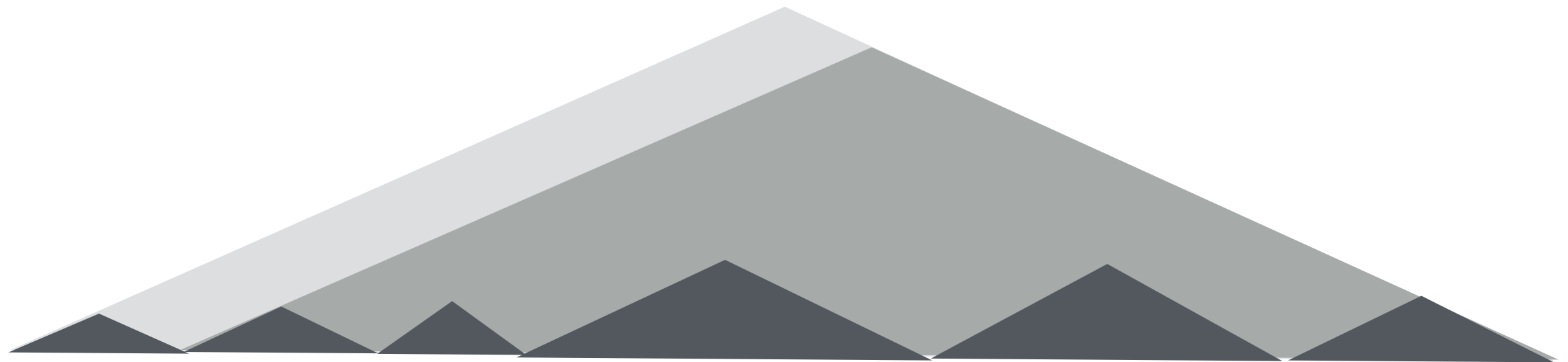
AM-PRP

AM-MNR

...

PropBank

Mary **loaded** the truck with hay at the depot on Friday.



Mary loaded hay onto the truck at the depot on Friday.

PropBank

Mary **loaded** the truck with hay at the depot on Friday.

load.01

A0 loader

A1 bearer

A2 cargo

A3 instrument

AM-LOC

AM-TMP

AM-PRP

AM-MNR

...



Mary **loaded** hay onto the truck at the depot on Friday.

PropBank

Mary **loaded** the truck with hay at the depot on Friday.

load.01

A0 loader

A1 bearer

A2 cargo

A3 instrument

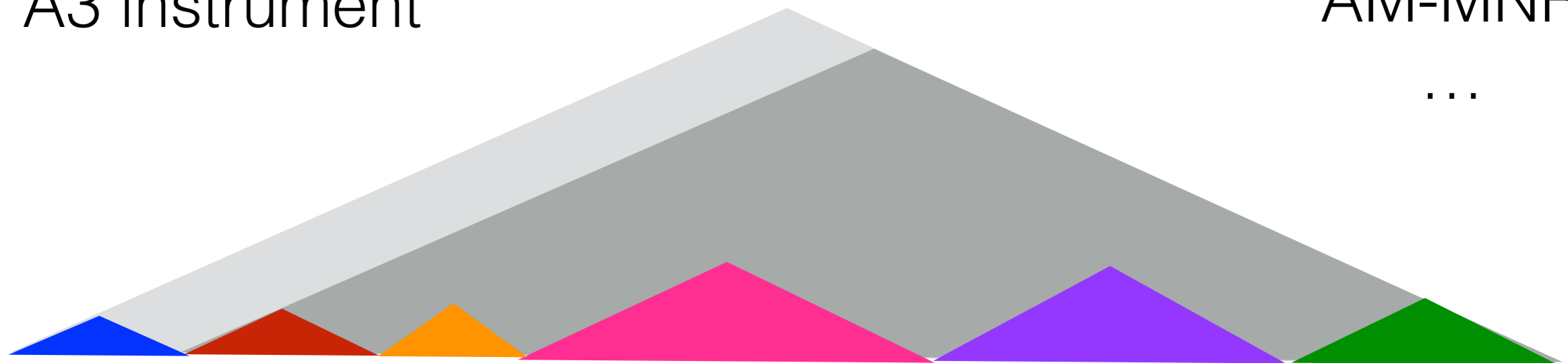
AM-LOC

AM-TMP

AM-PRP

AM-MNR

...



Mary **loaded** hay onto the truck at the depot on Friday.

PropBank

- Abstracts away from syntax to predicate-argument structures
- Predicate-argument lexicon + annotations of full WSJ PTB corpus and other data (such as OntoNotes)
- Originally **verbs** only (Kingsbury & Palmer 2002); now has many nouns, adjectives, light verb constructions, etc. (Bonial et al. 2014)
- Strongly **lexicalized**: no synonymy, hypernymy, etc. of predicates with different stems; very coarse-grained sense distinctions
- Phrase structure constituents of PTB(-style) trees



Argument Structure Alternations

- Mary opened **the door**.
The door opened.
- John slices **the bread** with **a knife**.
The bread slices easily.
The knife slices easily.
- **Mary** loaded **the truck** with **hay**.
Mary loaded **hay** onto **the truck**.
The truck was loaded with **hay** (by **Mary**).
Hay was loaded onto **the truck** (by **Mary**).
- **John** got **Mary** a present.
John got a present for **Mary**.
Mary got a present from **John**.



PropBank

Semantic Role Labeling

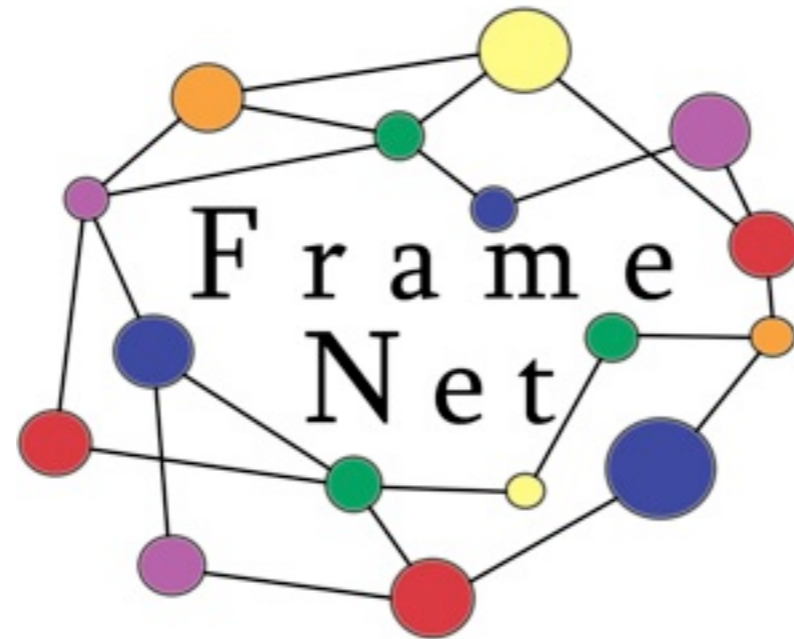
- Traditional pipeline:
 1. (Assume syntactic parse and predicate senses as given)
 2. **Argument identification:** select the predicate's argument phrases
 3. **Argument classification:** select a role for each argument
useful feature: predicate \rightarrow^* argument path in tree
- See Palmer et al. 2010 for a review

Limitation of PropBank

- Numbered roles (ARG0, ARG1, etc.) are predicate-specific.
- load.ARG1: beast of burden, whereas
- put.ARG1: thing put
- load.ARG1 corresponds to put.ARG2

Thematic Roles

- Linguists talk about general classes of semantic roles:
 - ▶ *Agent* = animate entity who is volitionally acting
 - ▶ *Theme* = participant that is undergoing motion, for example
 - ▶ *Patient* = participant that undergoes some internal change of state (e.g., breaking)
 - ▶ *Destination* = intended endpoint of motion
 - ▶ *Recipient* = party to which something is transferred
- The **VerbNet** resource uses these and a couple dozen other roles.
- But it is hard to come up with a small list of these roles that will suffice for all verbs.
- And there are correspondences that these roles do not expose: e.g., that someone who *buys* is on the receiving end of *selling*.



