

# Foundations of Natural Language Processing

## Lecture 16

### Semantic Role Labelling and Argument Structure

Alex Lascarides

(Slides based on those of Schneider, Koehn, Lascarides)

13 March 2020



## Argument Structure and Alternations

- Mary opened **the door**  
The door opened
- John slices **bread** with **a knife**  
This bread slices easily  
The knife slices cleanly
- Mary loaded **the truck** with **hay**  
Mary loaded **hay** onto the **the truck**  
The truck was loaded with **hay** (by Mary)  
The hay was loaded onto **the truck** (by Mary)
- John gave **a present** to **Mary**  
John gave **Mary** a present

## Language is Flexible

- Often we want to know *who did what to whom (when, where, how and why)*
- But the same event and its participants can have different syntactic realizations.

Sandy broke the glass. vs. The glass was broken by Sandy.  
She gave the boy a book. vs. She gave a book to the boy.

- Instead of focusing on syntax, consider the **semantic roles** (also called **thematic roles**) defined by each event.

## Stanford Dependencies

- Mary loaded the truck with hay.
- Hay was loaded onto the truck by Mary.

**Syntax is not enough!**

cf Mary ate the sandwich with Kim!

Add another family member

Relationship Status:

Interested in:

Looking for:

Networking

Political Views:

Religious Views:

- syntax  $\neq$  semantics
- The **semantic roles** played by different participants in the sentence are not trivially inferable from syntactical relations
- . . . though there are patterns!
- The idea of semantic roles can be combined with other aspects of meaning (beyond this course).

## Commonly used thematic roles

Role	Example
Agent	<i>The boy</i> kicked his toy
Theme	The boy kicked <i>his toy</i>
Experiencer	<i>The boy</i> felt sad
Result	The girl built <i>a shelf</i> with power tools
Instrument	The girl built a shelf <i>with power tools</i>
Source	She came <i>from home</i>
...	...

- J&M give definitions and additional roles

## Issues with thematic roles

- No universally agreed-upon set of roles
- Items with the “same” role (e.g., Instrument) may not behave quite the same
  - Sandy opened the door with a key    The key opened the door
  - Sandy ate the salad with a fork        \*The fork ate the salad
- The two main NLP resources for thematic roles avoid these problems by defining very fine-grained roles:
  - Specific to individual verbs only (PropBank)
  - Specific to small groups of verbs (FrameNet)

## Semantic role labelling

- The NLP task of identifying which words/phrases play which roles in an event.
- Supervised classification:
  - Resource data is **PropBank**: Repository of **frame files** for each verb (more shortly) plus annotations on constituents in Penn treebank with their semantic roles (wrt the relevant frame file).
  - Features are mostly related to syntactic structure and the particular words involved (i.e., assumes pipeline architecture)
- Current research focuses on reducing the need for training data (e.g., to work on non-English languages)

## PropBank

Penn treebank annotated with Arg0, Arg1 etc, and verb with its sense; so specific semantic role recoverable.

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

...



## Example Frame Roles: *load*

Mary loaded the truck with hay at the depot on Friday

- *load*: load.01 'cause to be burdened'
  - Arg0-PAG: loader,agent
  - Arg1-GOL: beast of burden
  - Arg2-PPT: cargo
  - Arg3-MNR: instrument
- *load\_up*: load.02 'phrasal cause to be burdened'  
Frame roles are the same as load.01
- *load*: load.03 'fix, set up to cheat'
  - Arg0-PAG: cheater
  - Arg1-GOL: thing loaded (dice, the deck, etc)
  - Arg2-PPT: with what
- All sentences can have temporal, spatial adjuncts (AM-TMP, AM-LOC). . .

## Semantic Role Labelling

Traditional pipeline:

1. Either assume or compute syntactic parse and predicate senses
2. **Argument identification** (deterministic): select the predicate's argument phrases (by parsing the parse tree)
3. **Argument classification**: select a role for each argument (wrt to the frame role for the predicate's sense).
  - Useful feature: predicate-to-argument path in the tree (e.g., NP-S-VP-V).

## Problems

- Numbered roles are predicate specific:
  - load.01.ARG1: beast of burden
  - put.01.ARG1: thing.put
  - put.01.ARG2: beast of burden.
- FrameNet tries to generalise via verb classes; but less treebank data.

## Paraphrase

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

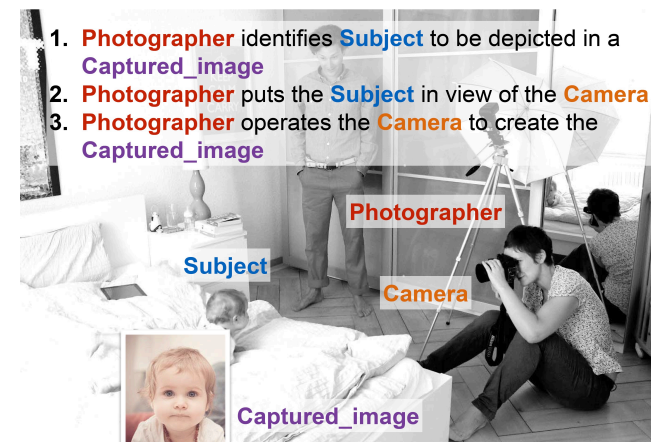
## Paraphrase

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



## Photo Labelled. . .

photograph vs. take picture vs. snap picture. . .



## Idealised Stanford Dependencies

- James snapped a photo of me with Sheila  
*nsubj*(snap, James)  
*doj*(snap, photo)  
*prep\_of*(photo,me)  
*prep\_with*(me,Sheila)  
*det*(photo,a)
- Sheila and I had our picture taken with James  
*nsubjpass*(taken,Sheila)  
*nsubjpass*(taken,I)  
*conj\_and*(Sheila,I)  
*aux*(taken,had)  
*doj*(taken,picture)  
*poss*(picture,our)  
*prep\_with*(picture,James)

Here, *agent* is the complement introduced with *by* in a passive construction. . .

## Example: Create\_physical\_artwork

### Definition:

A **Creator** creates an artefact 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**.  
I took his picture and told him it came out well.

### Frame Elements:

**Core:** creator, representation

**Non-Core** manner, location\_of\_representation . . .

## FrameNet: Meanings are reletavised to scenes!

- Tries to capture relationships among word and phrase meanings by assigning them the same frame (and so captures paraphrases).
- $\approx$  1000 frames represent scenarios.
  - Most are associated with lexical units (predicates); but some are phrases
- Frames are explained with textual descriptions and linguistic examples.

## FrameNet Resources

- FrameNets for several languages
- Some (limited!) data annotated with Frame elements from FrameNet
- SEMAFOR is a frame-semantic parser
  - Ongoing research at CMU, Google, Edinburgh. . .

## Summary

- Grammatical relations on their own don't determine who did what to whom
- You need to (also) know about word and phrase meanings and how they relate to grammatical roles
- There is flexibility in how a verb realises its participants syntactically (connected with the kind of event that the verb denotes)
- One must exploit those patterns to obtain NL understanding (e.g., predict entailments, paraphrases etc).