Automatic frame-semantic role labeling

Swabha Swayamdipta

Carnegie Mellon University / University of Washington

Outline

1. Task of frame-SRL

2. Primary Subtasks

    a. Target Identification

    b. Frame Identification

    c. Frame-Element Identification

3. Advanced Modeling

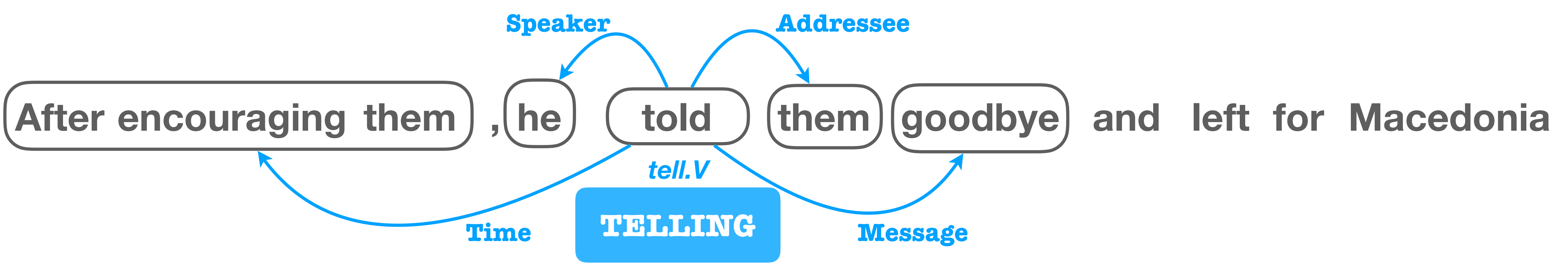
4. Looking forward: Multilingual Extensions

Frame-Semantic Role Labeling (frame-SRL)

After encouraging them , he told them goodbye and left for Macedonia .

Frame-Semantic Role Labeling (frame-SRL)

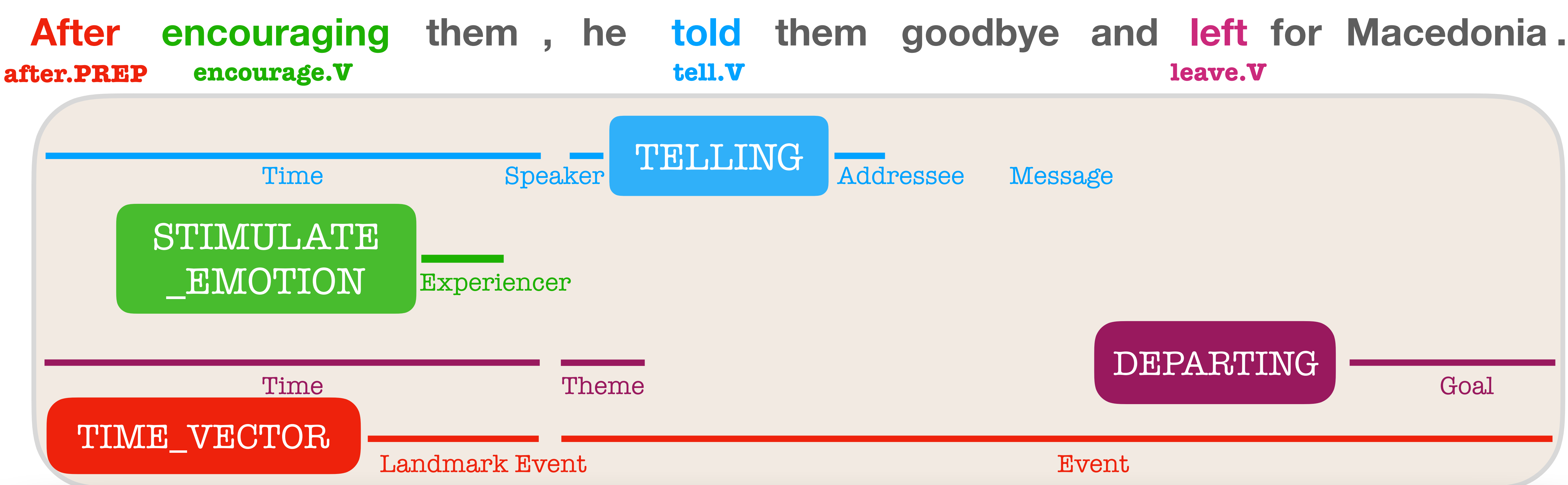
Sentence ⟶ Graph



- **Nodes**: tokens / spans in the sentence. Could represent both targets and arguments.
- **Node Labels**: lexical units (LUs) and frames.
- **Edges**: Between target nodes and argument nodes
- **Edge Labels**: roles of arguments / frame-elements

Frame-Semantic Graphs: Overlapping Nodes

Frame-Semantic Graphs: Overlapping Nodes



Frame-SRL: Subtasks

Frame-SRL: Subtasks

1. **Target Identification**

After encouraging them , he told them goodbye and left for Macedonia .

after.PREP encourage.V tell.V leave.V

Frame-SRL: Subtasks

1. **Target Identification**
2. **Frame Identification**

After encouraging them , he told them goodbye and left for Macedonia .

after.PREP encourage.V tell.V leave.V

TELLING

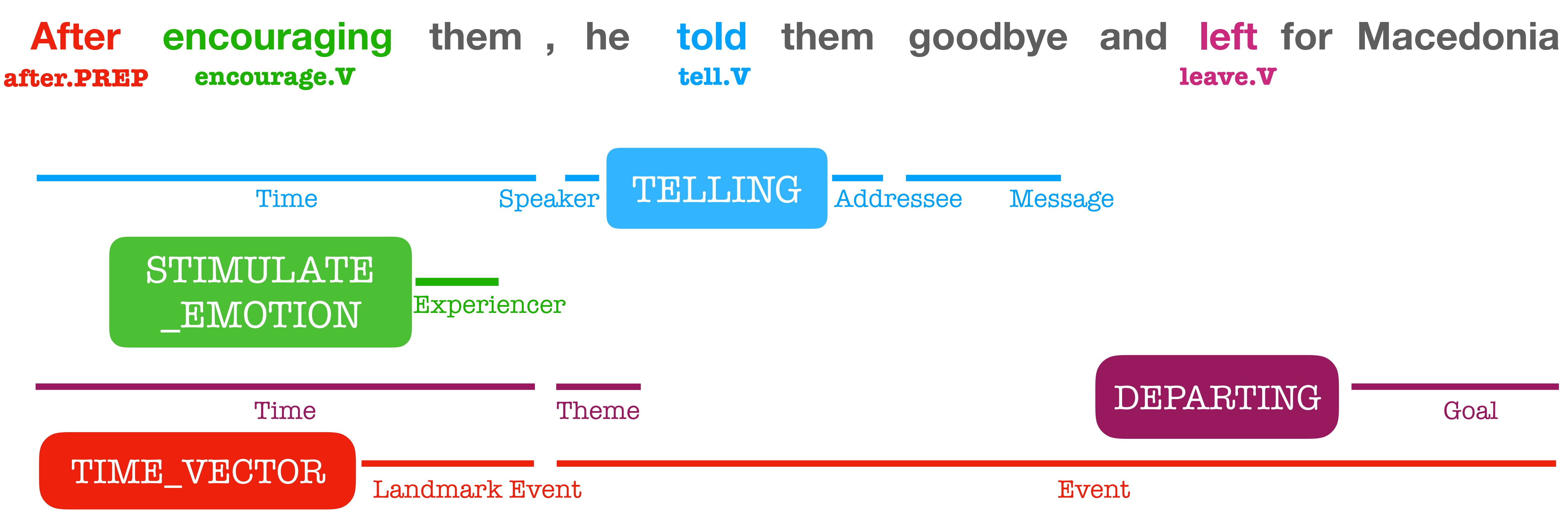
STIMULATE_EMOTION

DEPARTING

TIME_VECTOR

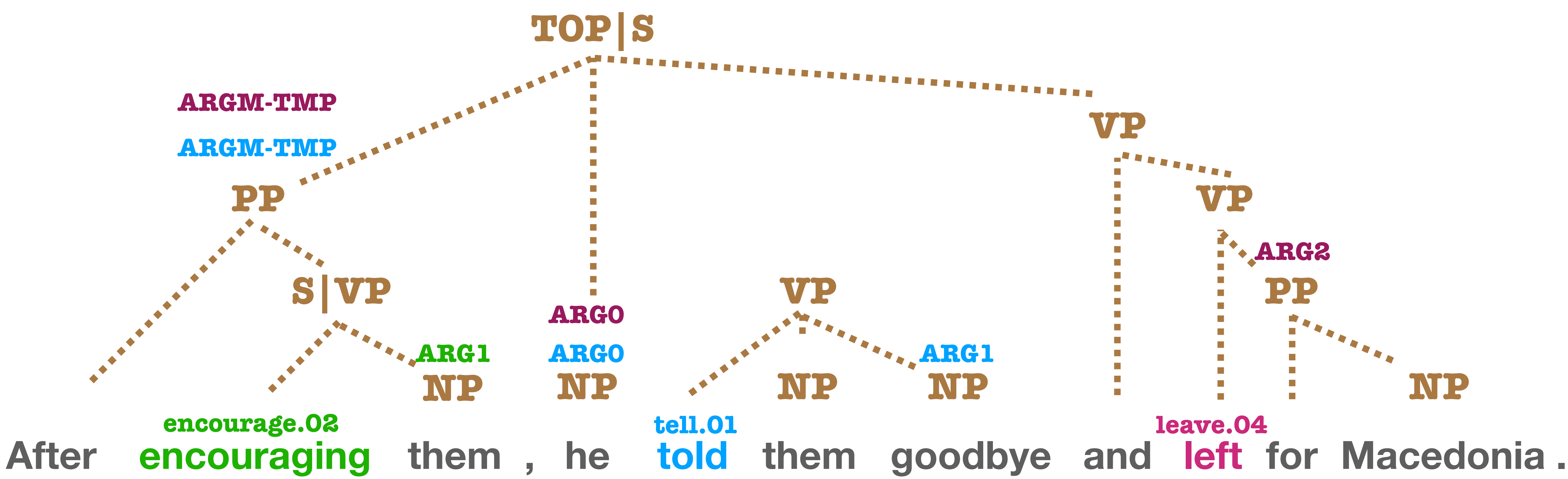
Frame-SRL: Subtasks

1. **Target Identification**
2. **Frame Identification**
3. **Frame-Elements Identification**



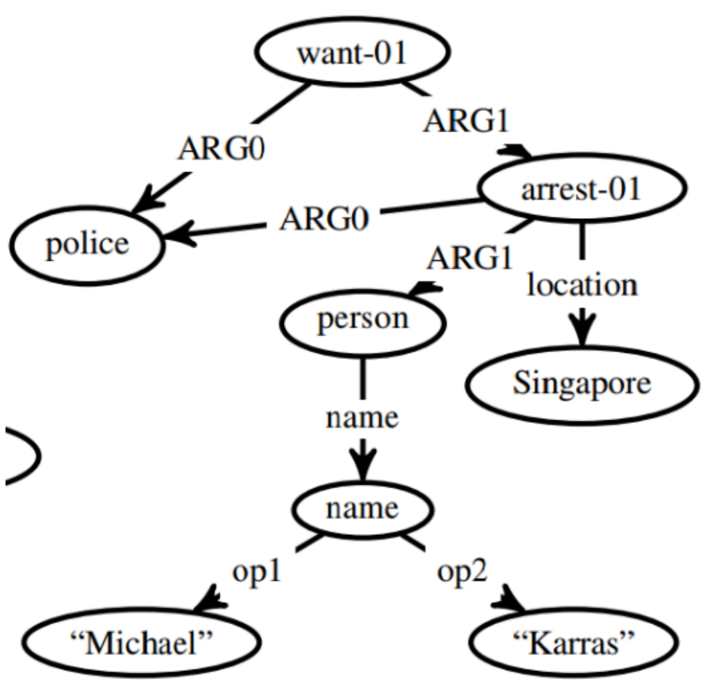
A close relative: PropBank SRL

1. ~~Target~~ Predication Identification
2. ~~Frame~~ Sense Identification
3. ~~Frame-Elements~~ Argument Identification



Related tasks

Abstract Meaning Representation



Banarescu et. al. (2013)

QA -SRL

A much larger super eruption in Colorado **produced** over 5,000 cubic kilometers of material.