Berkeley FrameNet

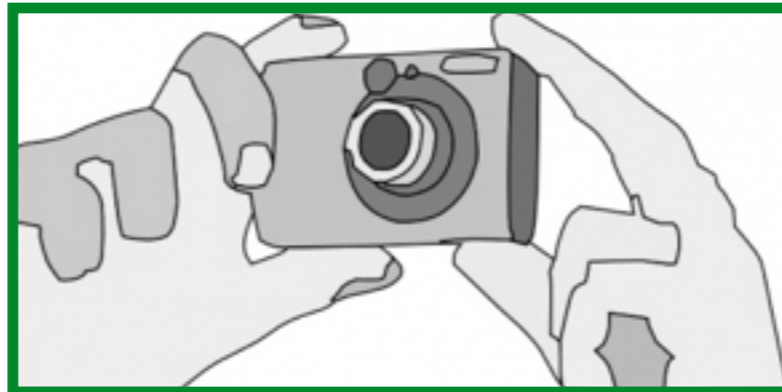
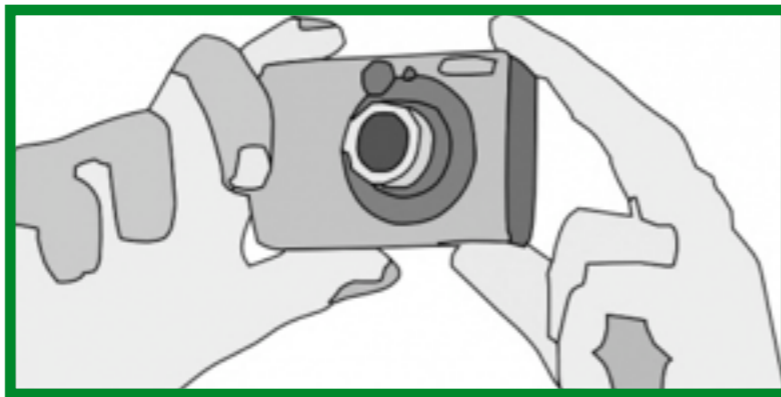
<https://framenet.icsi.berkeley.edu/>

Paraphrase

- James snapped a photo of me with Sheila.
- Sheila and I had our picture taken by James.

What's in common

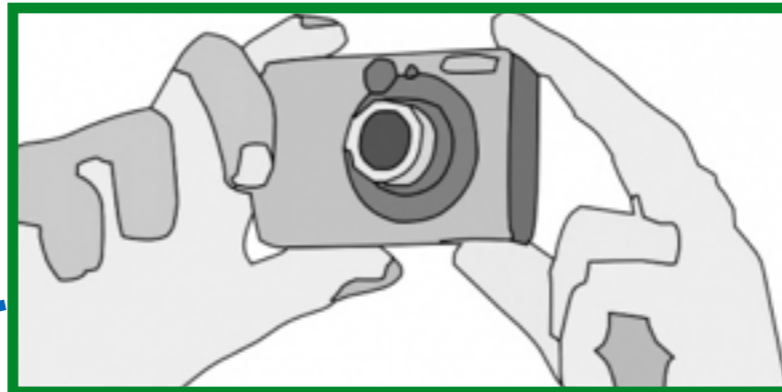
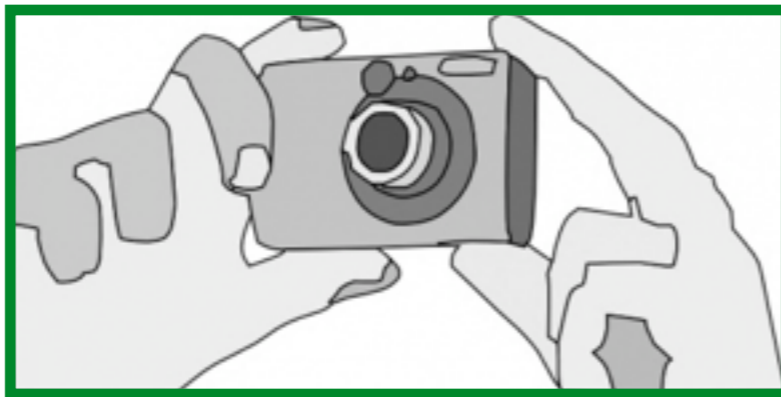
- James **snapped a photo** of me with Sheila.



- Sheila and I had our **picture taken** by James.

What's in common

- James **snapped a photo** of me with Sheila.



- Sheila and I had our **picture taken** by James.

Idealized Stanford Dependencies

- James snapped a photo of me with Sheila.

nsubj(snap, James)
dobj(snap, photo)
prep_of(photo, me)
prep_with(me, Sheila)
det(photo, a)

- Sheila and I had our picture taken by James.

nsubjpass(taken, Sheila)
nsubjpass(taken, I)
conj_and(Sheila, I)
aux(taken, had)
dobj(taken, picture)
poss(picture, our)
agent(taken, James)

Frame Semantics

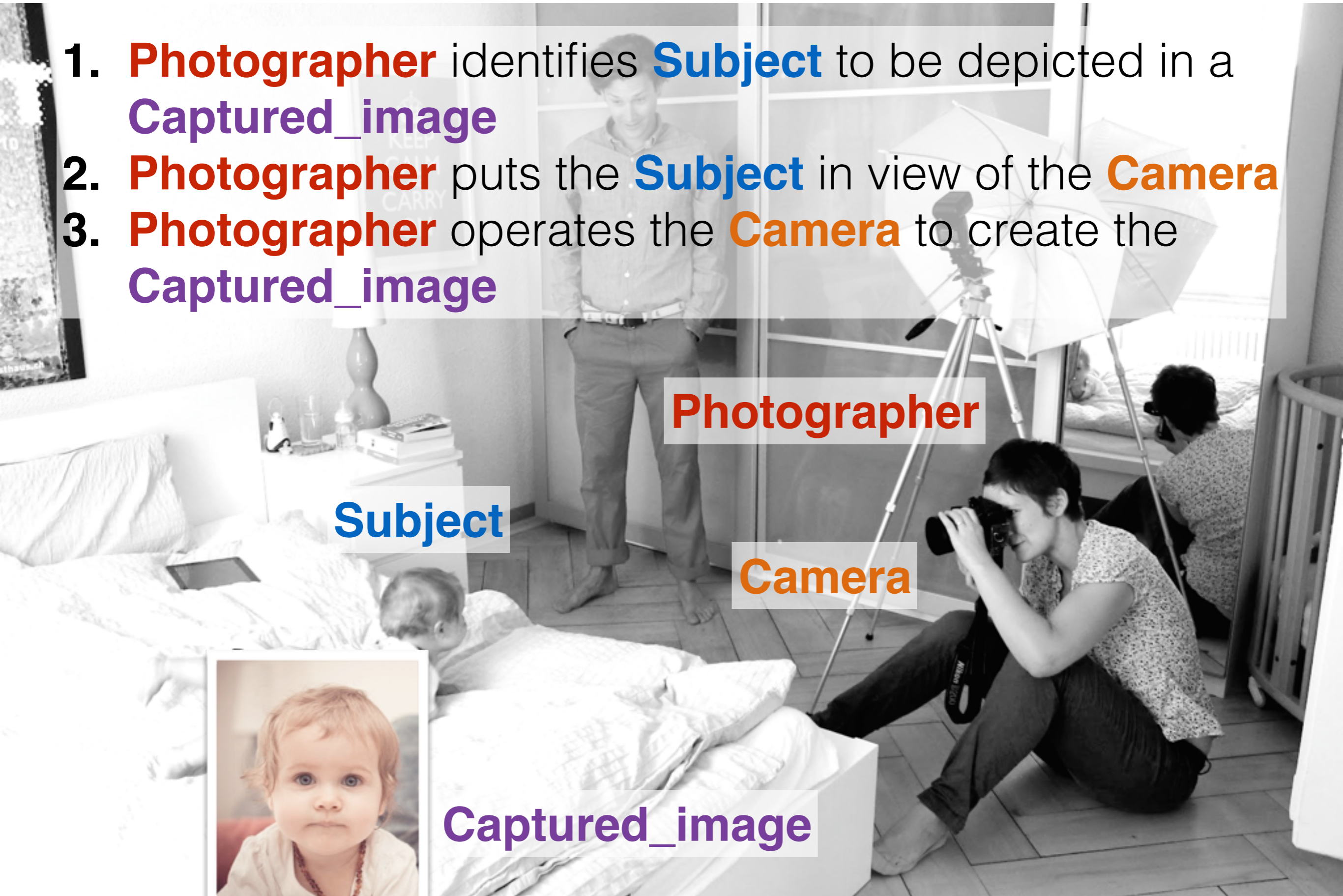
“MEANINGS ARE RELATIVIZED
TO SCENES”

