Semantics as a Foreign Language

Gabriel Stanovsky and Ido Dagan

EMNLP 2018

Semantic Dependency Parsing (SDP)

- A collection of three semantic formalisms (Oepen et al., 2014;2015)

Semantic Dependency Parsing (SDP)

- A collection of three semantic formalisms (Oepen et al., 2014;2015)
    a. DM (derived from MRS) (Copestake et al., 1999, Flickinger, 2000)

Semantic Dependency Parsing (SDP)

- A collection of three semantic formalisms (Oepen et al., 2014;2015)
    a. DM (derived from MRS)
    b. Prague Semantic Dependencies (PSD) (Hajic et al., 2012)

Semantic Dependency Parsing (SDP)

- A collection of three semantic formalisms (Oepen et al., 2014;2015)
    a. DM (derived from MRS)
    b. Prague Semantic Dependencies (PSD)
    c. Predicate Argument Structures (PAS) (Miyao et al., 2014)

Semantic Dependency Parsing (SDP)

- A collection of three semantic formalisms (Oepen et al., 2014;2015)
    a. DM (derived from MRS)
    b. Prague Semantic Dependencies (PSD)
    c. Predicate Argument Structures (PAS)

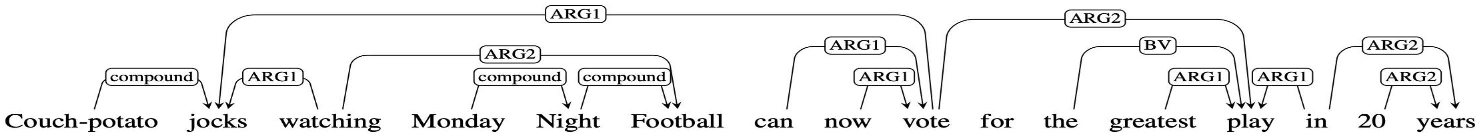
- Aim to capture **semantic** predicate-argument relations

Semantic Dependency Parsing (SDP)

- A collection of three semantic formalisms (Oepen et al., 2014;2015)
    a. DM (derived from MRS)
    b. Prague Semantic Dependencies (PSD)
    c. Predicate Argument Structures (PAS)

- Aim to capture **semantic** predicate-argument relations

- Represented in a graph structure

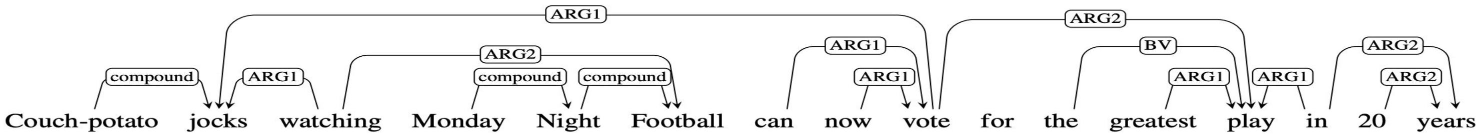


Semantic Dependency Parsing (SDP)

- A collection of three semantic formalisms (Oepen et al., 2014;2015)
    a. DM (derived from MRS)
    b. Prague Semantic Dependencies (PSD)
    c. Predicate Argument Structures (PAS)

- Aim to capture **semantic** predicate-argument relations

- Represented in a graph structure
    a. Nodes: single words from the sentence

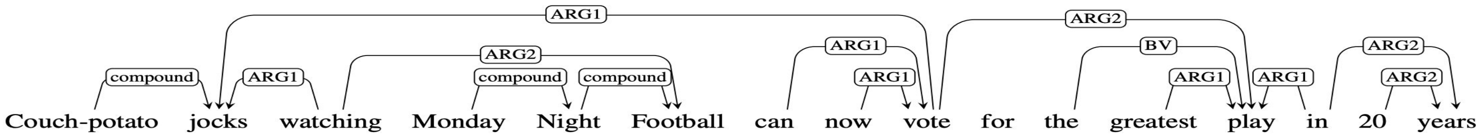


Semantic Dependency Parsing (SDP)