Produced	What produced something?	A much larger super eruption
	Where did something produce something?	in Colorado
	What did something produce?	over 5,000 cubic kilometers of material

He et. al. (2015)
Fitzgerald et. al. (2018)

Semantic proto-roles

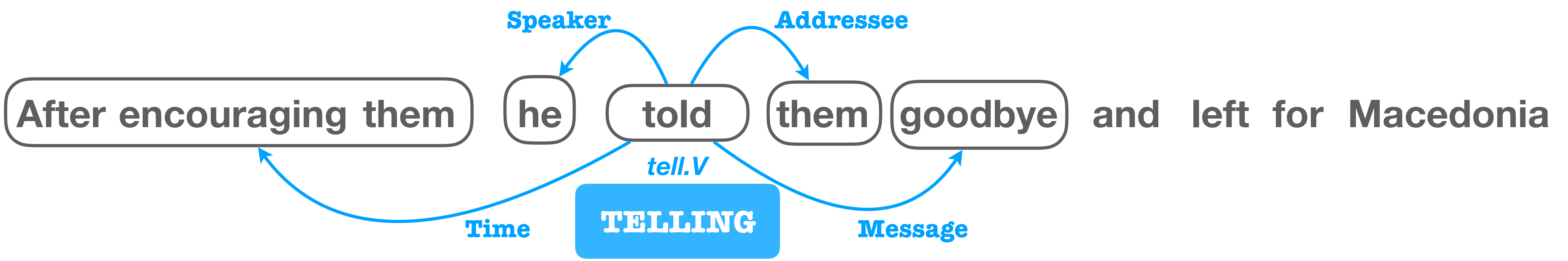
Sentences	**Property**	(A)	(B)	(C)
(A) **She** was untrained and, in one botched job *killed* a client.	instigated	5	5	5
(B) **The antibody** then *kills* the cell.	volitional	2	1	5
(C) **An assassin in Colombia** *killed* a federal judge on a Medellin street.	awareness	3	1	5
PropBank KILL.01, ARG_0-PAG: killer	sentient	5	1	5
VerbNet MURDER-42.1-1, AGENT: ACTOR in an event who initiates and carries out the event intentionally or consciously, and who exists independently of the event	moved	3	3	3
	phys_existed	5	5	5
	created	1	1	1
	destroyed	1	3	1
FrameNet KILLING, KILLER/CAUSE: (The person or sentient entity) / (An inanimate entity or process) that causes the death of the VICTIM.	changed_poss	1	1	1
	changed_state	3	3	3
	stationary	3	3	3

Reisinger et. al. (2015)

A little bit of history

- Pioneered by Gildea & Jurafsky (CL 2002) on an earlier version of FrameNet version 1.0.
- Development of PropBank (Kingsbury & Palmer, 2002; Palmer et. al., 2005)
- CoNLL shared tasks expedited the development on PropBank-style SRL.
    - Shared tasks in 2005, 2008, 2009, 2012
- SemEval 2009 Shared Task 19 (Baker, Ellsworth & Erk, 2007) sparked interest in automatic frame-SRL.

Frame-SRL data



- Full-text annotations (78 documents in FN 1.5)
    - Train (47) / Dev (8) / Test (23)
- Exemplars
- Mapping between LUs and frames
- Mapping between frames and frame-elements
- (Multi) Inheritance between frames
- Phrase Types / Grammatical Functions

Model architectures

Model architectures

- Most common approach: Supervised learning

Model architectures

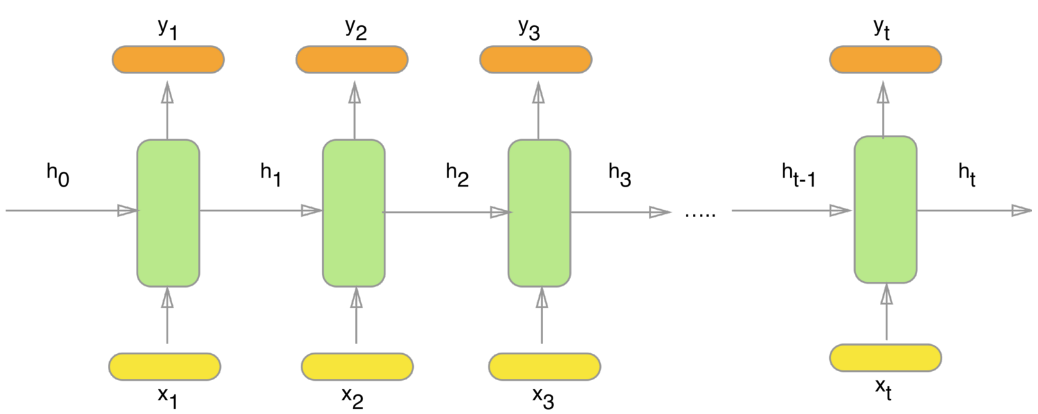
- Most common approach: Supervised learning
- Linear models - most models prior to 2015
    - May use distributional representations

Model architectures

- Most common approach: Supervised learning
- Linear models - most models prior to 2015
    - May use distributional representations
- Non-linear / neural models

Model architectures

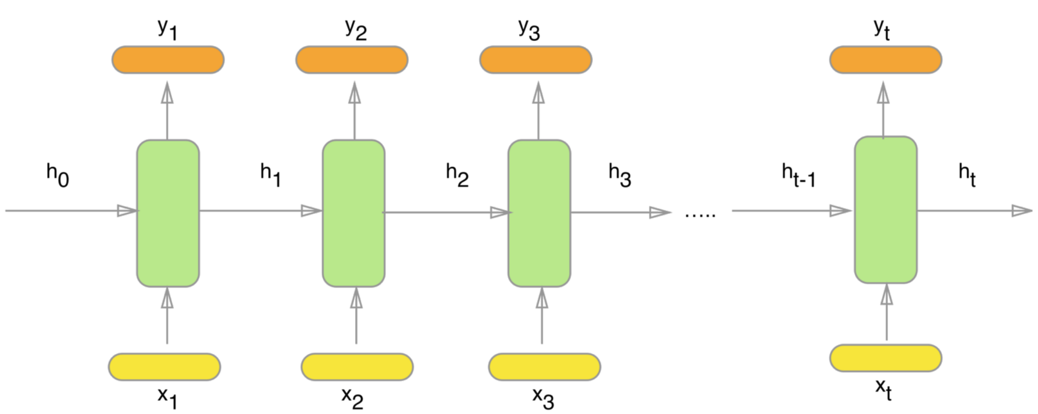
- Most common approach: Supervised learning
- Linear models - most models prior to 2015
    - May use distributional representations
- Non-linear / neural models



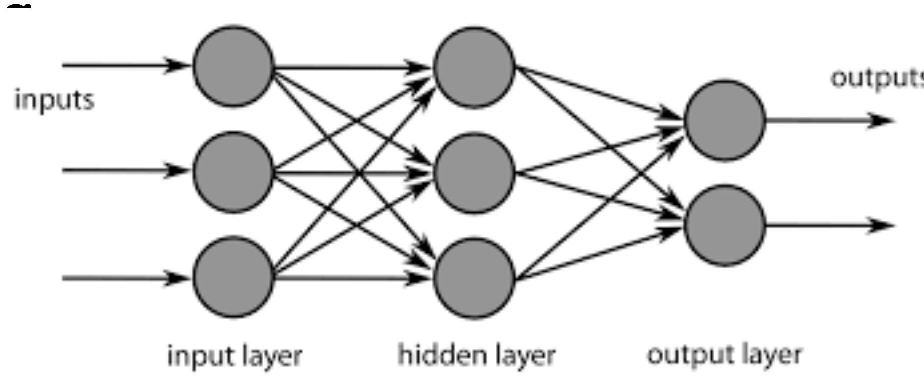
Recurrent Neural Nets

Model architectures

- Most common approach: Supervised learning
- Linear models - most models prior to 2015
    - May use distributional representations
- Non-linear / neural models



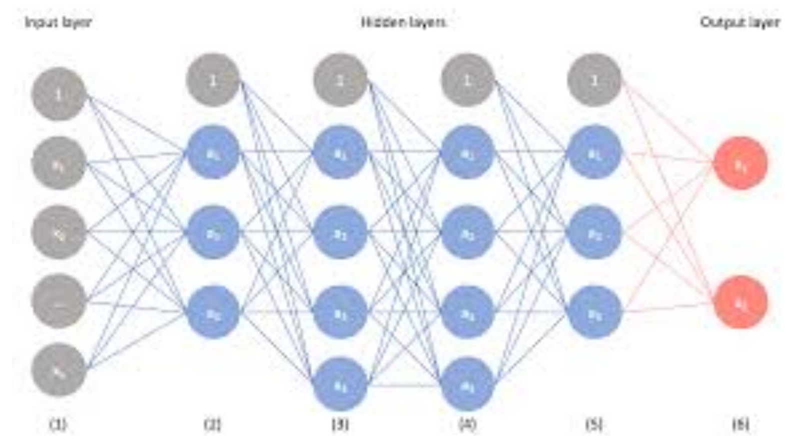
Recurrent Neural Net



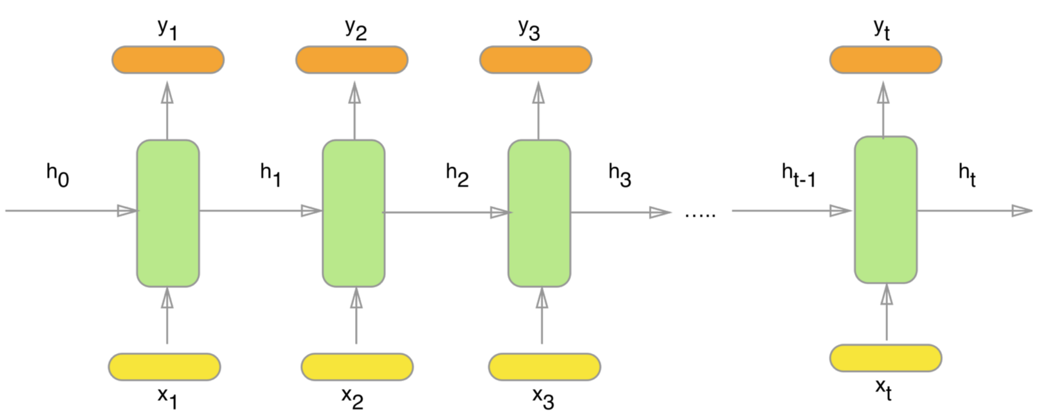
Feed-forward Nets

Model architectures

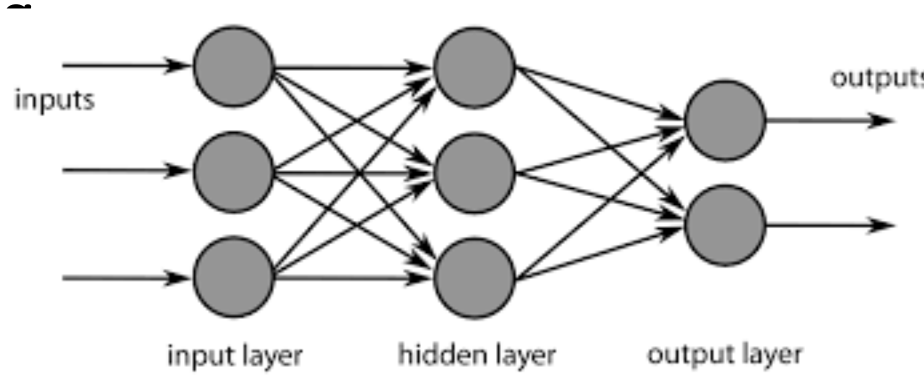
- Most common approach: Supervised learning
- Linear models - most models prior to 2015
    - May use distributional representations
- Non-linear / neural models