(Fillmore 1977)





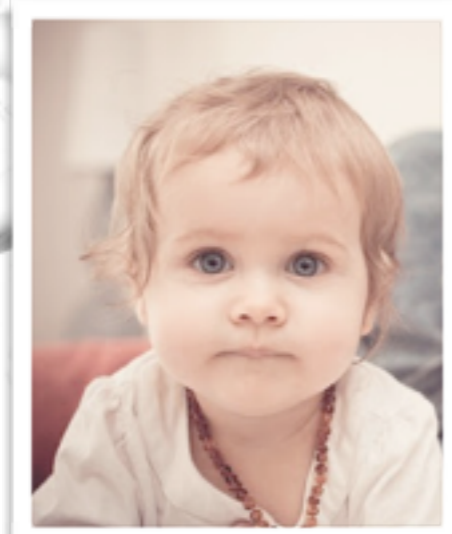
1. **Photographer** identifies **Subject** to be depicted in a **Captured_image**
2. **Photographer** puts the **Subject** in view of the **Camera**
3. **Photographer** operates the **Camera** to create the **Captured_image**



Subject

Photographer

Camera



Captured_image

1. **Photographer** identifies **Subject** to be depicted in a **Captured_image**
2. **Photographer** puts the **Subject** in view of the **Camera**
3. **Photographer** operates the **Camera** to create the **Captured_image**

Photographer

time

manner

Subject

duration

location

Camera

frequency

reason

Captured_image

photograph.v take ((picture)).v snap picture.v

frame name

textual definition explaining the scene and how the **frame elements** relate to one another

Core

non-core

Frame

FEs

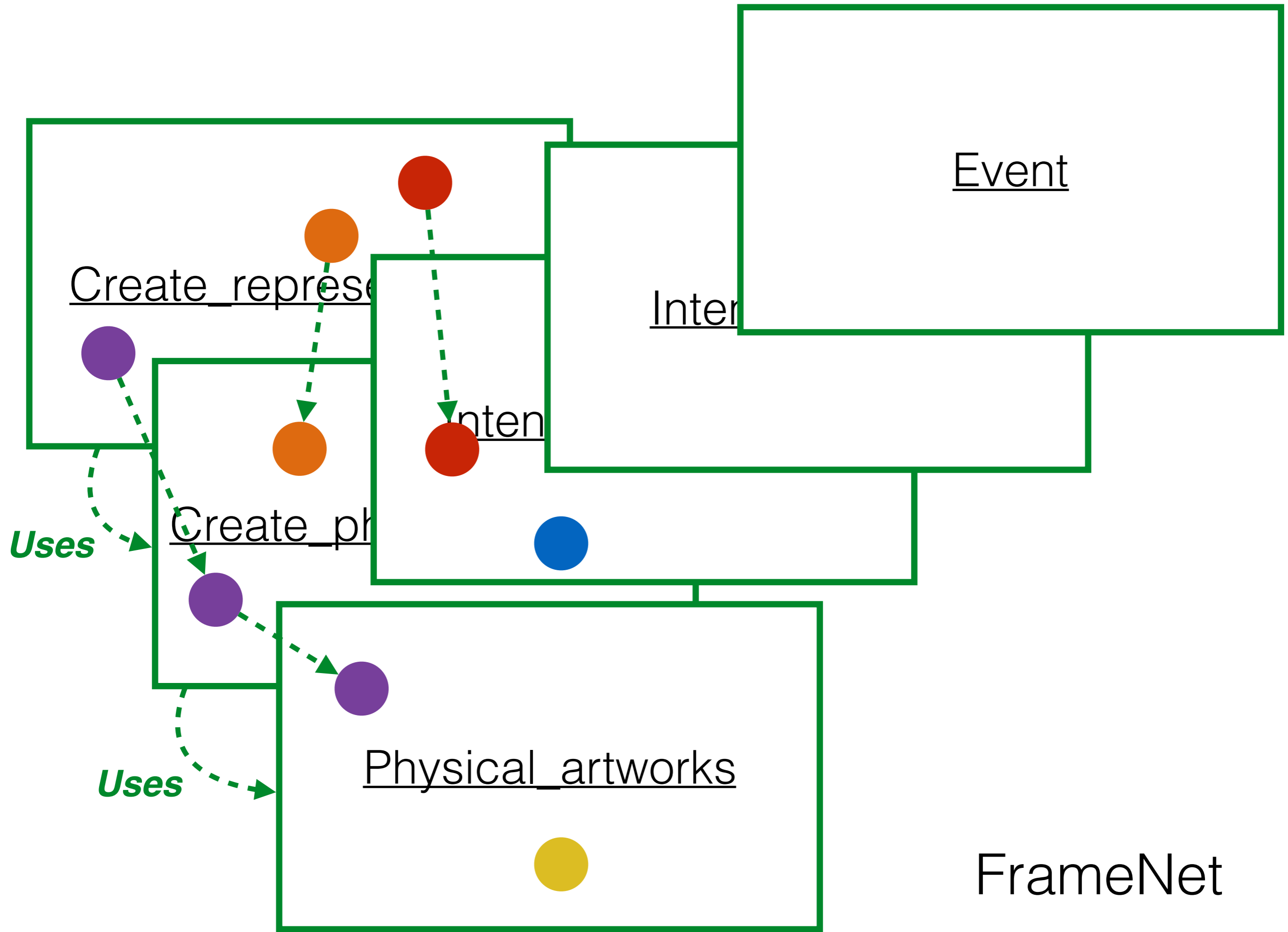
Elements

predicate1.v

predicate2.n

predicate3.a





FrameNet: Lexicon

- ~1000 **frames** represent scenarios. Most are associated with **lexical units** (a.k.a. **predicates**). Berkeley FrameNet currently has 13k LUs (5k nouns, 5k verbs, 2k adjectives).
- **Frame elements** (a.k.a. **roles**) represent participants/components of those scenarios. **Core** vs. **non-core**.
- Frames and their corresponding roles are linked together in the lexicon.
- Frames are explained with textual descriptions.