- A collection of three semantic formalisms (Oepen et al., 2014;2015)
    a. DM (derived from MRS)
    b. Prague Semantic Dependencies (PSD)
    c. Predicate Argument Structures (PAS)

- Aim to capture **semantic** predicate-argument relations

- Represented in a graph structure
    a. Nodes: single words from the sentence
    b. Labeled edges: semantic relations, according to the paradigm



Outline

Outline

- SDP as Machine Translation
  - Different formalisms as foreign languages
  - **Motivation**: downstream tasks, inter-task analysis, extendable framework

Outline

- SDP as Machine Translation
  - Different formalisms as foreign languages
  - **Motivation**: downstream tasks, inter-task analysis, extendable framework
  - **Previous work** explored the relation between MT and semantics
    (Wong and Mooney, 2007), (Vinyals et al., 2015), (Flanigan et al., 2016)

Outline

- SDP as Machine Translation
  - Different formalisms as foreign languages
  - **Motivation**: downstream tasks, inter-task analysis, extendable framework
  - **Previous work** explored the relation between MT and semantics
    (Wong and Mooney, 2007), (Vinyals et al., 2015), (Flanigan et al., 2016)

- Model
  - Seq2Seq
  - Directed graph linearization

Outline

- SDP as Machine Translation
  - Different formalisms as foreign languages
  - **Motivation**: downstream tasks, inter-task analysis, extendable framework
  - **Previous work** explored the relation between MT and semantics (Wong and Mooney, 2007), (Vinyals et al., 2015), (Flanigan et al., 2016)
- Model
  - Seq2Seq
  - Directed graph linearization
- Results
  - Raw text to SDP (near state-of-the-art)
  - Novel inter-task analysis

Outline

- **SDP as Machine Translation**
  - Different formalisms as foreign languages
  - **Motivation**: downstream tasks, inter-task analysis, extendable framework
  - **Previous work** explored the relation between MT and semantics
    (Wong and Mooney, 2007), (Vinyals et al., 2015), (Flanigan et al., 2016)
- Model
  - Seq2Seq
  - Linearization
- Results
  - Raw text -> SDP (near state-of-the-art)
  - Novel inter-task analysis