Convolutional Neural Nets

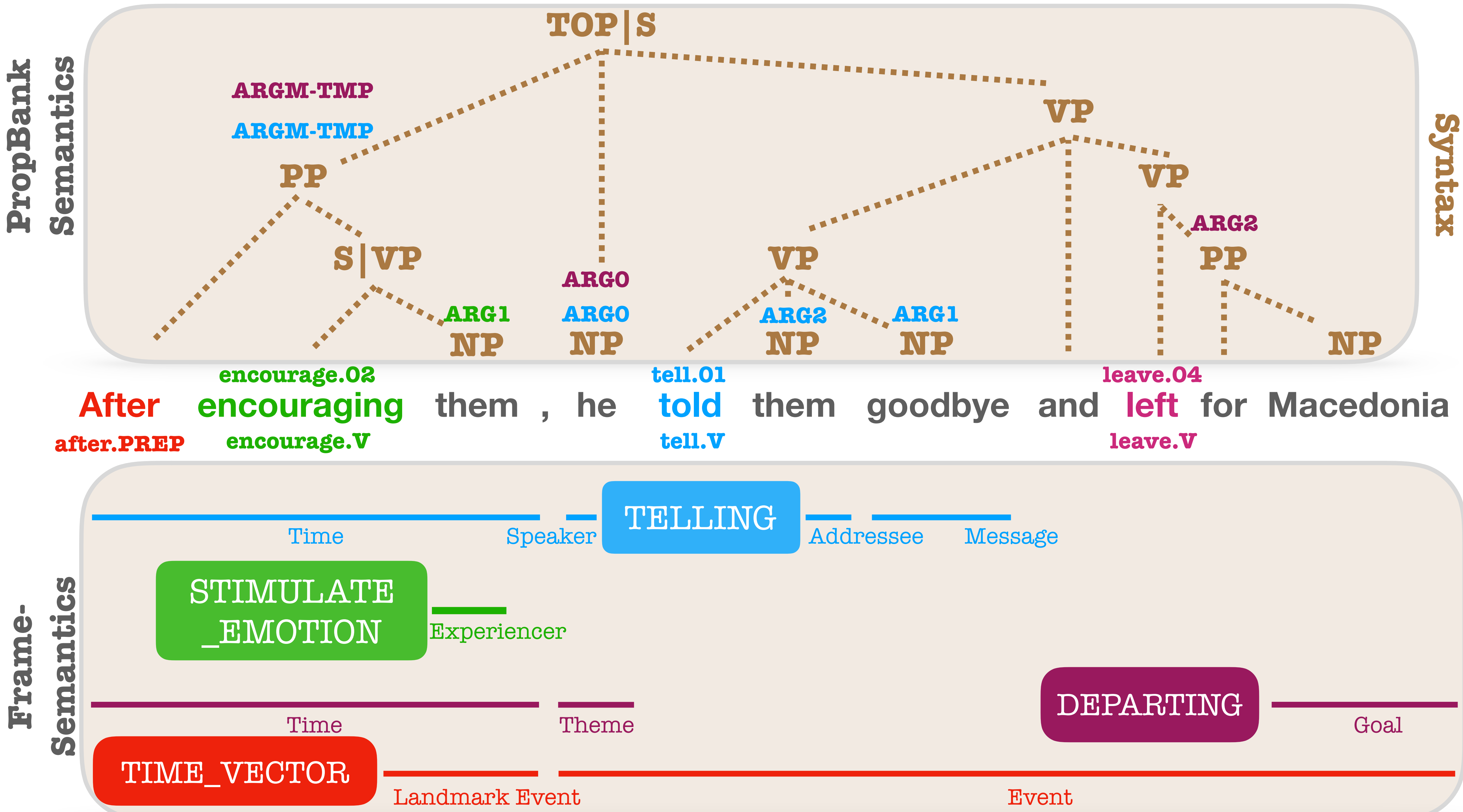


Recurrent Neural Net



Feed-forward Nets

Role of syntax



Why automatic frame-SRL?

- Information extraction (Surdeanu, et al., 2003)
- Textual entailment (Tatu & Moldovan, 2005; Burchardt & Frank, 2006)
- Text categorization (Moschitti, 2008)
- Question answering (Narayanan & Harabagiu, 2004; Frank, et. al., 2007; Moschitti, et. al., 2007; Shen & Lapata, 2007)
- Machine Translation (Wu & Fung, 2009a, 2009b, Marchegianni et. al., 2017) and its evaluation (Giménez & Màrquez, 2007)
- Text-to-scene generation (Coyne et. al., 2012)
- Dialog systems (Chen et. al., 2013)
- Social network extraction (Agrawal et. al., 2014)
- Knowledge Extraction from Twitter (Søgaard et. al., 2015)

Summary of Part 1

- Frame-SRL as a graph induction task
    - Subtasks
- Related tasks - Propbank SRL, QA-SRL etc.
- What's in the dataset?
- Supervised Learning: Shift from linear to non-linear models
- Syntax is key!

Outline

1. Task of frame-SRL

2. Primary Subtasks

   a. Target Identification

   b. Frame Identification

   c. Frame-Element Identification

3. Advanced Modeling

4. Looking forward: Multilingual Extensions

Target Identification

After	encouraging	them	,	he	told	them	goodbye	and	left	for	Macedonia	.
after.PREP	encourage.V				tell.V				leave.V			

- Predict "semantically salient" tokens as targets in the sentence.
- Also, identify the lexical units (LUs) = lemma + POS tag of targets
    - There might be ambiguity here! Example "**encourage.V**" vs "**encouraging.A**"
- Average in FN 1.5: 6 targets per sentence.

Target ID: challenges

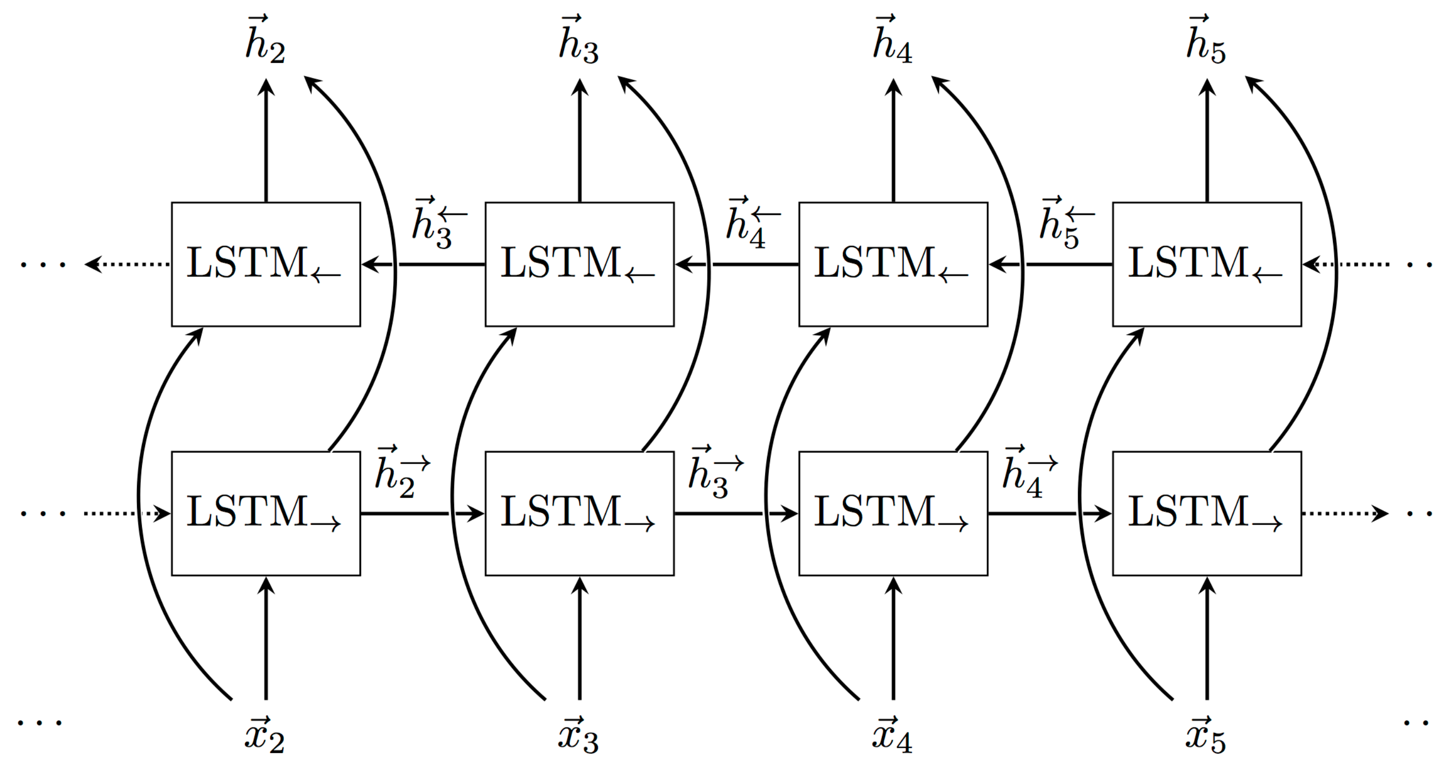
- Data sparsity, cannot use exemplar data.
- No simple POS tag based bijection, unlike in PropBank, where targets are almost always verbs.
- FrameNet: Verbs, nouns, adjectives and prepositions can be targets, BUT not always!
- Multi-word expressions also considered valid targets. About 4% of all targets in FN 1.5.
  - Span "**tell apart**" gets labeled with LU "**tell_apart.V**"
- Targets can be discontinuous
  - Span "**there would have been**" gets labeled as LU "**there_be.V**".

Target ID: model based on heuristics

- **have** was retained only if had an object,
- **be** was retained only if it was preceded by **there**,
- **will** was removed in its modal sense,
- **of course** and **in particular** were removed,
- the prepositions **above, against, at, below, beside, by, in, on, over,** and **under** were removed unless their head was marked as locative,
- **after** and **before** were removed unless their head was marked as temporal,
- **into, to,** and **through** were removed unless their head was marked as direction,
- **as, for, so,** and **with** were always removed,
- because the only sense of the word of was the frame PARTITIVE, it was removed unless it was preceded by **only, member, one, most, many, some, few, part, majority, minority, proportion, half, third, quarter, all**, or **none**, or it was followed by **all, group, them,** or **us**,
- all targets marked as support verbs for some other target were removed.

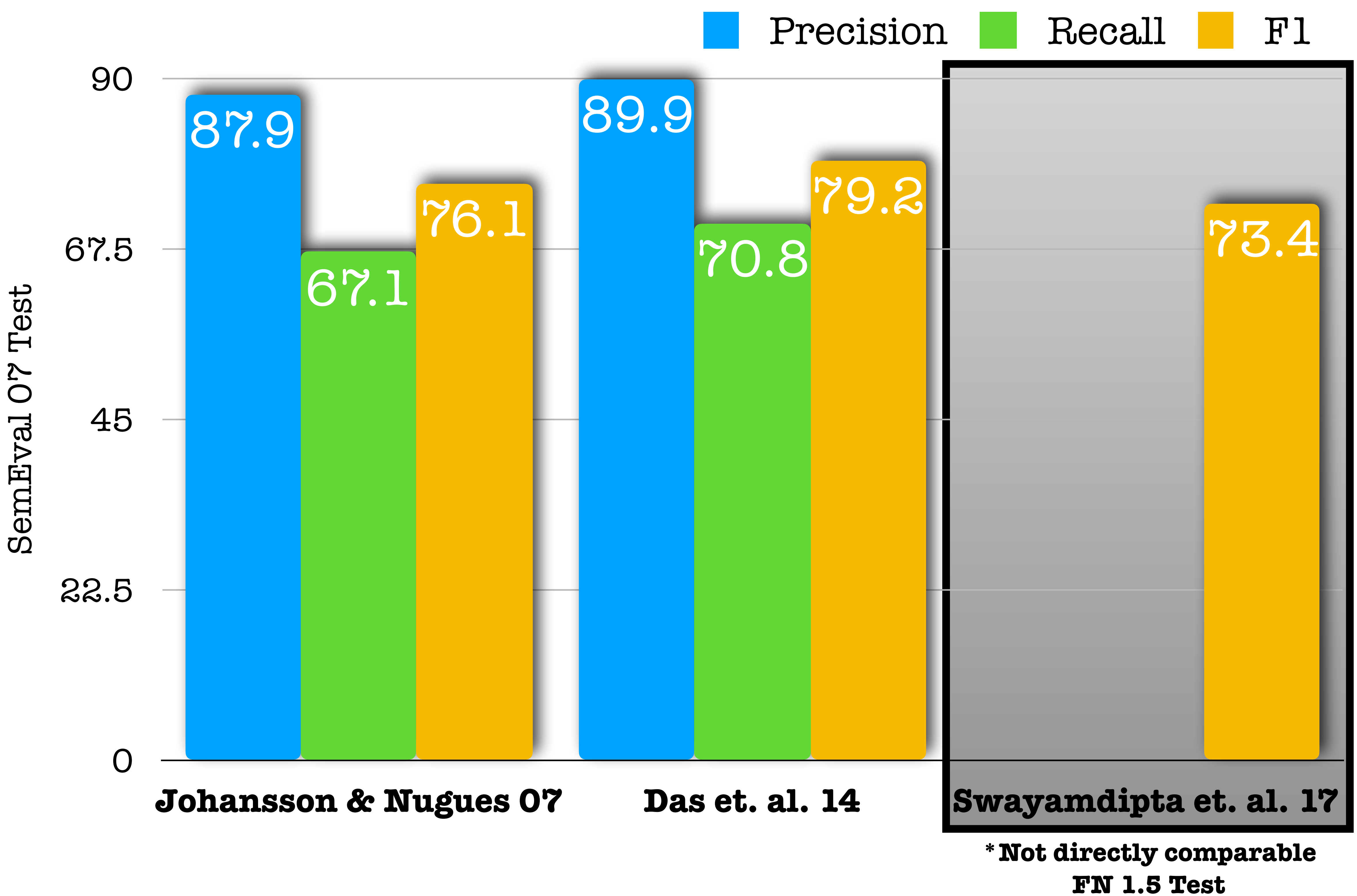
Johansson & Nugues (SemEval, 2007)
Das et. al. (CL 2014)

Target ID: Neural Model



- Bidirectional RNNs (Open-SESAME; Swayamdipta et. al., 2017)
- Does significantly worse than heuristics-based model.