Create_physical_artwork

Definition:

A **Creator** creates an artifact that is typically an iconic **Representation** of an actual or imagined entity or event. The **Representation** may also be evocative of an idea while not based on resemblance.

Diagrams must be **clearly DRAWN** on construction paper. **CNI**

I TOOK his picture and told him that if it came out well I would make him a copy.

In about 1305 and 1306 **Giotto PAINTED** a notable series of 38 frescoes.

FEs:

Core:

Creator [cre]

Semantic Type: Sentient

An individual or individuals that bring the **Representation** into existence.

Supposedly, **the artist DREW** the picture from memory.

Representation [rep]

The entity that is created to represent either iconically or abstractly.

Most of us know where we **TOOK** a photo but have a harder time remembering the time we took it.

Non-Core:

Depictive [dep]

This FE describes the **Creator** as being in some state during the creation of the **Representation**.

Descriptor []

A characteristic of the **Creator** or the **Representation**.

Lexical Units:

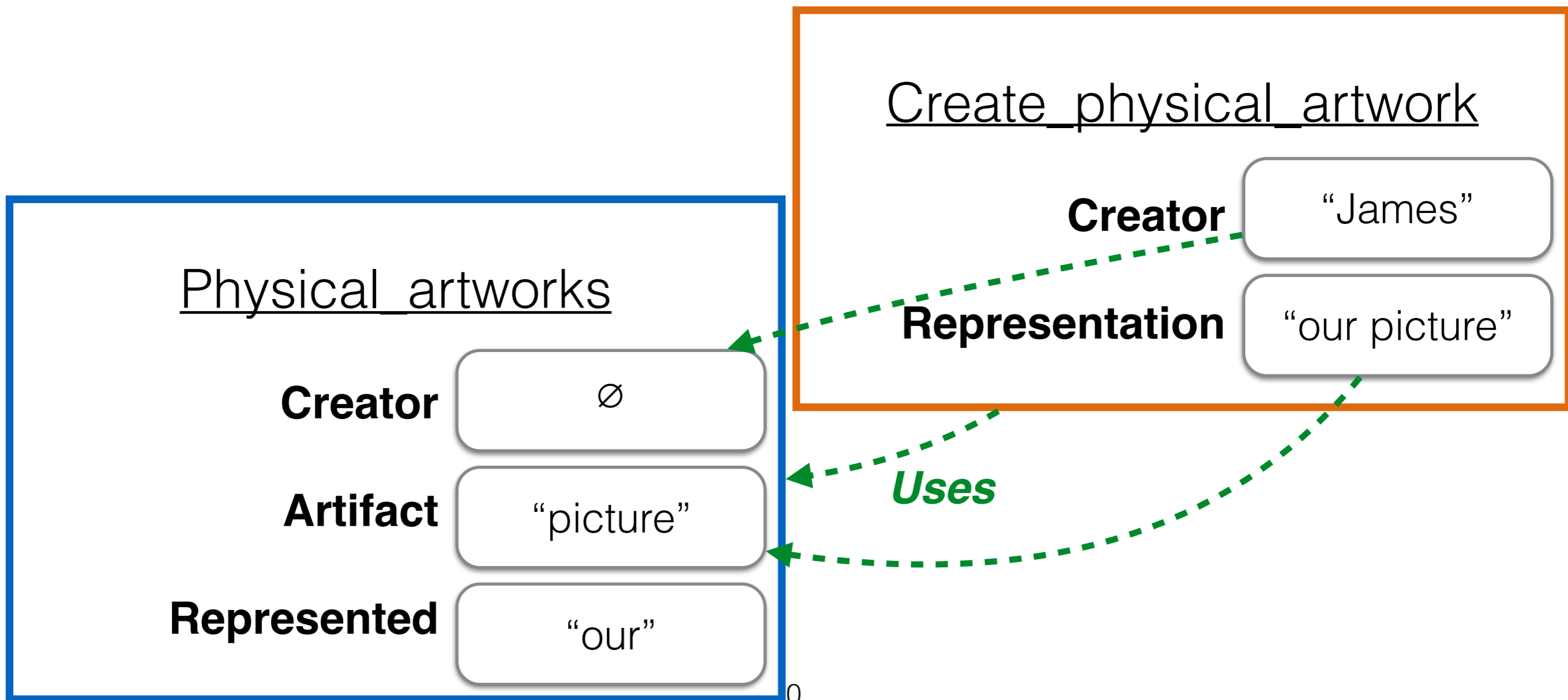
artist.n, cast.v, draw.v, paint.v, sculpt.v, take_((picture)).v

Created by 605 on 11/21/2005 03:47:00 PST Mon

<u>Lexical Unit</u>	<u>LU Status</u>	<u>Lexical Entry Report</u>	<u>Annotation Report</u>	<u>Annotator ID</u>	<u>Created Date</u>
artist.n	Created	Lexical entry	Annotation	361	03/28/2007 03:10:10 PDT Wed
cast.v	Created	Lexical entry		597	06/09/2008 01:41:45 PDT Mon
draw.v	Finished_Initial	Lexical entry	Annotation	605	11/21/2005 05:28:34 PST Mon
paint.v	Finished_Initial	Lexical entry	Annotation	605	11/21/2005 05:26:23 PST Mon
sculpt.v	Created	Lexical entry		597	05/23/2008 02:55:21 PDT Fri
take_((picture)).v	Created	Lexical entry		605	11/21/2005 05:29:24 PST Mon

FrameNet Annotations

- Sheila and I had our **picture taken** by James.



Languages with FrameNets



SRL Demos

- Mateplus/Framat (PropBank, FrameNet):
<http://www.coli.uni-saarland.de/~mroth/demo.html>
- SEMAFOR (FrameNet):
<http://demo.ark.cs.cmu.edu/parse>
- AllenNLP (PropBank): <http://demo.allennlp.org/>

Advanced
Topic



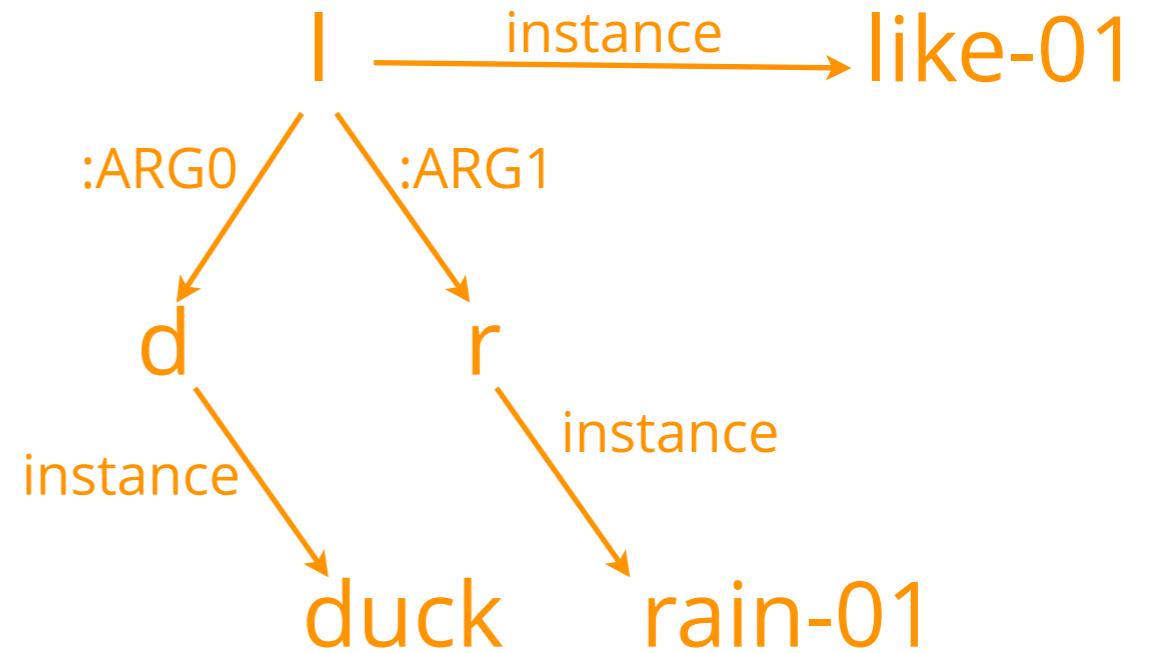
Abstract Meaning Representation

(Banarescu et al., LAW 2013)

A graph-based representation of lexical **concepts** and typed **relations** between those concepts that are denoted by an English sentence.