Semantic Dependencies as MT

Source

Raw sentence

Target

Semantic Dependencies as MT

Source

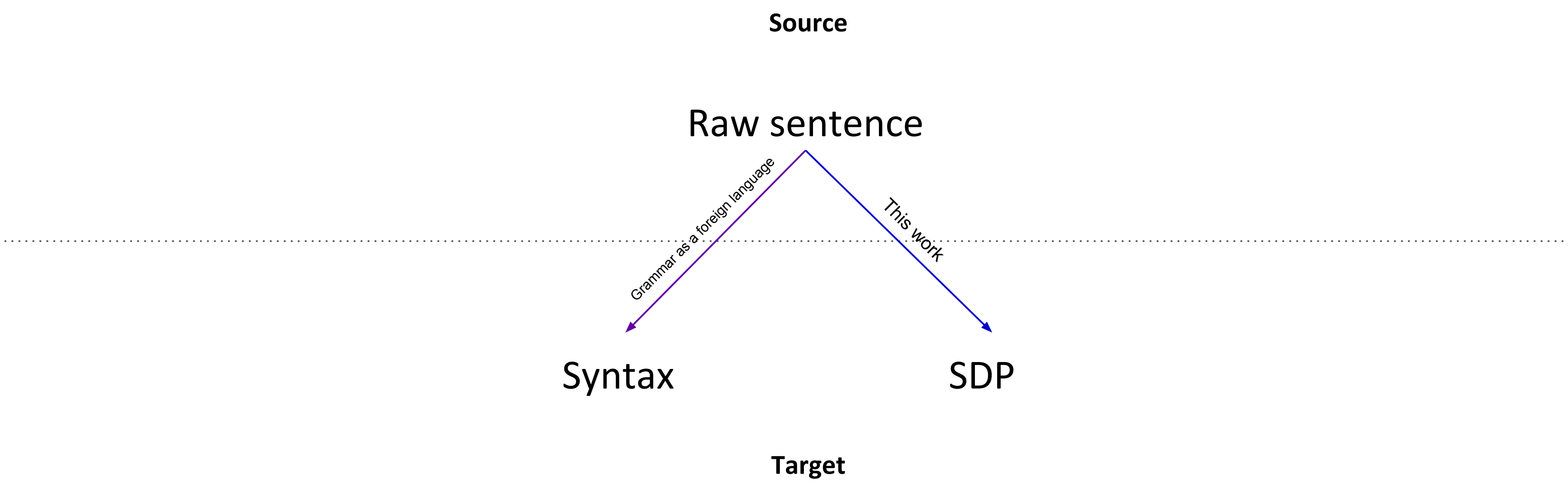
Raw sentence

Grammar as a foreign language

Syntax

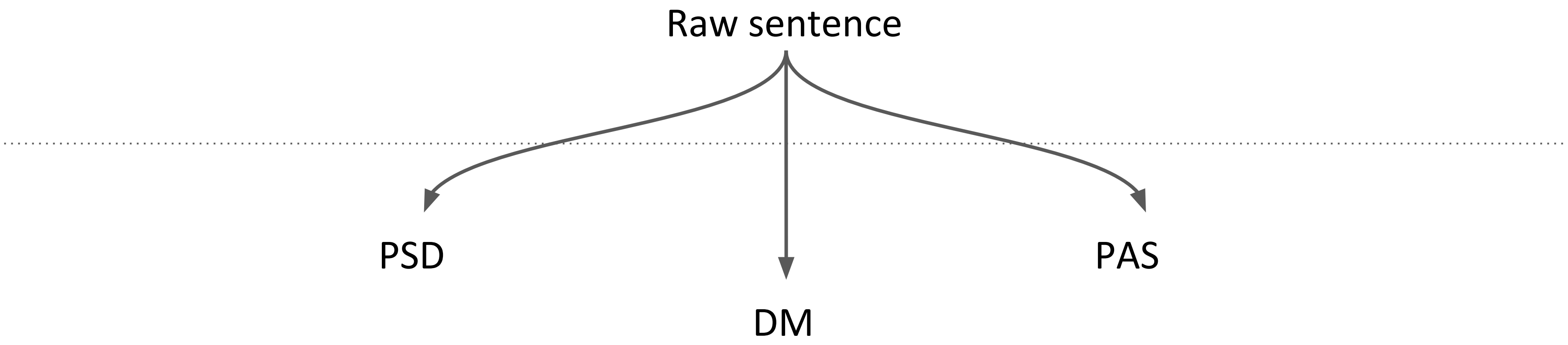
Target

Semantic Dependencies as MT



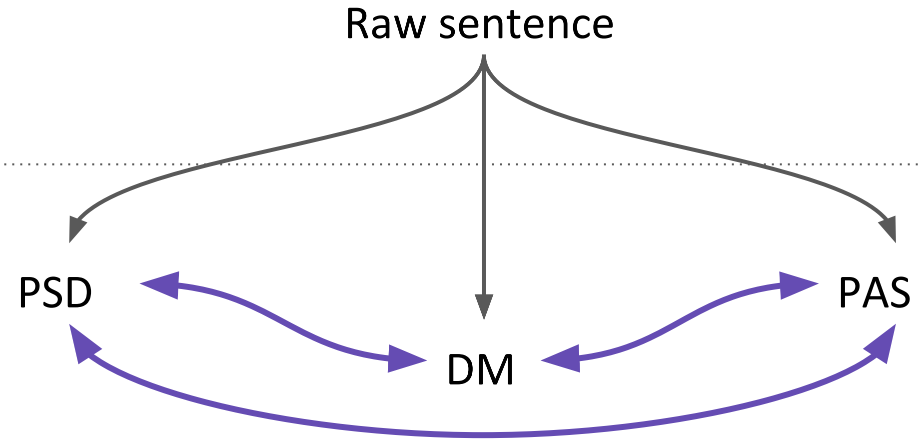
Semantic Dependencies as MT

- Standard MTL: 3 tasks