Target ID: Evaluation



Outline

1. Task of frame-SRL

2. Primary Subtasks

   a. Target Identification

   b. Frame Identification

   c. Frame-Element Identification

3. Advanced Modeling

4. Looking forward: Multilingual Extensions

Frame Identification

After	encouraging	them	,	he	told	them	goodbye	and	left	for	Macedonia	.
after.PREP	encourage.V				tell.V				leave.V			
TIME_ VECTOR	STIMULATE _EMOTION				TELLING				DEPARTING			

- Given a target (lexical unit) token in the sentence, identify the frame evoked by it.
- On an average, about 2 frames per lexical unit.
- Lexical units play a critical role here, because of the mapping between lexical units and frames.
  - Errors in identifying lexical units / targets directly impact frame identification.

Frame ID models

- Simple Classification

$$\textbf{frame} = \arg \max_{\textbf{frame} \in \textbf{LU}} p(\textbf{frame} \mid \textbf{target}, \textbf{LU}, \textbf{sentence})$$

- When LU is ambiguous:
    - treat it as another unknown
    - Learn a distribution for it

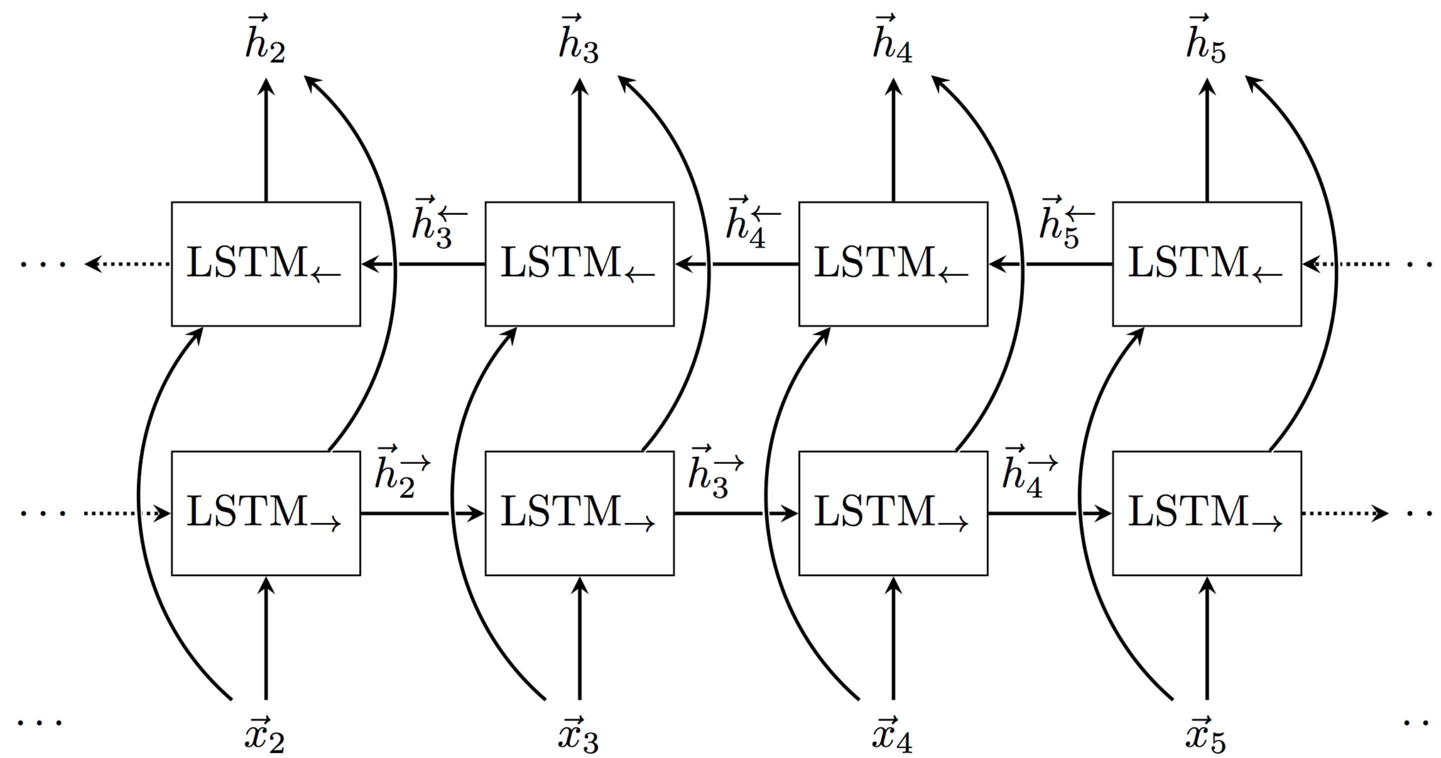
Linear Frame ID Models

- With features from syntax (Das et. al., CL 2014)

  - the POS of the parent of the head word of t_i
  - * the set of syntactic dependencies of the head word[21] of t_i
  - * if the head word of t_i is a verb, then the set of dependency labels of its children
  - the dependency label on the edge connecting the head of t_i and its parent
  - the sequence of words in the prototype, w_ℓ
  - the lemmatized sequence of words in the prototype
  - the lemmatized sequence of words in the prototype and their part-of-speech tags π_ℓ
  - WordNet relation[22] ρ holds between ℓ and t_i
  - WordNet relation[22] ρ holds between ℓ and t_i, and the prototype is ℓ
  - WordNet relation[22] ρ holds between ℓ and t_i, the POS tag sequence of ℓ is π_ℓ, and the POS tag sequence of t_i is π_t

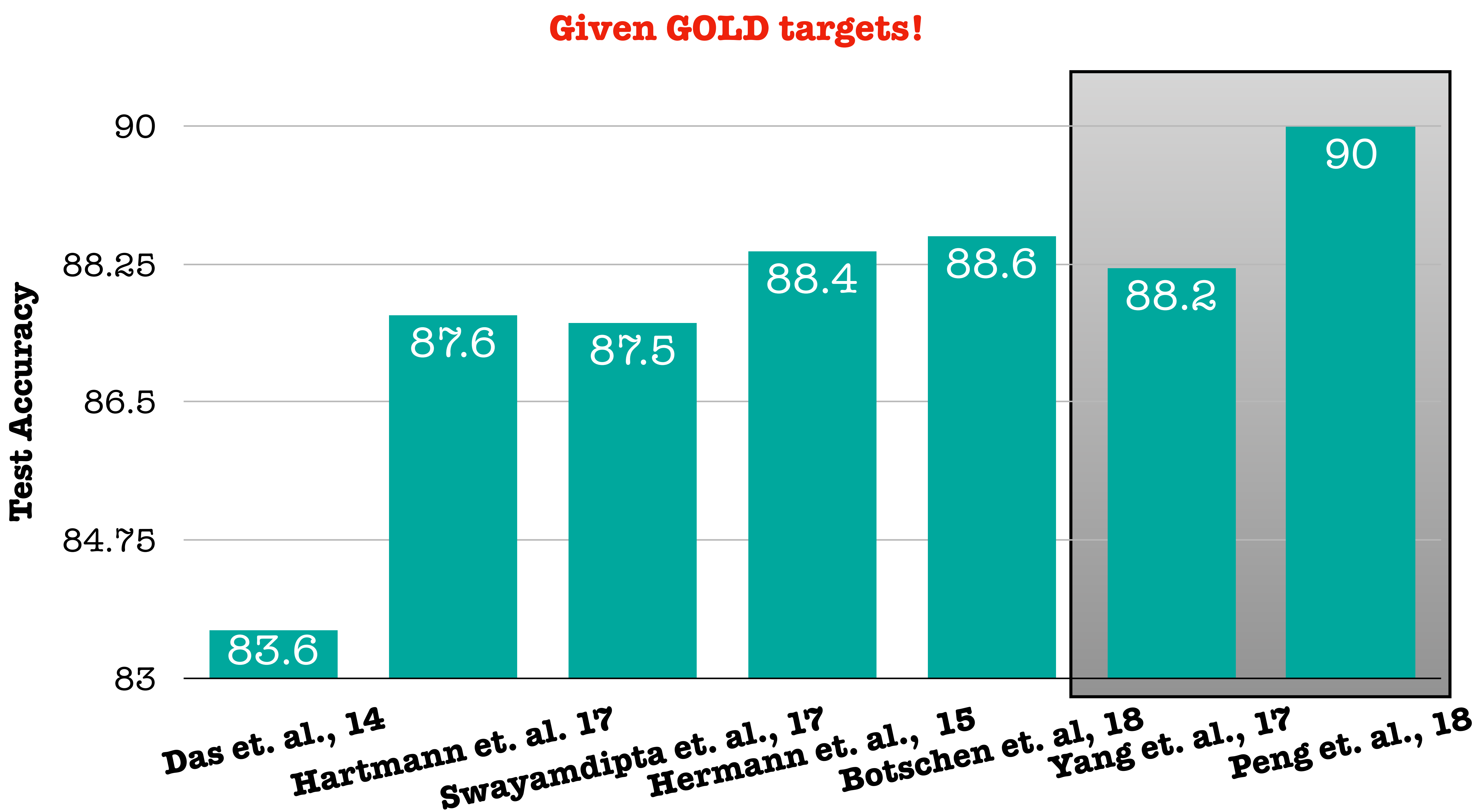
- With distributional semantics (Hermann et. al., ACL 2014)

Neural Models for Frame ID



- Bidirectional LSTM (Swayamdipta et. al., 2017)
- Feed-forward neural nets (Hermann et. al., 2015)