AMR integrates several aspects of lexical/relational meaning—abstracting away from the grammatical details—in a **single structure** designed to support rapid corpus annotation and data-driven NLP.

AMRs

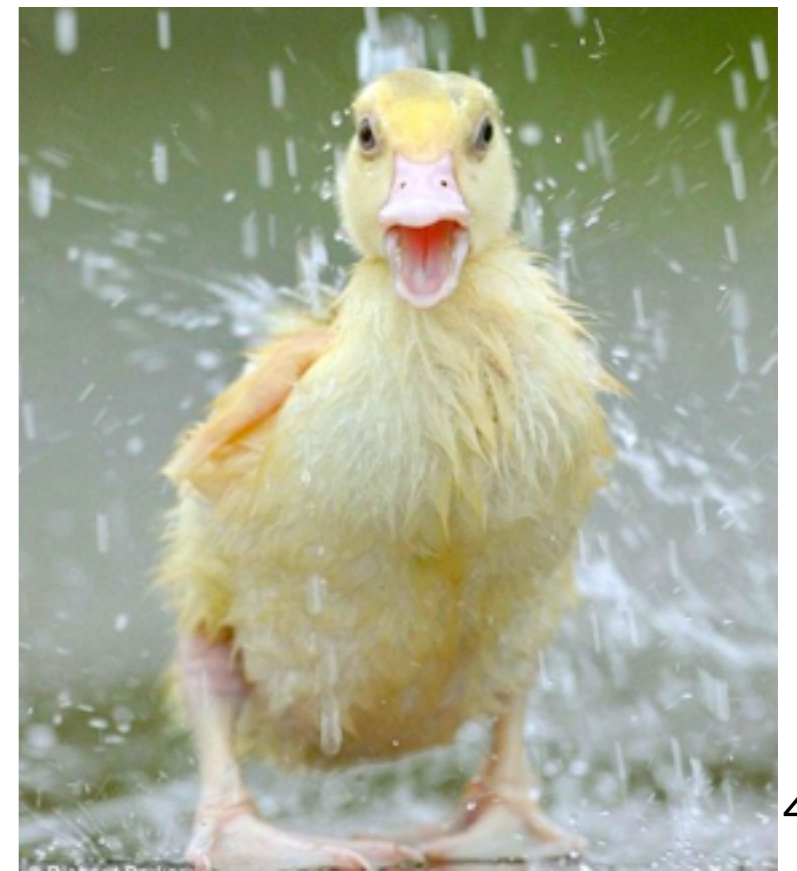


(I / like-01

:ARG0 (d / duck)

:ARG1 (r / rain-01))

- ▶ ducks like rain
- ▶ the duck liked that it was raining



(l / like-01

:ARG0 (d / duck)

:ARG1 (r / rain-01))

AMRs

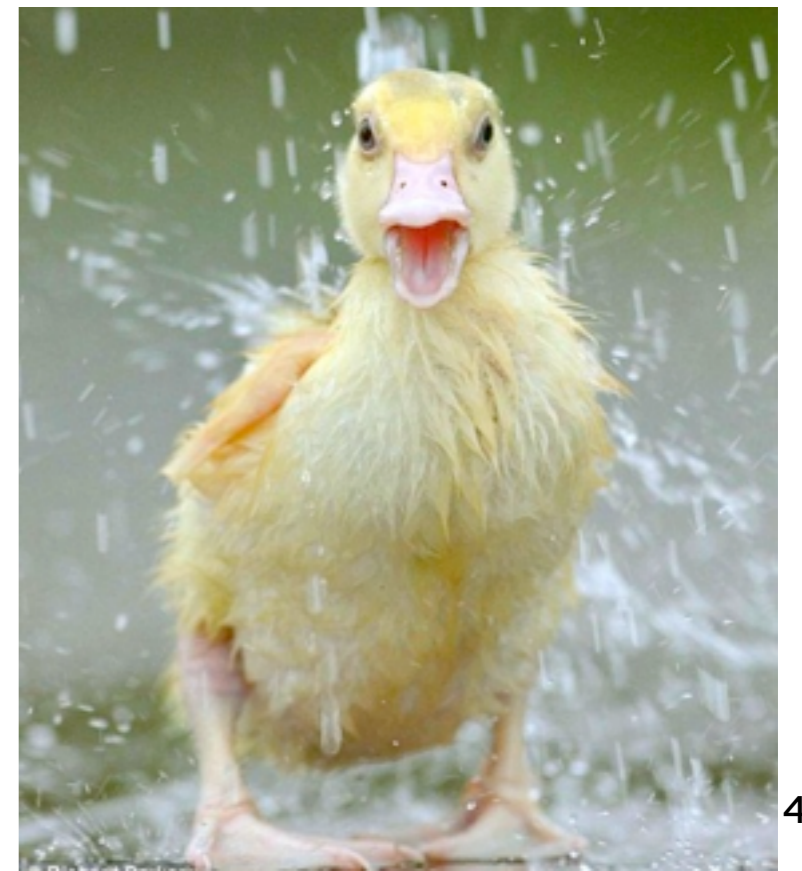
(s2 / see-01

:ARG0 (i / i)

:ARG1 (d / duck

:poss (s / she)))

▶ I saw her duck



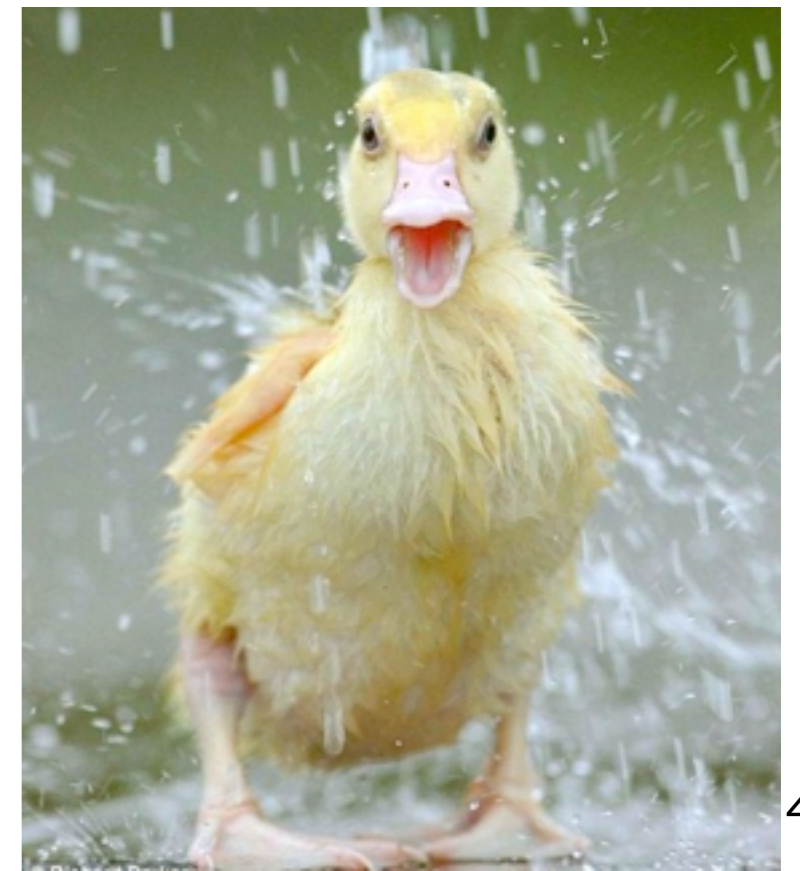
(l / like-01
:ARG0 (d / duck)
:ARG1 (r / rain-01))

