

# Unsupervised Induction of Frame-Semantic Representations

Ashutosh Modi  
amodi@mmci.uni-saarland.de

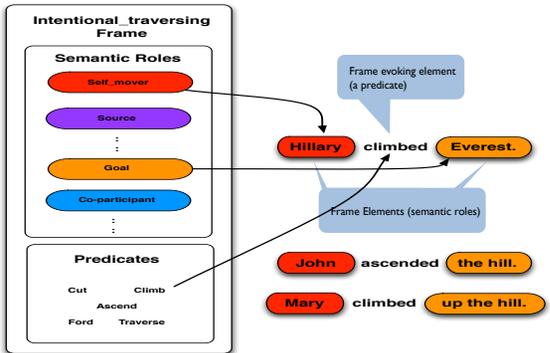
Ivan Titov  
titov@mmci.uni-saarland.de

Alexandre Klementiev  
aklement@mmci.uni-saarland.de



## Task Definition

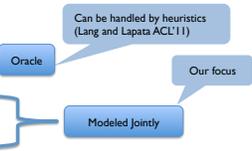
In frame semantics, a semantic frame is a conceptual structure describing a situation, object, or event along with associated properties and participants.



In this work we focus on inducing frames and their roles from unlabeled data.

For a collection of sentences:

1. Identify predicates
2. Identify arguments
3. Label predicates with frames
4. Label arguments with roles



### Frame Induction

- Cluster predicates (each cluster is a frame)

#### Key Signals

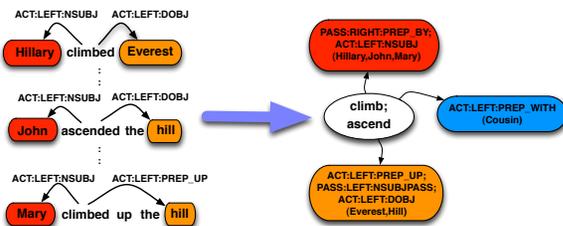
- Related predicates have:
    - Similar argument fillers
    - Similar mapping between syntax and semantics
- Related to Levin classes

### Role Induction

- Associate arguments with syntactic signatures (**argkeys**)
- Cluster argkeys (each argkey cluster is a role)

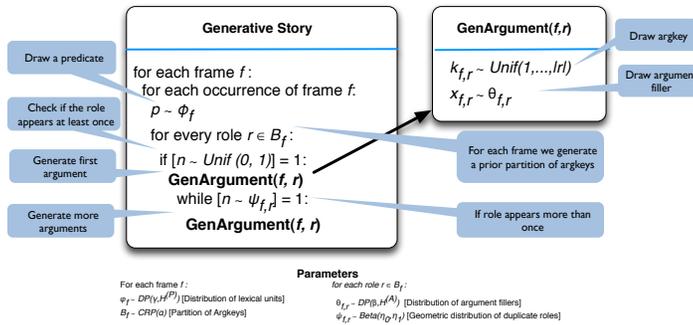
#### Key Signals

- Argkeys with similar argument fillers clustered together
- Most roles occur once (per predicate occurrence)



## The Model

Our unsupervised model simultaneously induces frames (clusters of predicates) and roles (clusters of argkeys) by exploiting distributional context.



- Related to Titov and Klementiev, ACL'11, EACL'12
- Can be extended to share alternation patterns across frames (as in EACL'12)
- Can be extended to induce cross-cutting clusters of argument fillers and multi-word expressions (as in ACL'11)

### Assumptions

- We consider verbal predicates only
- Every verb belongs to a single frame, i.e. we do not model polysemy
- Priors encode sparsity of selectional preferences and predicates over frames

### Approximate MAP inference

- Initialize with one frame per predicate
- Iteratively merge frames (greedy search)
- Each frame merge involves greedy role alignment

Pruned with a coarse cosine heuristic.

## Evaluation

Evaluation done on the **FrameNet corpus**: 158,048 sentences with 3,474 unique verbal predicates and 722 gold frames.

### Qualitative Evaluation

Induced Frames	FrameNet frames corresponding to the verbs
(rush::dash::tiptoe)	rush : [Self motion][150] [Fluidic motion][19] dash : [Self motion][100] tiptoe : [Self motion][114]
(ratify::sign::accede)	ratify : [Ratification][41] sign : [Sign agreement][81] [Hiring][18] [Text Creation][1] accede : [Sign Agreement][31]
(crane::lean::bustle)	crane : [Body movement][26] lean : [Change posture][70] [Placing][22] [Posture][12] bustle : [Self motion][55]
(cool::heat::warm)	cool : [Cause temperature change][27] heat : [Cause temperature change][52] warm : [Cause temperature change][41] [Inchoative change of temperature][16]
(encourage::intimidate::confuse)	encourage : [Stimulus focus][49] intimidate : [Stimulus focus][26] confuse : [Stimulus focus][45]
(plait::braid::dye)	plait : [Hair configuration][11] [Grooming][12] braid : [Hair configuration][7] [Clothing parts][6] [Rope manipulation][4] dye : [Processing materials][18]
(sell::purchase)	sell : [Commerce sell][107] purchase : [Commerce buy][93]
(forestall::shush)	forestall : [Thwarting][12] shush : [Silencing][6]

### Quantitative Evaluation

#### METRICS

**Purity (PU)** : Extent to which predicted cluster occurrences share the same gold label  
**Collocation (CO)** : Extent to which gold label is assigned to a single cluster  
**F1** : Harmonic mean of PU and CO

#### Role labeling

	PU	CO	F1
Our Approach	78.9	71.0	74.8
No Frame Induction	79.2	70.7	74.7
Syntactic Baseline	69.9	73.3	71.6

#### Frame labeling

	PU	CO	F1
Our Approach	77.9	31.4	44.7
No Clustering	80.8	29.0	42.7