Frame ID: Evaluation



Outline

1. Task of frame-SRL

2. Primary Subtasks

    a. Target Identification

    b. Frame Identification

    c. Frame-Element Identification

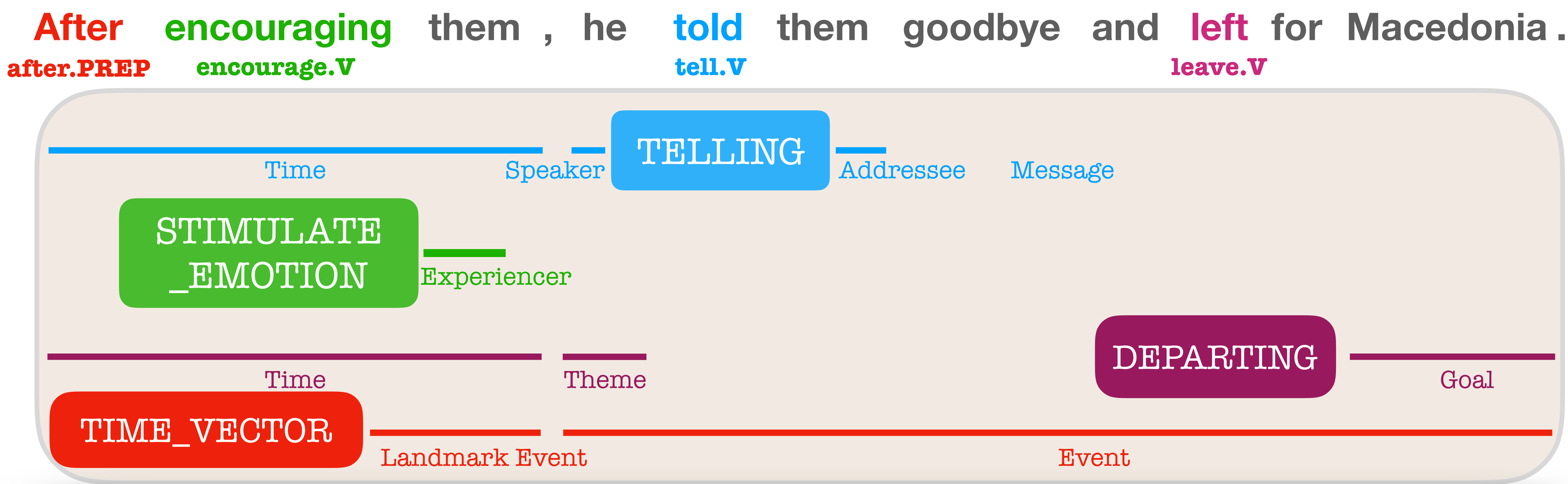
3. Advanced Modeling

4. Looking forward: Multilingual Extensions

Argument Identification

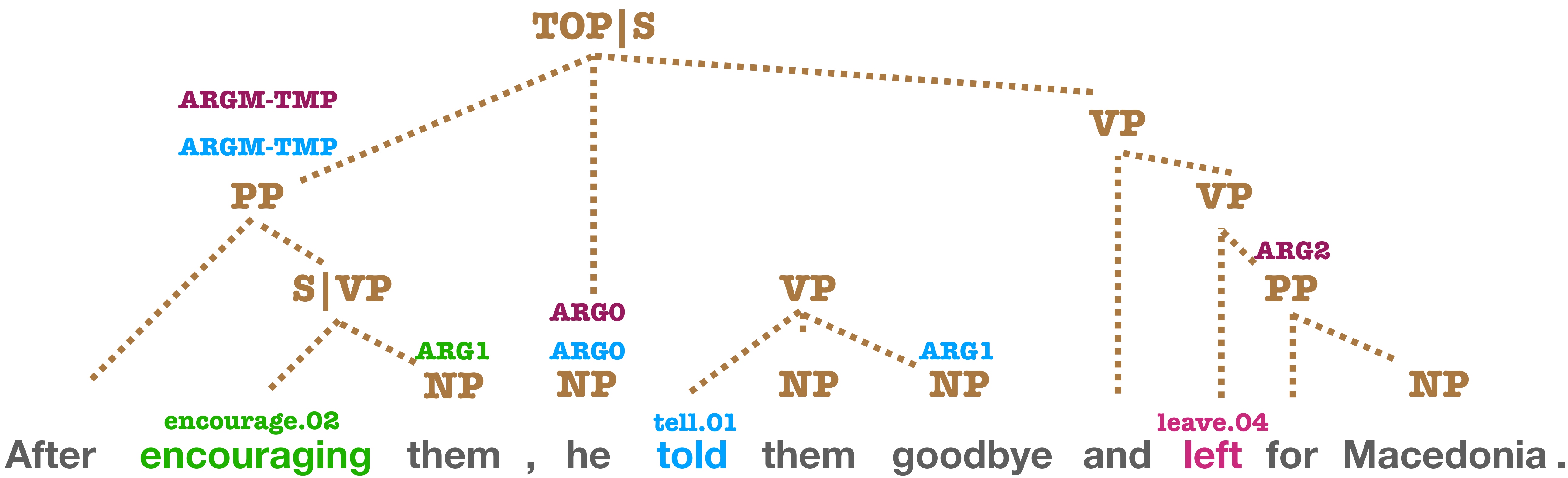
- Given a target and the frame it evokes, identify
    - all the spans in the sentence which are arguments to the frame,
    - and their respective labels (frame-elements)

Argument Identification



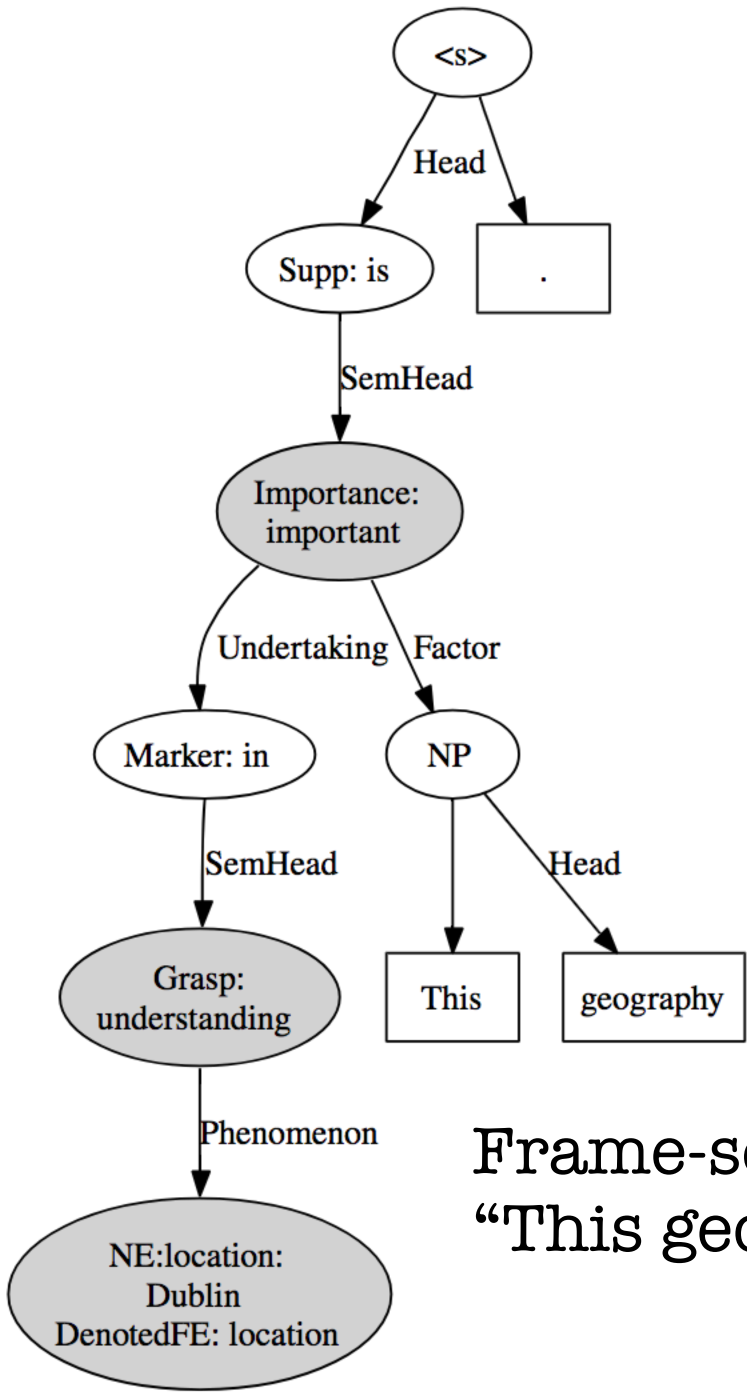
- Given a target and the frame it evokes, identify
  - all the spans in the sentence which are arguments to the frame,
  - and their respective labels (frame-elements)

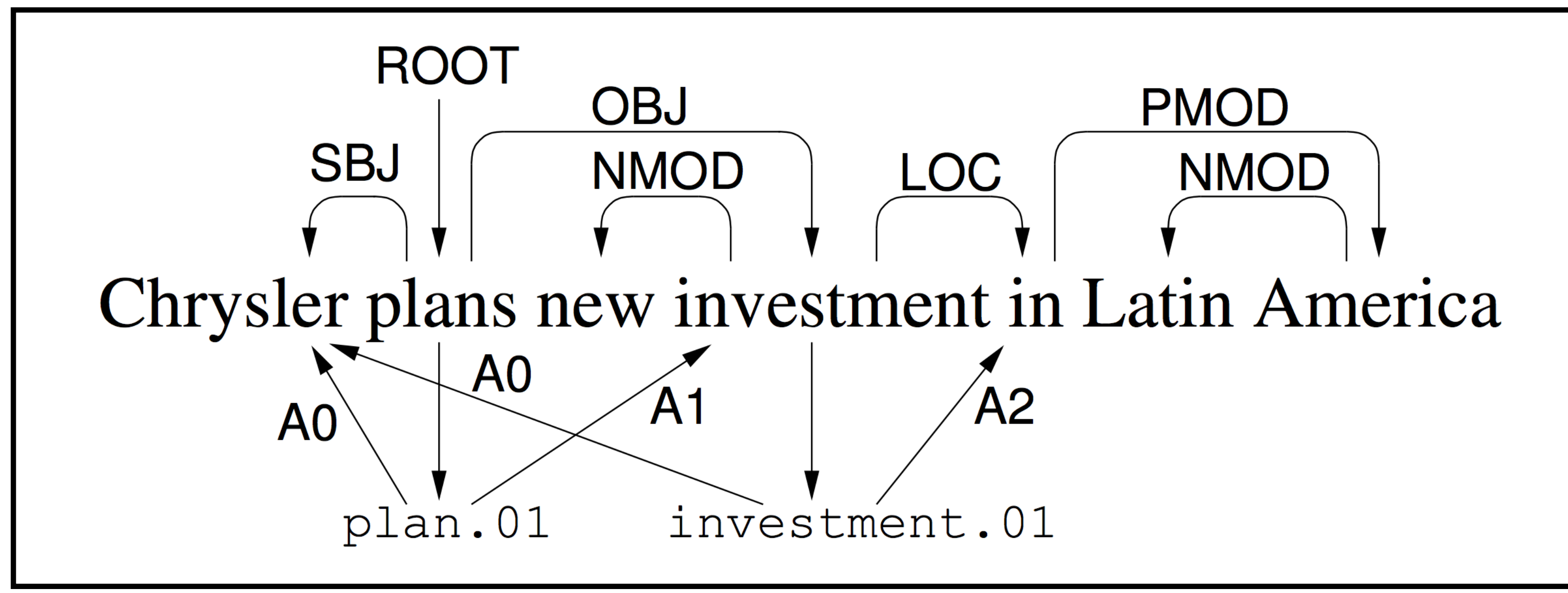
PropBank vs FrameNet arguments



- Primary difference between PropBank SRL and Frame SRL arg ID:
    - PropBank role labels (ARG_0 - ARG_n) are uniform across predicates.
    - ARG_0 and ARG_1 correspond to Dowty's (1991) proto-agent and proto-patient, respectively.
    - Higher ARG_n have verb-specific definitions.

Arg ID: Dependency Graph Variant





PropBank-style dependency graph for sentence, along with syntactic dependencies.

Johannson & Nugues (ACL, 2008)

Frame-semantic dependency graph for sentence "This geography is important for understanding Dublin.".

Baker et. al. (SemEval, 2007)

Arg ID: Basics

- Each predicate/ target (and its frame) considered independently
- Arguments as spans
- Arguments as sequences

Linear Arg ID models

$$\textbf{role} = \arg\max_{\textbf{role} \in \textbf{frame}} p(\textbf{role} \mid \textbf{frame}, \textbf{LU}, \textbf{target}, \textbf{span})$$

- Span classification task
- Candidate spans pruned by syntactic rules
- Features rely heavily on syntax

Features with both null and non-null variants: These features come in two flavors: if the argument is null, then one version fires; if it is overt (non-null), then another version fires.

- ● some word in t has lemma λ
- ◐ some word in t has lemma λ, and the sentence uses PASSIVE voice
- ◐ the head of t has subcategorization sequence $\tau = \langle \tau_1, \tau_2, \ldots \rangle$
- ● the head of t has c syntactic dependents
- ● some word in t has POS π
- ◐ some word in t has lemma λ, and the sentence uses ACTIVE voice
- ◐ some syntactic dependent of the head of t has dependency type τ
- ● bias feature (always fires)

Span content features: apply to overt argument candidates.

- ○ POS tag π occurs for some word in s
- ○ the first word of s has POS π
- ○ the last word of s has POS π
- ○ the head word of s has syntactic dependency type τ
- ● w_{s_2} and its closed-class POS tag π_{s_2}, provided that $|s| \geq 2$
- ○ the head word of s has lemma λ
- ○ the last word of s: $w_{s_{|s|}}$ has lemma λ
- ● $w_{s_{|s|}}$, and its closed-class POS tag $\pi_{s_{|s|}}$, provided that $|s| \geq 3$
- ◐ lemma λ is realized in some word in s, the voice denoted in the span (ACTIVE or PASSIVE)
- ○ the head word of s has POS π
- ● $|s|$, the number of words in the span
- ○ the first word of s has lemma λ
- ● the first word of s: w_{s_1}, and its POS tag π_{s_1}, if π_{s_1} is a closed-class POS
- ● the syntactic dependency type τ_{s_1} of the first word with respect to its head
- ● τ_{s_2}, provided that $|s| \geq 2$
- ● $\tau_{s_{|s|}}$, provided that $|s| \geq 3$
- ◐ lemma λ is realized in some word in s
- ◐ lemma λ is realized in some word in s, the voice denoted in the span, s's position with respect to t (BEFORE, AFTER, or OVERLAPPING)

Syntactic features: apply to overt argument candidates.

- ○ dependency path: sequence of labeled, directed edges from the head word of s to the head word of t
- ○ length of the dependency path

Span context POS features: for overt candidates, up to 6 of these features will be active.