AMRs

(s2 / see-01
:ARG0 (i / i)
:ARG1 (d / duck
:poss (s / she)))

(s2 / see-01
:ARG0 (i / i)
:ARG1 (d / duck-01
:ARG0 (s / she)))

▶ I saw her duck (alternate interpretation)

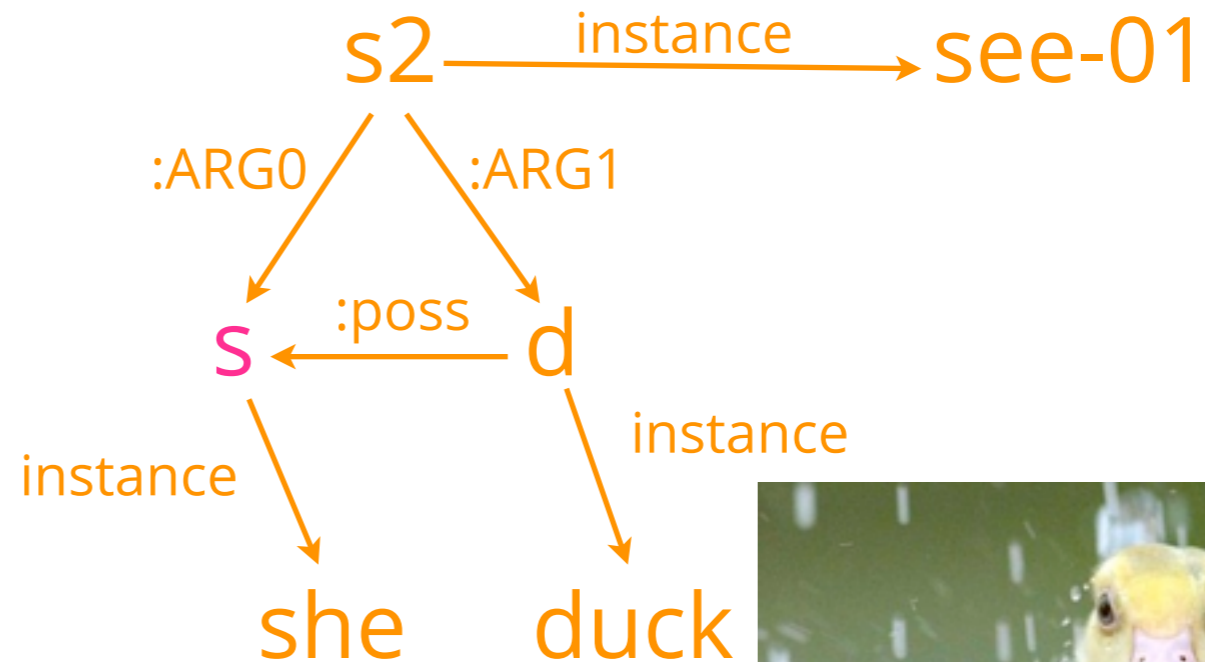


(l / like-01
:ARG0 (d / duck)
:ARG1 (r / rain-01))

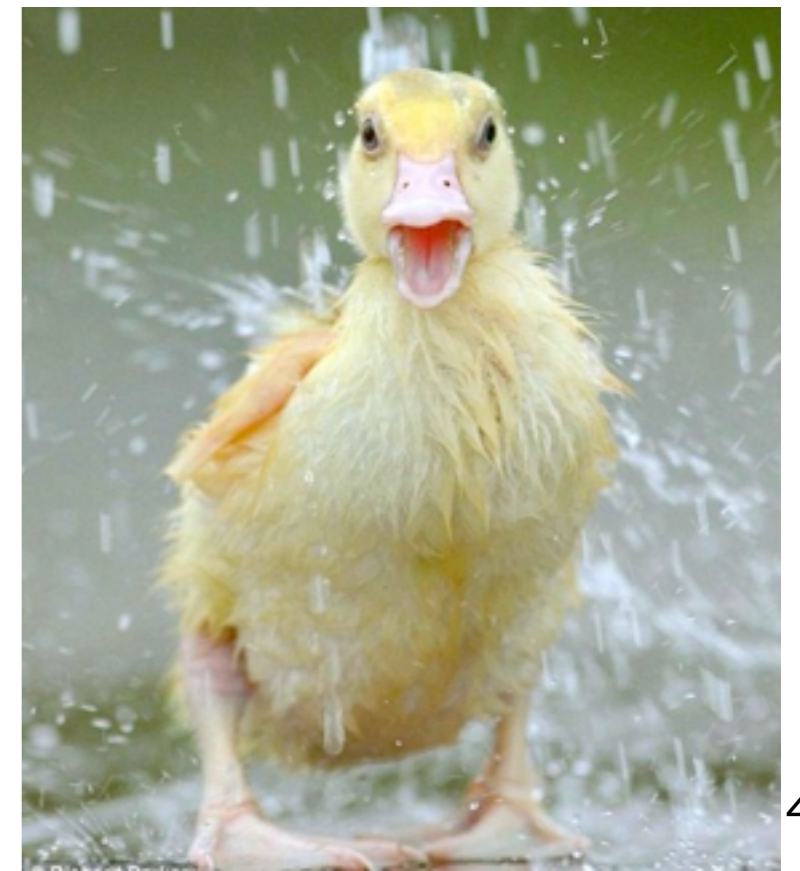
AMRs

(s2 / see-01
:ARG0 (i / i)
:ARG1 (d / duck
:poss (s / she)))

(s2 / see-01
:ARG0 (s / she)
:ARG1 (d / duck
:poss s))