Semantic Dependencies as MT

- Standard MTL: 3 tasks



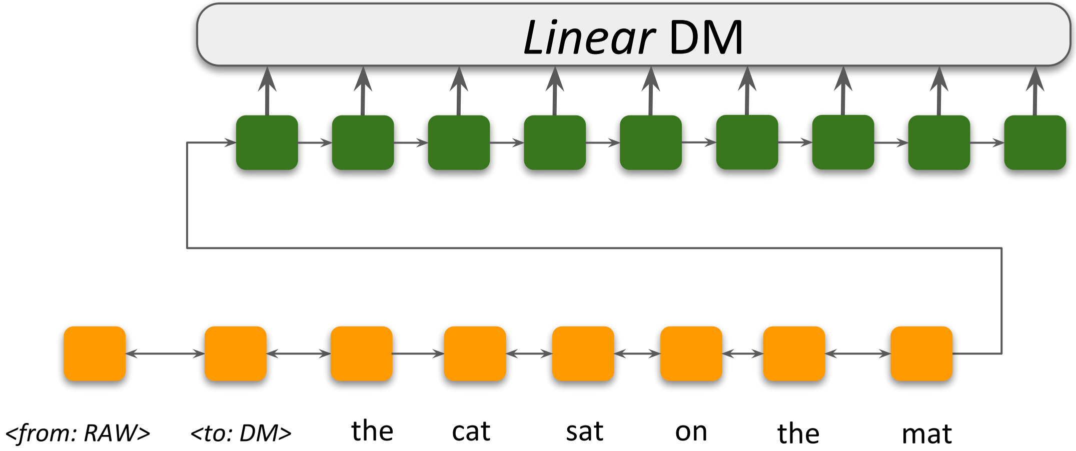
- *Inter-task* translation (9 tasks)

Outline

- SDP as Machine Translation
    - Motivation: downstream tasks
    - Different formalisms as foreign languages
- **Model**
    - Seq2Seq
    - Linearization
- Results
    - Raw text -> SDP (near state-of-the-art)
    - Novel inter-task analysis