- ○ a word with POS π occurs up to 3 words before the first word of s
- ○ a word with POS π occurs up to 3 words after the last word of s

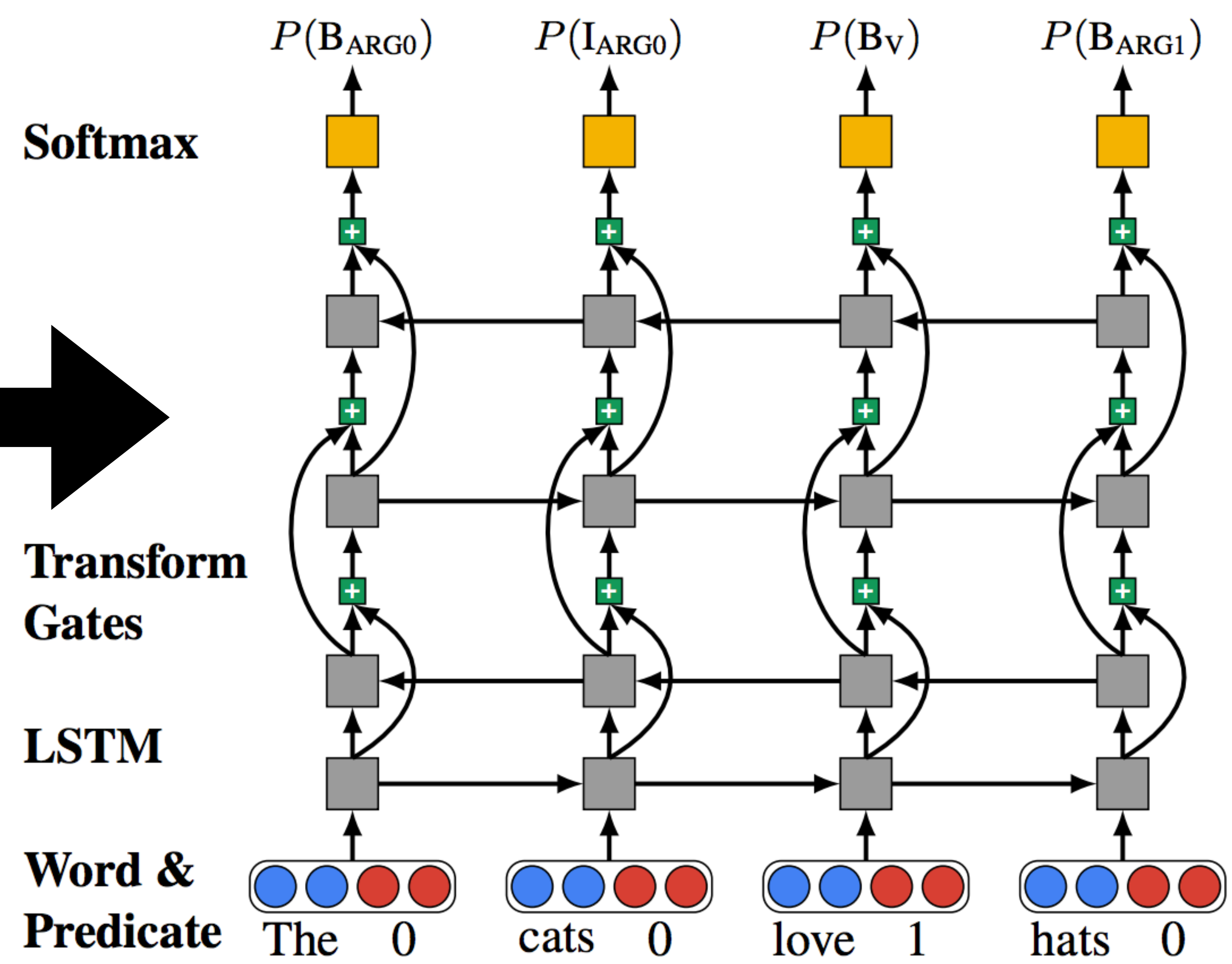
Ordering features: apply to overt argument candidates.

- ● the position of s with respect to the span of t: BEFORE, AFTER, or OVERLAPPING (i.e. there is at least one word shared by s and t)
- ○ linear word distance between the nearest word of s and the nearest word of t, provided s and t do not overlap
- ○ target-argument crossing: there is at least one word shared by s and t, at least one word in s that is not in t, and at least one word in t that is not in s
- ○ linear word distance between the middle word of s and the middle word of t, provided s and t do not overlap

Das et. al. (CL 2014)

Neural models for Arg ID

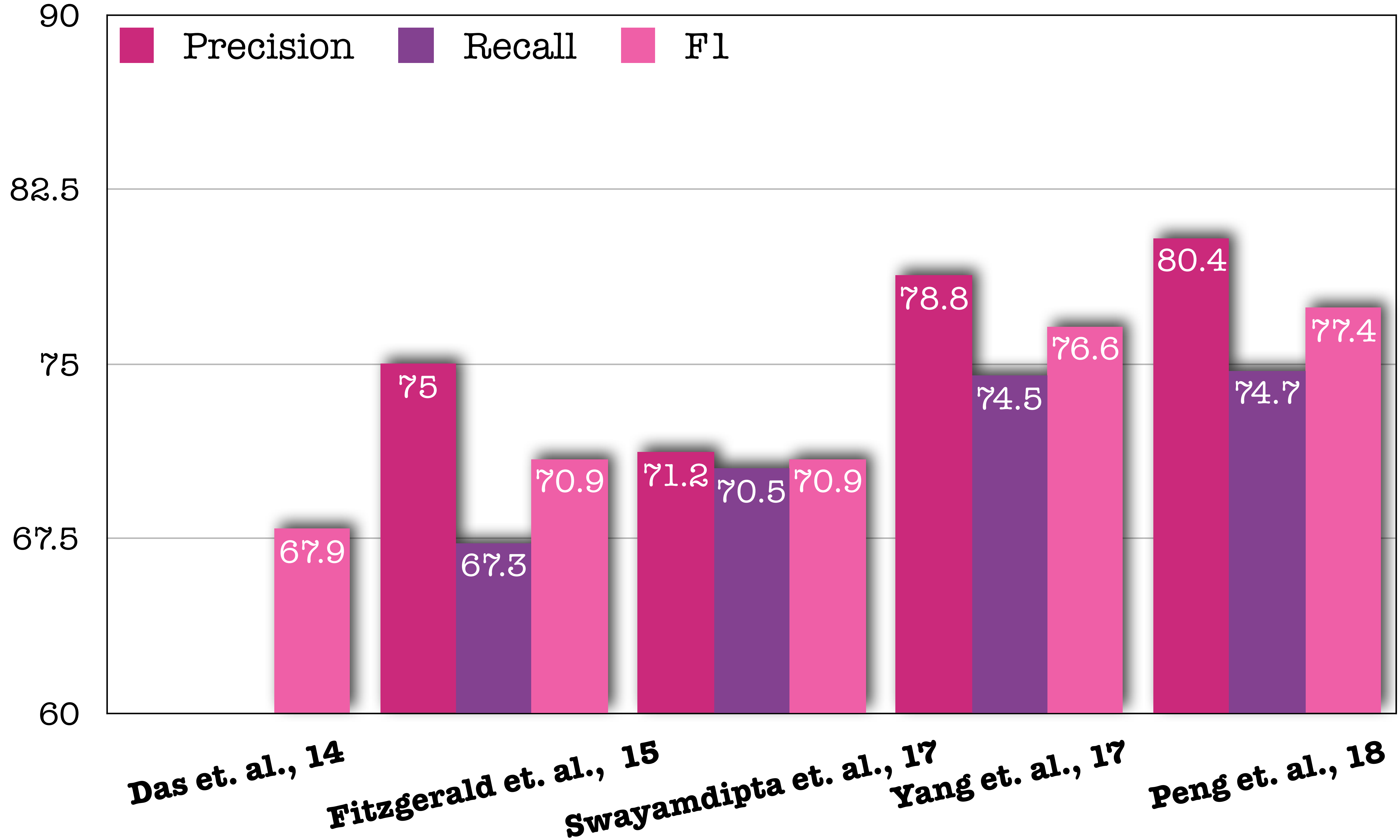
- BIO specification
- Deep bidirectional, highway (Zhou & Xu,2015; He et. al. ACL 2017)
- Transformers (Tan et. al., AAAI 2018; Strubell et. al., 2018)



He et. al. (2017)

Frame + Arg ID: Evaluation

Given GOLD targets!



End-to-end
Frame SRL evaluation

Summary of Part 2

- Subtasks have their own intricacies.
- Automation is coming along fast, we have seen big gains.
- Non-uniformity of evaluation is an issue.
- Can we do better than individual tasks? [Part 3]

Outline

1. Task of frame-SRL

2. Primary Subtasks

    a. Linear

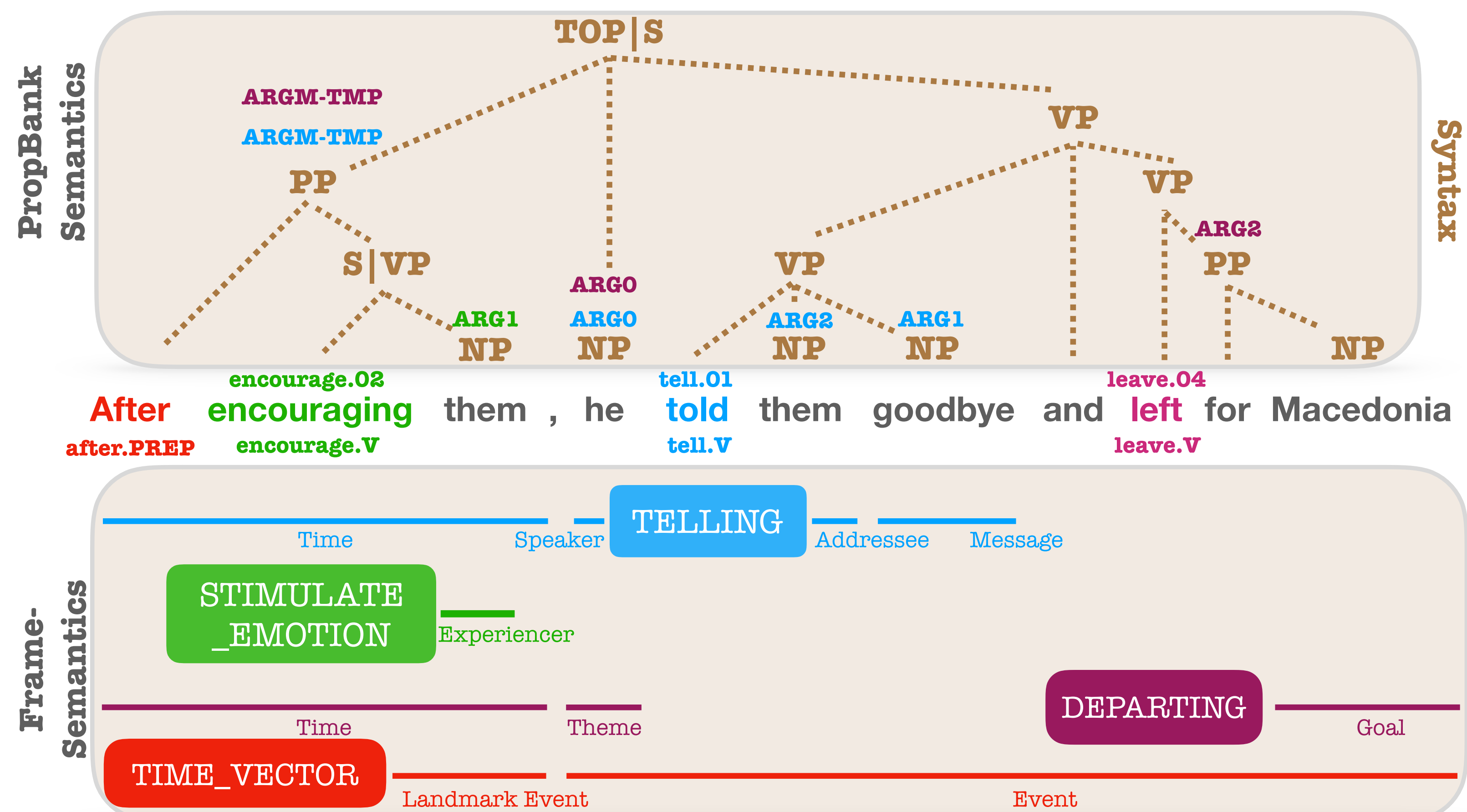
    b. Neural models

3. Advanced Modeling

4. Looking forward: Multilingual Extensions

Role of syntax (revisited)

- Traditional feature design (Das et. al. 2014)
- Heuristics for potential argument identification (Das et. al., 2014)
- Continuous valued features (Roth & Lapata, 2016; Yang et. al., 2017)
- Constraints during decoding (He et. al., 2017)



Learning strategies

- Data likelihood
- Encourage recall of spans. (Cost-augmented objectives Swayamdipta et. al., 2017, 2018; Peng et. al., 2018)
- Encourage argument sequences to match argument spans. (KL divergence - Yang et. al., 2017)

Prediction: Variants

- Constrained prediction (Das et. al., CL 2014)
- Greedy prediction (Henderson et. al., 2013)
- Joint prediction of all arguments (Das et. al., 2012)
- Joint prediction of frames and frame-elements (Yang et. al. EMNLP 2017; Peng et. al. NAACL 2018)
- Joint prediction of predicates, senses and arguments (Labeled Span Graphs, He et. al., ACL 2018)

Constrained Prediction

- Constraints:
    - Non-repetition of core frame-elements
    - Non-overlap of argument spans
    - Pairwise inclusions
    - Pairwise exclusions
- ILP formulation with constraints / Dual Decomposition (Rush et. al., 2010; Das, Martins & Smith, 2012)
- Other inference algorithms typically only respect non-overlap constraint.