▶ She saw her (own) duck

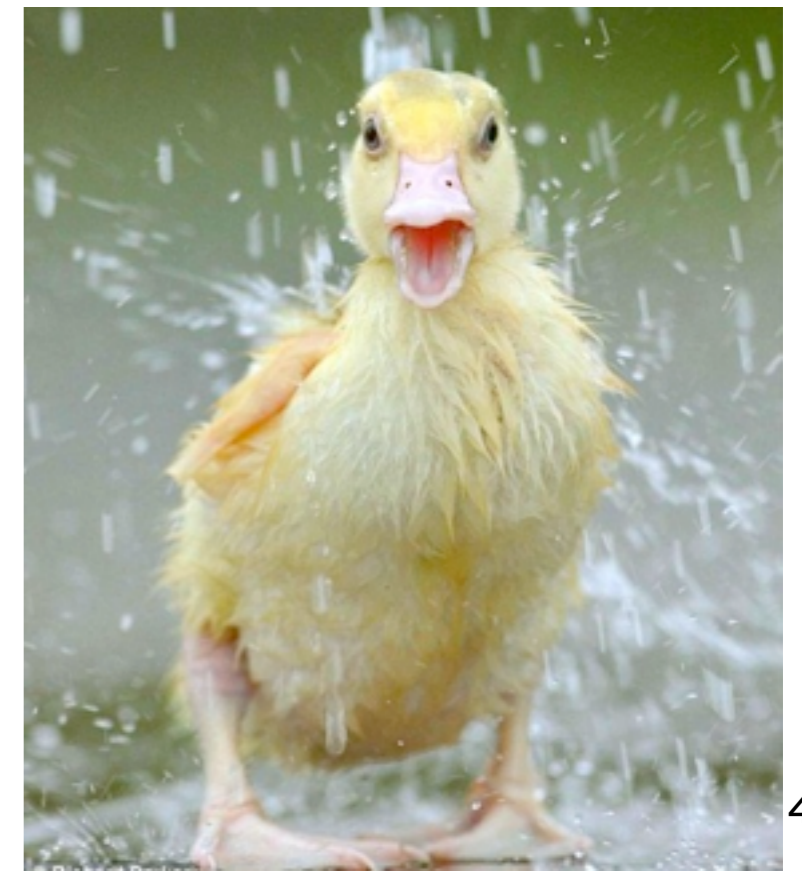
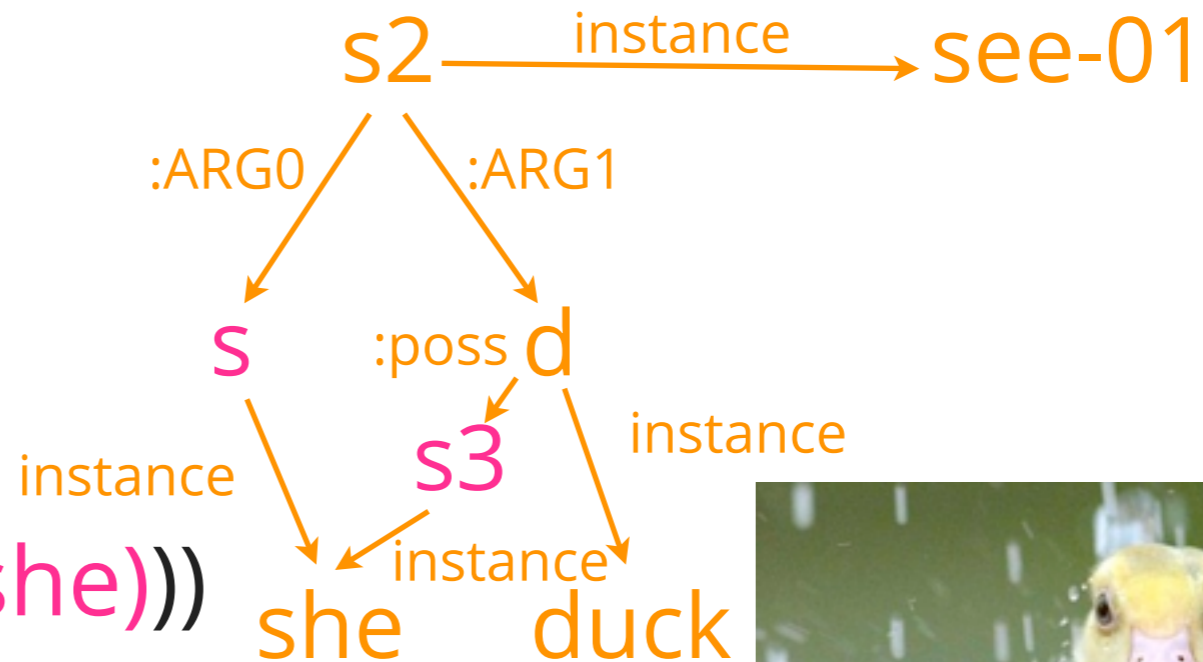


(l / like-01
:ARG0 (d / duck)
:ARG1 (r / rain-01))

AMRs

(s2 / see-01
:ARG0 (i / i)
:ARG1 (d / duck
:poss (s / she)))

(s2 / see-01
:ARG0 (s / she)
:ARG1 (d / duck
:poss (s3 / she)))



▶ She saw her (someone else's) duck

(l / like-01

AMRs

:ARG0 (d / duck)

:ARG1 (r / rain-01))

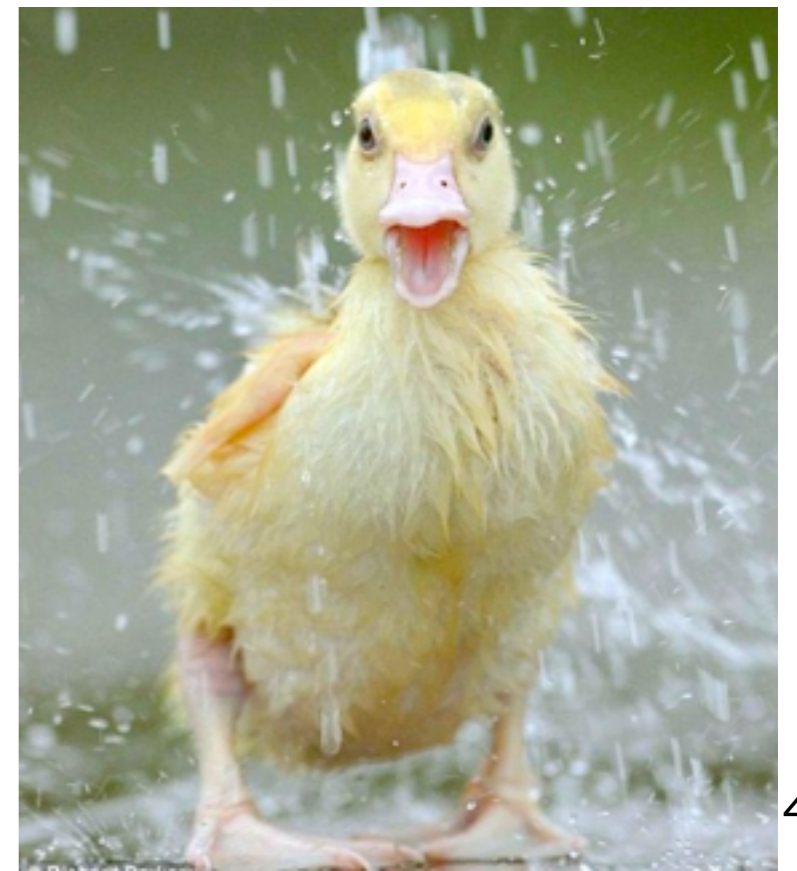
(h / happy

:domain (d / duck

:ARG0-of (l / like-01

:ARG1 (r / rain-01))))

- ▶ Ducks who like rain are happy



(l / like-01

AMRs

:ARG0 (d / duck)

:ARG1 (r / rain-01))