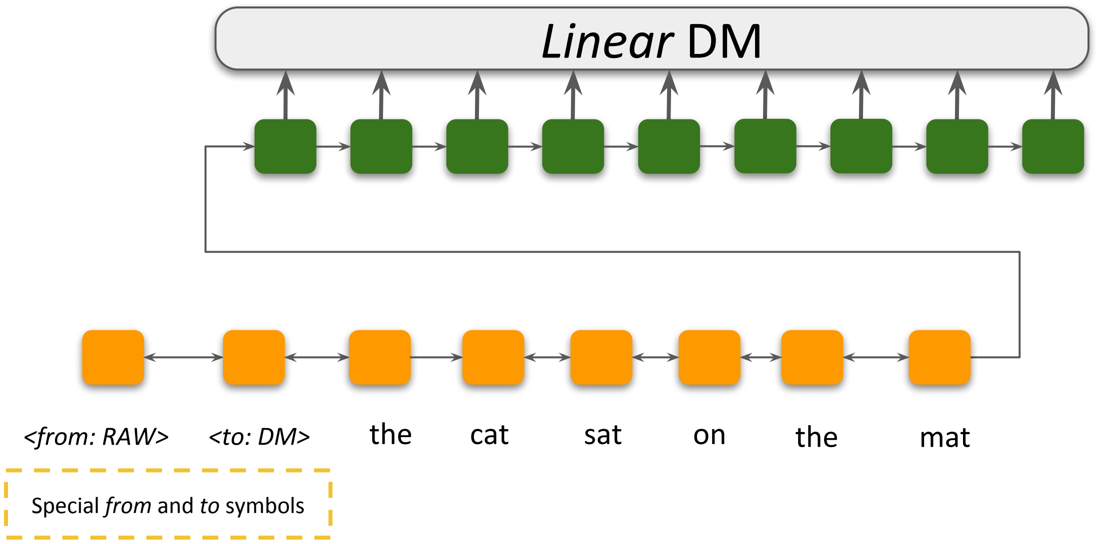
Our Model Ⅰ: Raw -> SDP^x

- Seq2Seq translation model:
  - Bi-LSTM encoder-decoder with attention



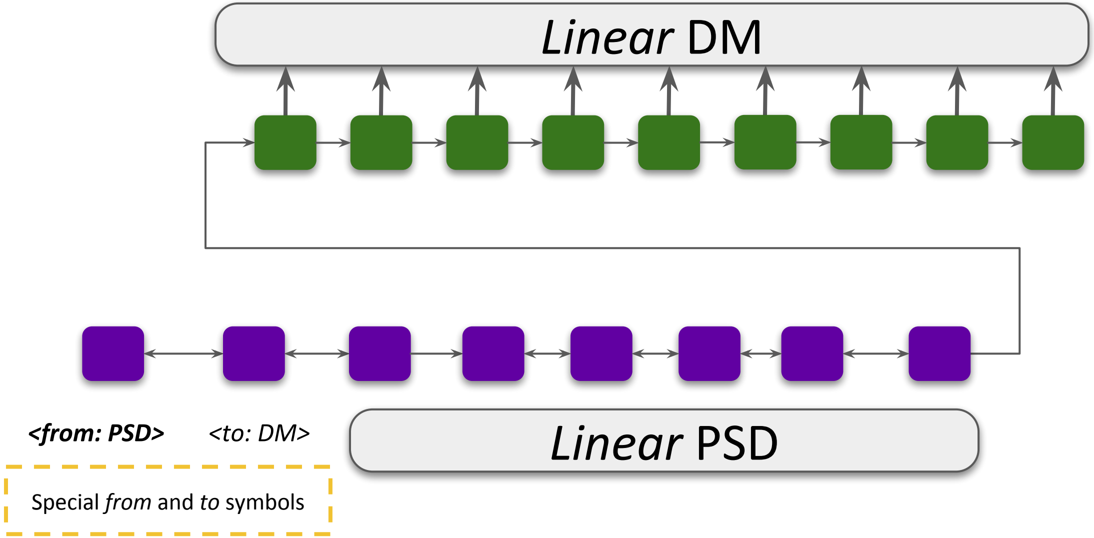
Our Model Ⅰ: Raw -> SDP^x

- Seq2Seq translation model:
  - Bi-LSTM encoder-decoder with attention



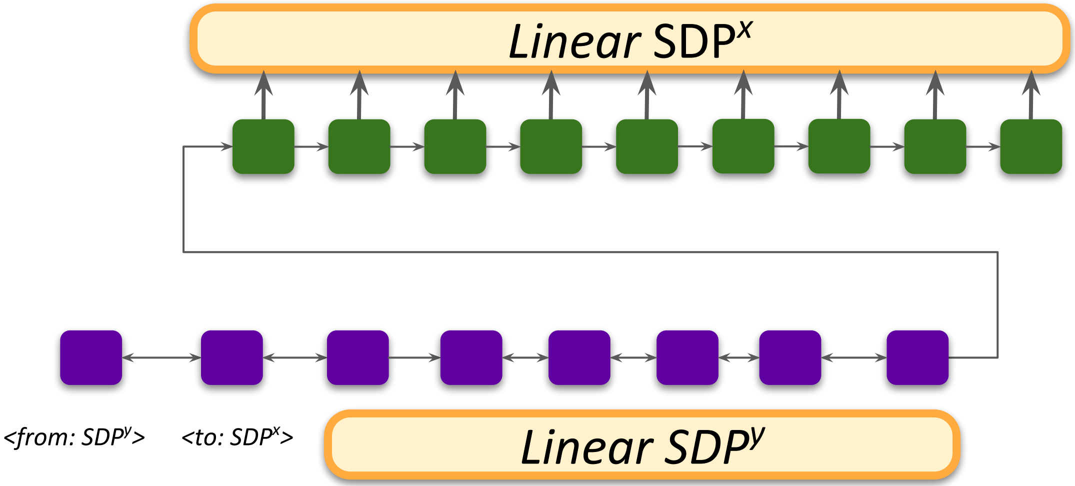
Our Model Ⅱ: SDP^y -> SDP^x

- Seq2Seq translation model:
    - Bi-LSTM encoder-decoder with attention



Our Model

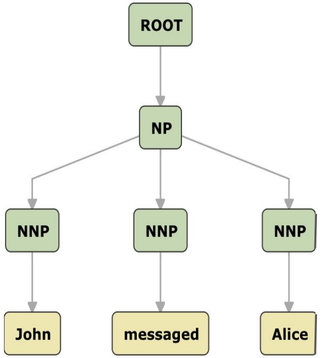
Seq2seq prediction requires a **1:1 linearization function**



Linearization: Background

- Previous work used bracketed tree linearization

  (Vinyals et al., 2015; Konstas et al., 2017; Buys and Blunsom, 2017)

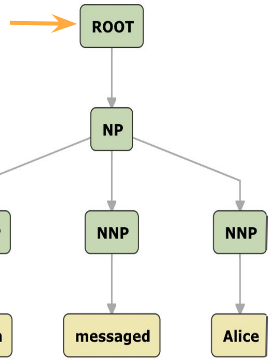


Linearization: Background

- Previous work used bracketed tree linearization

  (ROOT

  (Vinyals et al., 2015; Konstas et al., 2017; Buys and Blunsom, 2017)