Greedy prediction with transition-based models

- Dense graph with syntactic and semantic dependencies
- Linear-time algorithm to generate the entire graph (Henderson et. al., CL 2013)
- Neural model using variant of recurrent neural networks (Swayamdipta et. al., 2017)

Jointly predict all arguments for predicate

- Discriminative log-linear models (Toutanova, Haghighi, and Manning 2005), with global features and reranking framework.
- AD^3 with global constraints (Das et. al. 2012)

Joint prediction for frames and frame-elements

- Frames and frame-elements are mutually informative.
- Some heuristics about which frames and frame-elements can appear together (Yang et. al. 2017)
- More generic formulation (Peng et. al. 2018)
- Inference with AD^3 - A variant of dual decomposition

Multitask learning

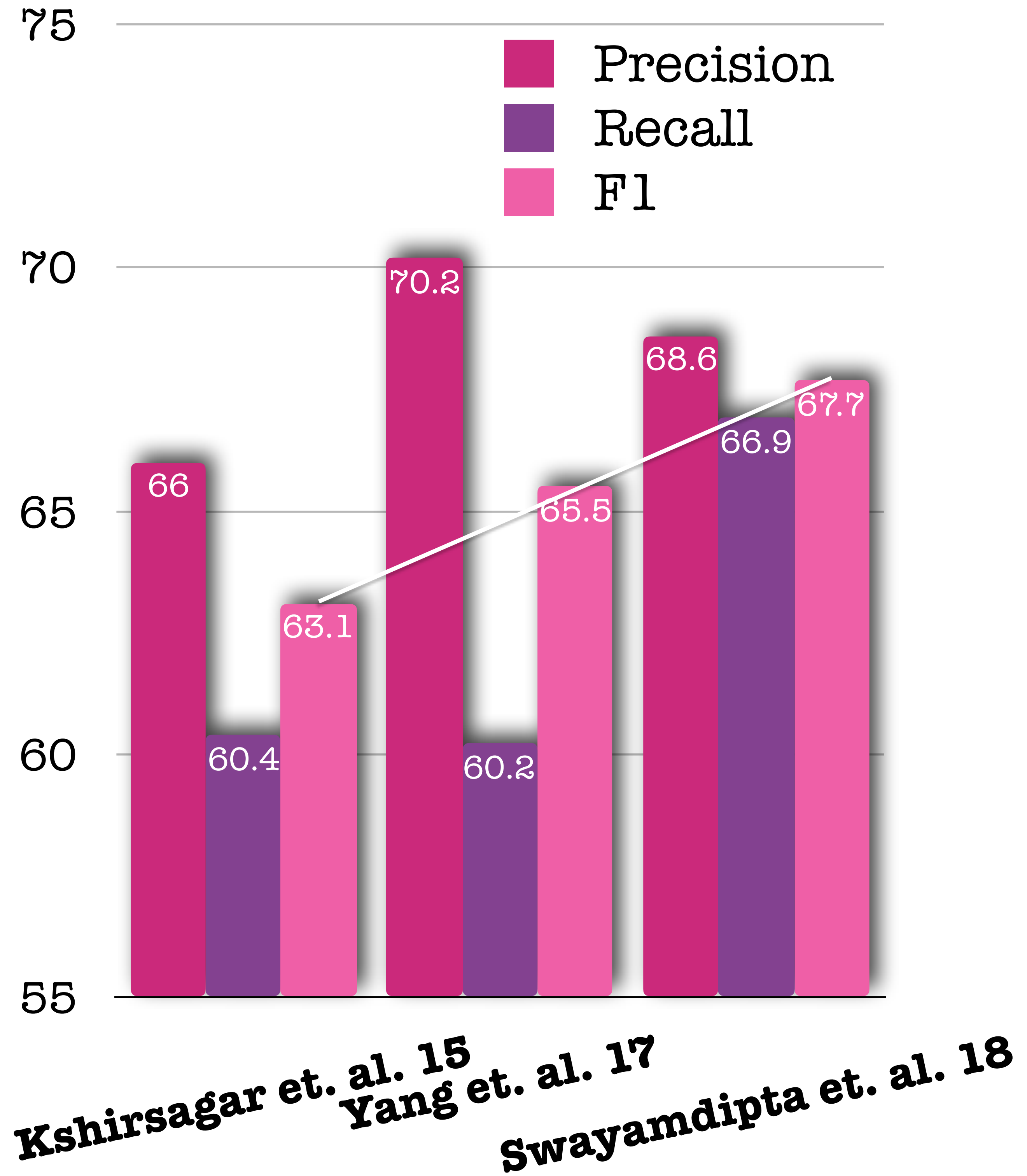
- With syntax
  - Full syntactic tree(CoNLL shared tasks 2008, 2009)
  - Only relevant parts -Scaffolding (Swayamdipta et. al., EMNLP 2018)
- With multiple semantic formalisms:
  - PropBank + FrameNet (Kshirsagar et. al. NAACL 2015; Fitzgerald et. al. EMNLP 2015)
  - Semantic dependencies + FrameNet (Peng et. al. NAACL 2018)

Syntactic + semantic graph

- CoNLL 2008 (Surdeaneau et. al., 2008)
- CoNLL 2009, multilingual (Hajivc et. al., 2009)
- Pipelined vs joint models
- Joint models
    - Graph-based
    - Transition-based

Syntactic Scaffolding

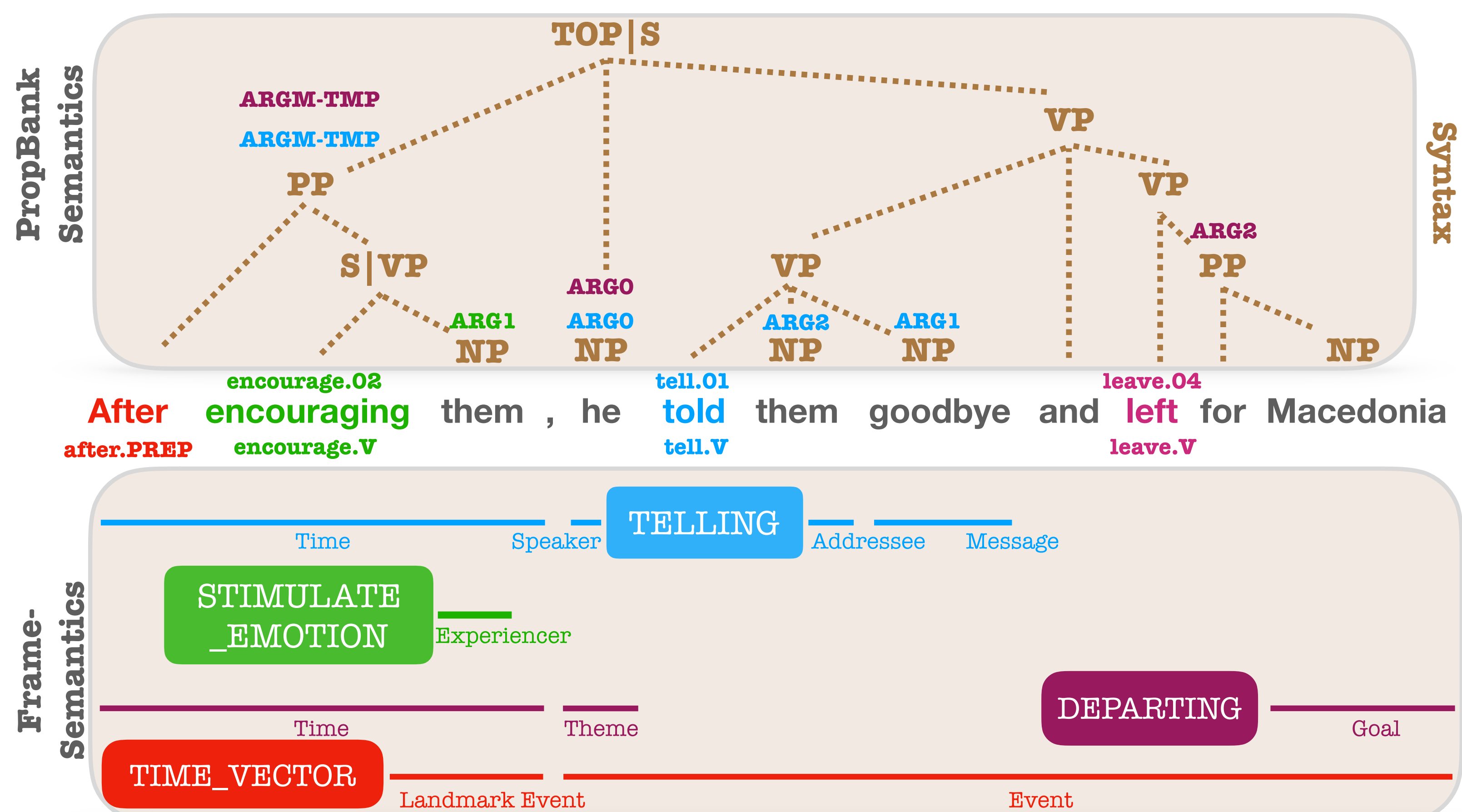
- Instead of learning entire syntactic trees, only learn parts of the tree which are relevant for SRL.
- Learns better representations of spans.
- “Scaffold” - can be discarded after training.
- Also helps other semantic tasks such as coreference resolution.



Swayamdipta et. al. (EMNLP, 2018)

PropBank + Frame-SRL

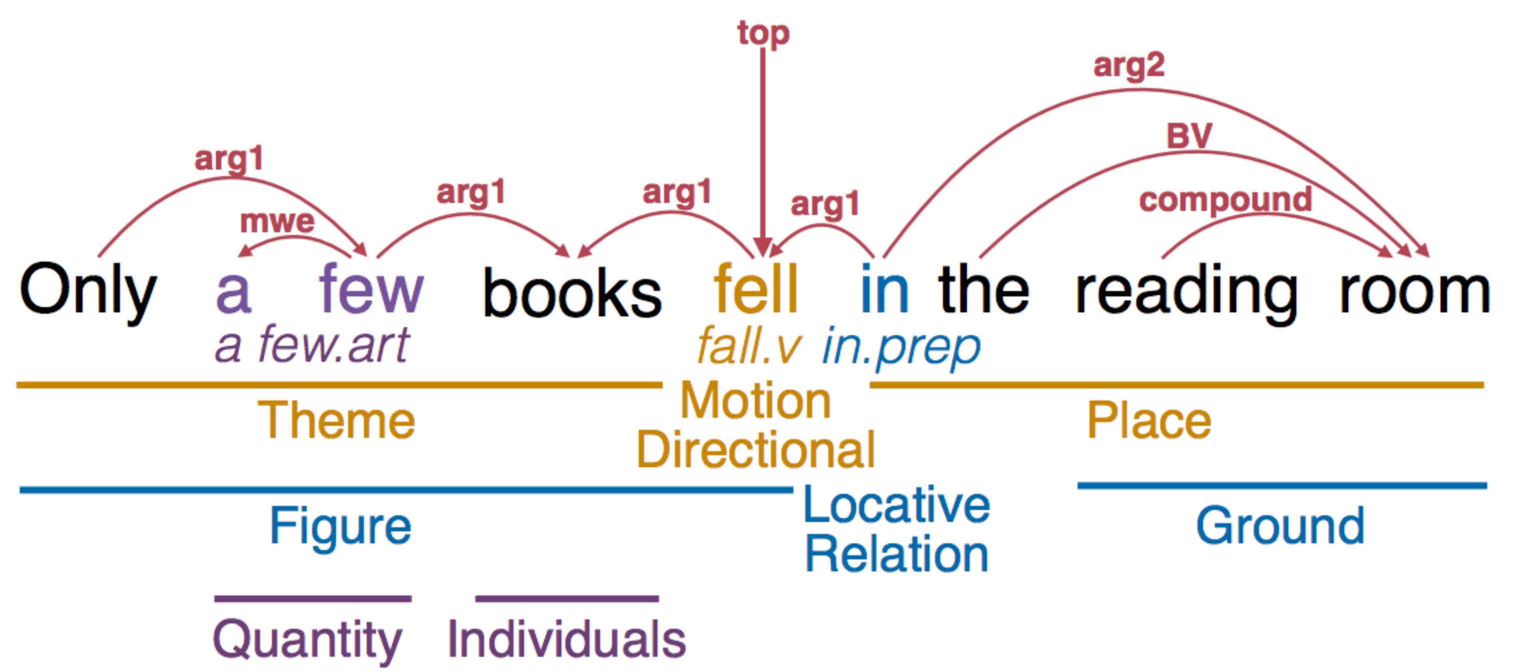
- Significant overlaps between formalisms
- PropBank much larger than FrameNet, so helps FrameSRL performance



Fitzgerald et. al. (2015)
Kshirsagar et. al. (2015)

Semantic Dependencies + Frame-SRL

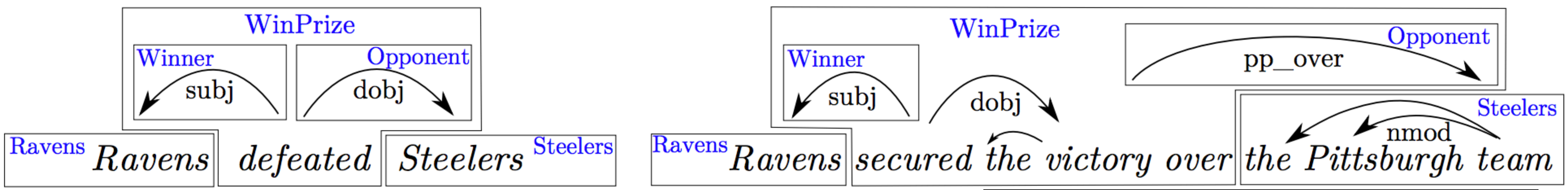
- Disjoint formalisms : span-based vs. dependency-based.
- Treat semantic dependencies as “latent” when predicting frame-elements.