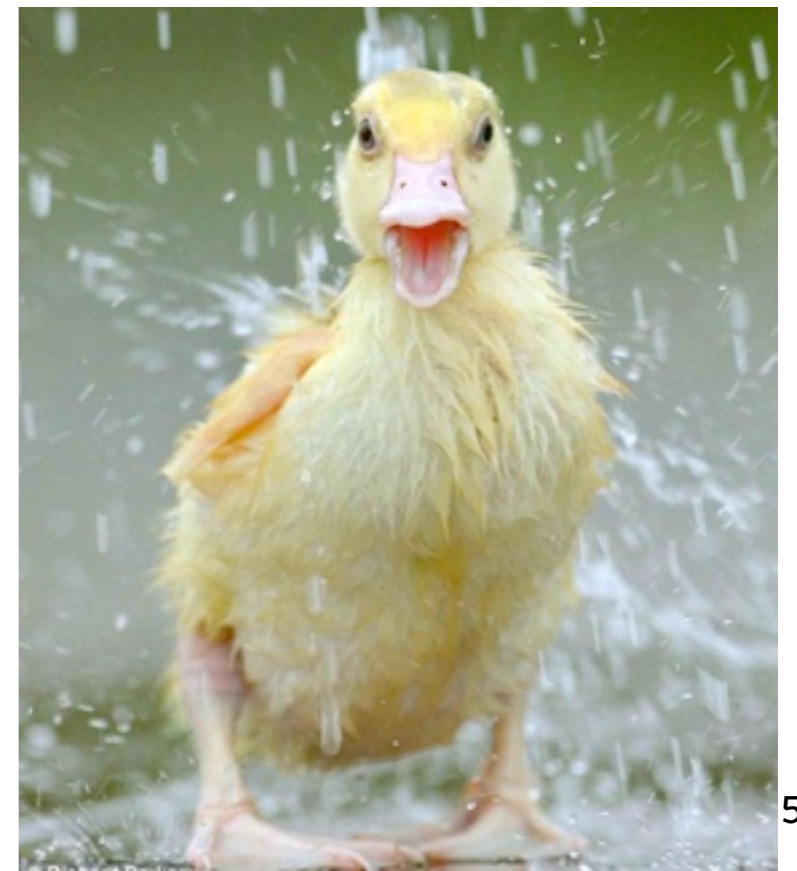
(h / happy

:domain (d / duck

:ARG0-of (l / like-01

:ARG1 (r / rain-01))))

- ▶ Ducks who like rain are happy



(l / like-01

:ARG0 (d / duck)

:ARG1 (r / rain-01))

AMRs

(h / happy

:domain (d / duck

:ARG0-of (l / like-01

:ARG1 (r / rain-01))))

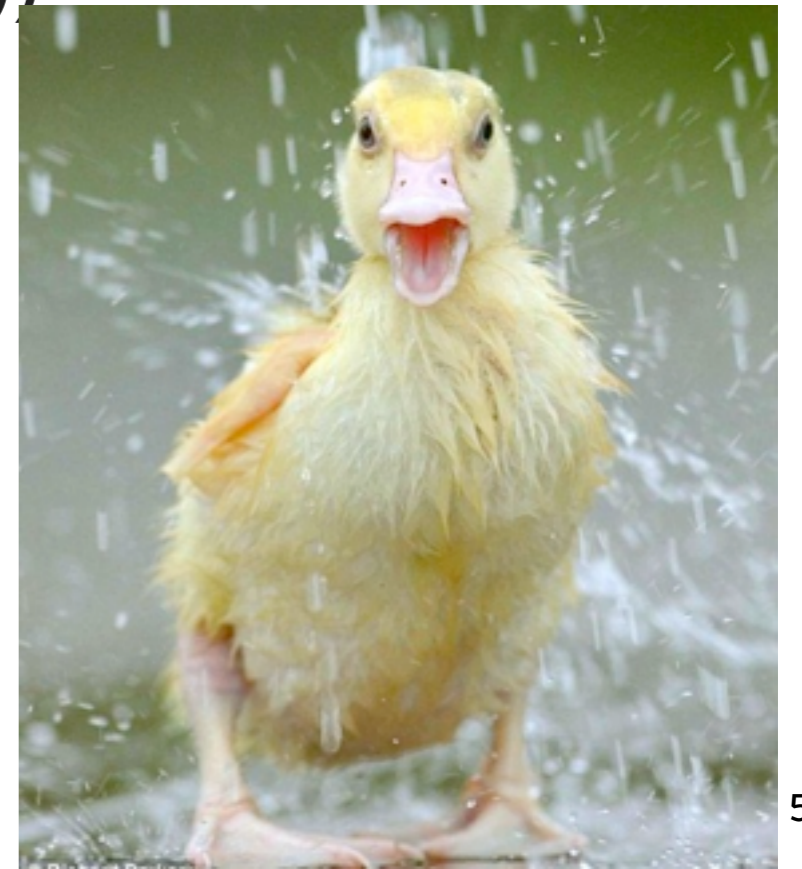
(l / like-01

:ARG0 (d / duck

:domain-of/:mod (h / happy))

:ARG1 (r / rain-01))

▶ Happy ducks like rain



Police release security footage of the man they believe assaulted a 12-year-old in her home.

```
(r / release-01
  :ARG0 (p / police)
  :ARG1 (f / footage
    :mod (s / security)
    :topic (m / man
      :ARG0-of (a / assault-01
        :ARG1 (g / girl
          :age (t / temporal-quantity :quant 12
            :unit (y / year)))
        :ARG1-of (b / believe-01
          :ARG0 p)
        :location (h / home
          :poss g))))))
```

Police release security footage of the man they believe assaulted a 12-year-old in her home.

```
(r / release-01
  :ARG0 (p / police)
  :ARG1 (f / footage
    :mod (s / security)
    :topic (m / man
      :ARG0-of (a / assault-01
        :ARG1 (g / girl
          :age (t / temporal-quantity :quant 12
            :unit (y / year)))
        :ARG1-of (b / believe-01
          :ARG0 p)
          :location (h / home
            :poss g))))))
```

AMR Features

- PropBank predicate-argument semantics
- name & value entities; entity linking (wikification)
- coreference **entities & events**
- modality, negation, questions
- relations between nominals ***history teacher* → (p / person :ARG0-of (t / teach-01 :ARG1 (h / history)))**
- canonicalization of content words (remove inflectional morphology, convert adv → adj → noun → verb where possible)
- ...*all in a single graph!* ***his trial* → (t / try-02 :ARG1 (h / he))**

AMR Assets

- Snazzy annotation tool
- Evaluation method (smatch)
- Extensive documentation (guidelines, help pages in tool, heuristics in tool)
- Tutorial: <https://github.com/nschneid/amr-tutorial>
- Close coordination with PropBank
- Annotation sites: CU, ISI, SDL, LDC
- Data: ~40,000 AMRs released (as of 2016)

Abstract Meaning Representation (AMR)

(Banarescu et al., LAW 2013)

A graph-based representation of lexical **concepts** and typed **relations** between those concepts that are denoted by an English sentence.

AMR integrates several aspects of lexical/relational meaning—abstracting away from the grammatical details—in a **single structure** designed to support rapid corpus annotation and data-driven NLP.

(Flanigan et al., ACL 2014)



AMR Parsing: JAMR

- Open source system from CMU
- Pipeline:
 1. **Preprocessing:** dependency parsing, NER
 2. **Concept identification:** map word sequences to graph fragments
 3. **Relation identification:** connect the fragments into a rooted DAG (novel MSCG algorithm)
- See Flanigan et al. 2014 for details



```
(g / girl
(a / assaulted @ 12-year-old
:age (t / temporal-quantity :quant 12
:unit (y / year)))
```

```
(a / assault-01
:ARG1 (g / girl
:age (t / temporal-quantity :quant 12
:unit (y / year))))
```

Summary

- For verbs (and other semantic predicates), there are complicated patterns of **argument structure**—how semantic **arguments/roles** correspond to syntactic slots.
- Lexicons formalize this in different ways: **PropBank**, **VerbNet**, **FrameNet**
 - Corpora annotated according to each of these lexicons for training **semantic role labelers**.
 - **FrameNet** is the richest theory (deep frames), but that imposes practical limits on the size of the lexicon and annotated corpora.
 - **PropBank** has good coverage of English verbs, and large amount of annotated corpora (WSJ + more!). But a bit superficial (verb-specific frames).
- PropBank event predicates are used in **AMR**, a meaning representation that also captures named entities, negation/modality, coreference, and other aspects of semantics in a graph for each sentence.