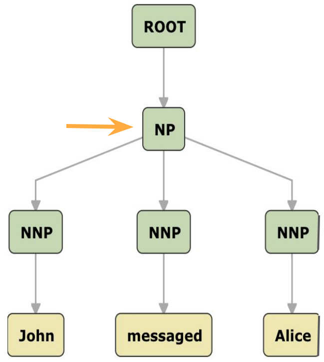


Linearization: Background

- Previous work used bracketed tree linearization

  (ROOT ***(NP***

  (Vinyals et al., 2015; Konstas et al., 2017; Buys and Blunsom, 2017)

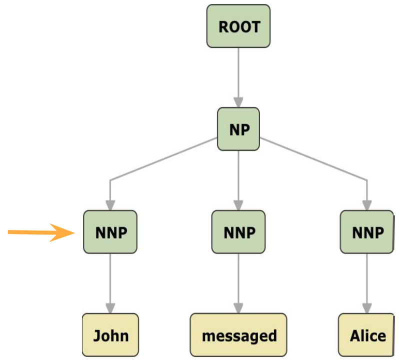


Linearization: Background

- Previous work used bracketed tree linearization

  *(ROOT (NP **(NNP***

  (Vinyals et al., 2015; Konstas et al., 2017; Buys and Blunsom, 2017)

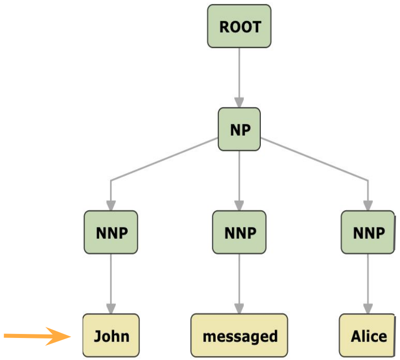


Linearization: Background

- Previous work used bracketed tree linearization

    (ROOT (NP (NNP **John**

    (Vinyals et al., 2015; Konstas et al., 2017; Buys and Blunsom, 2017)

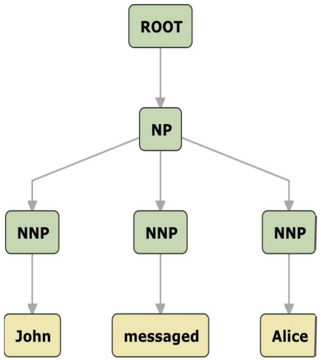


Linearization: Background

- Previous work used bracketed tree linearization

  (ROOT (NP (NNP **John** *)NNP)NP (VP* **messaged** *(NP* **Alice** *)NP)VP)ROOT*

  (Vinyals et al., 2015; Konstas et al., 2017; Buys and Blunsom, 2017)