Peng et. al. (2018)

Alternatives to supervised learning

Unsupervised approaches



Titov & Klementiev (2009)

- Two different syntactic trees with a common semantic representation.
- Clusters of syntactic structures correspond to semantic roles.

Semi-supervised approaches

- Pre-trained embeddings, based on language models.
- Seed examples and projection (Fursteanu & Lapata, 2012; Das et. al., CL 2014)

Opportunities!

- Frame-semantic parsing with heterogenous annotations (Kshirsagar et. al., 2015)
    - Frame Hierarchy
    - Exemplar annotations
- Grammatical Functions and Phrase Types, but only for gold arguments

Summary of Part 3

- Vanilla classifiers for subtasks can be improved on.
- Joint prediction
- Multi-task learning
- So, what’s stopping us?

Outline

1. Task of frame-SRL
2. Primary Subtasks
    a. Linear
    b. Neural models
3. Advanced Modeling
4. **Looking Forward / Multilingual Extensions**

Biggest challenges

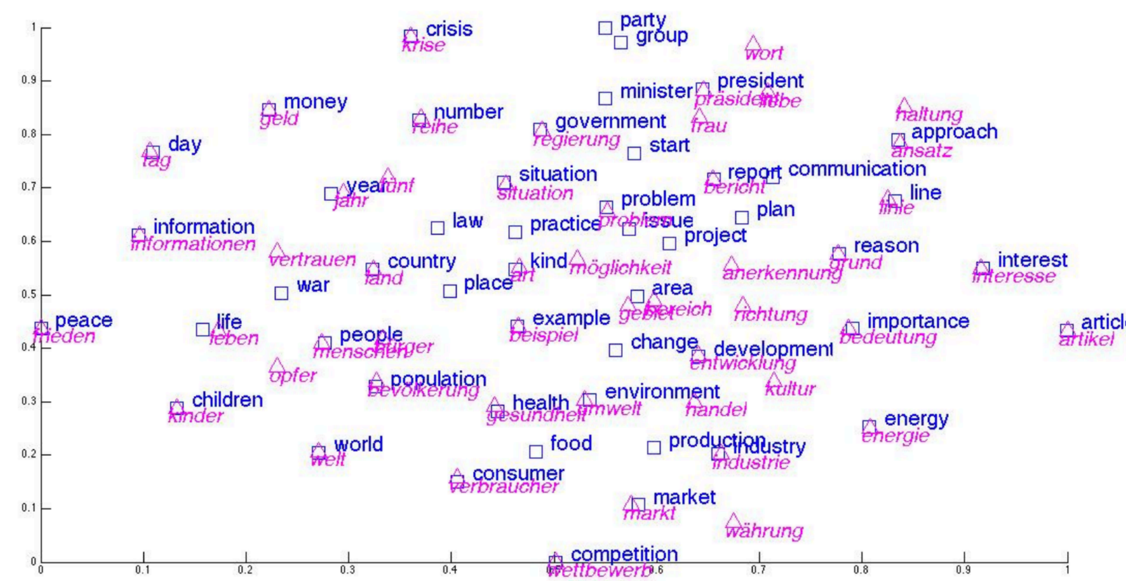
- Availability of data : More is better!
- Coverage
  - FrameNet+ (Pavlick et. al. ACL 2015)
  - Augmentation via Paraphrases (Rastogi & Van Durme, Workshops at ACL 2015)
- Domain Adaptation
  - Distributional semantics (Hartmann et. al. EACL 2017)

One glaring gap

One glaring gap

Multilingual SRL

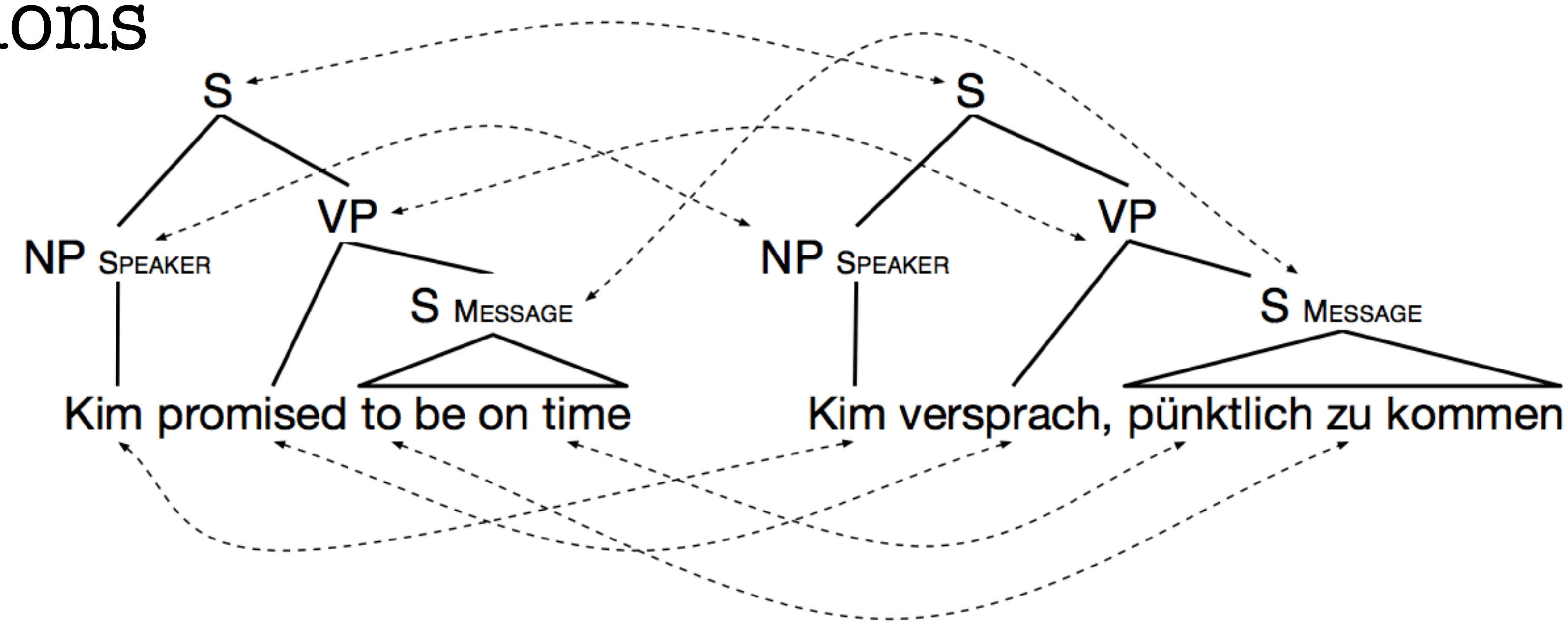
- Encouraged CoNLL 2009 for PropBank SRL
- Primary approach:
    - Build a single model for SRL
    - Apply to others via language-specific features / embeddings
- Multilingual models: Cross-lingual embeddings!



Luong et. al. (2015)

Cross-lingual annotation projection for SRL

- Needs parallel corpora
- Projection of annotations via lexical / syntactic alignments between sentences
- Not feasible without parallel data / highly accurate syntax



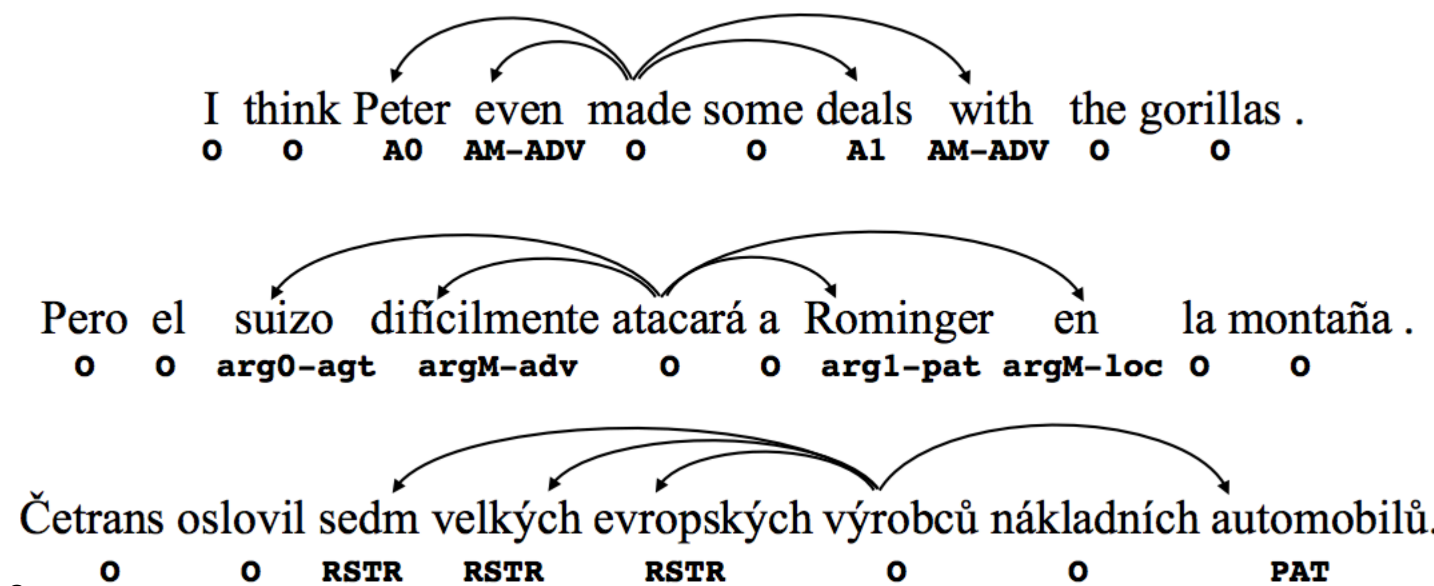
Pado & Lapata (2009)

Multilingual FN Efforts

- Annotation: WordNet + FrameNet (Hartmann & Gurevych, ACL 2013)
- Any language frame-semantic parsing (Johannsen et. al. 2015)
    - 9 languages in 2 domains
    - Using word-word translation
  - Inter-annotator agreement issues stemming from automatic target identification through word-word translation

Polyglot SRL

- Training data from pairs of languages merged
- Challenge: Differences in annotation schemes across languages.
- Multilingual word embeddings, learned from cross-lingual alignments (Ammar et. al., 2016)
- Maximum benefit reported for low-resource languages such as Catalan, when combined with English.



Mulcaire et. al. (ACL, 2018)

Applications of Multilingual FrameNet

- Translation using semantics as pivot.
- Cross-lingual transfer for downstream applications such as knowledge | information | relation extraction.
- Particular benefits for low-resource languages

Summary

- Part 1: Frame-SRL
  a. Graph induction
  b. Supervised Learning

- Part 2: Subtasks
  a. Target Identification
  b. Frame Identification
  c. Frame-Element Identification

- Part 3: Advanced Modeling

- Part 4: Looking Forward / Multilinguality