Linearization: Background

- Previous work used bracketed tree linearization

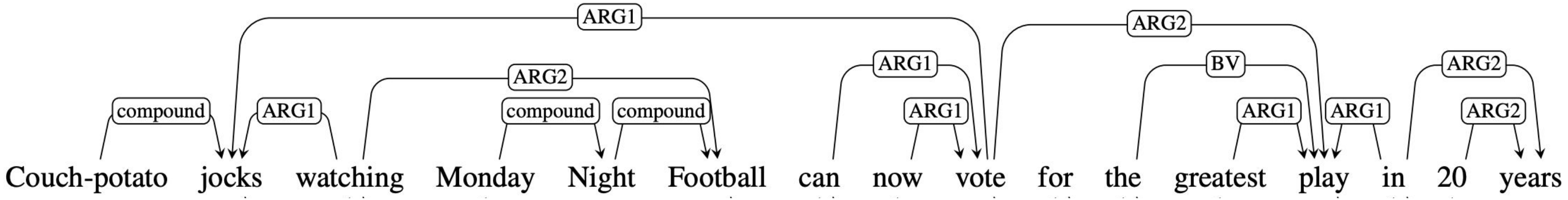
  (ROOT (NP (NNP **John** *)NNP)NP (VP* **messaged** *(NP* **Alice** *)NP)VP)ROOT*

  (Vinyals et al., 2015; Konstas et al., 2017; Buys and Blunsom, 2017)

- Depth-first representation **doesn't directly apply to SDP graphs**
  - Non-connected components
  - Re-entrencies

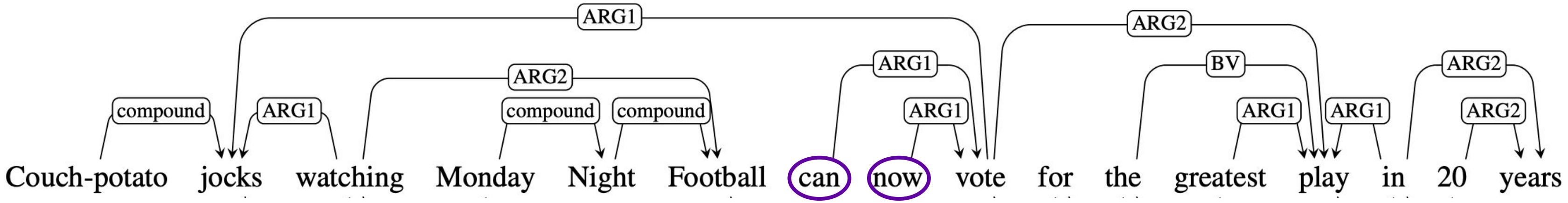
SDP Linearization (Connectivity)

- **Problem: No single root** from which to start linearization



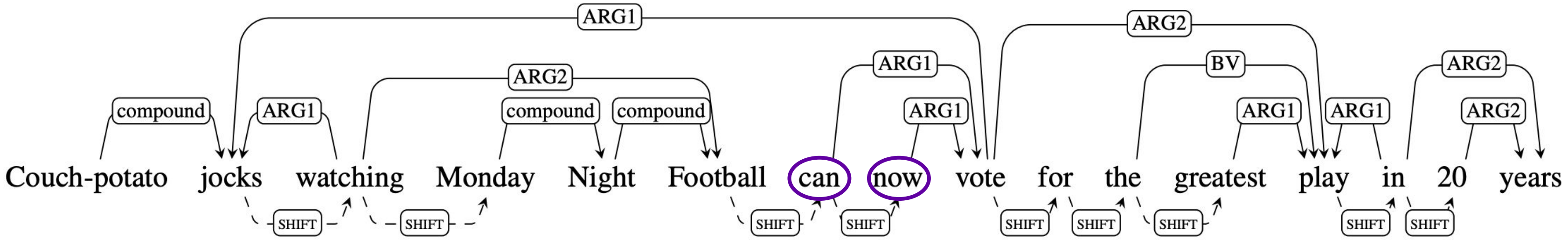
SDP Linearization (Connectivity)

- **Problem: No single root** from which to start linearization



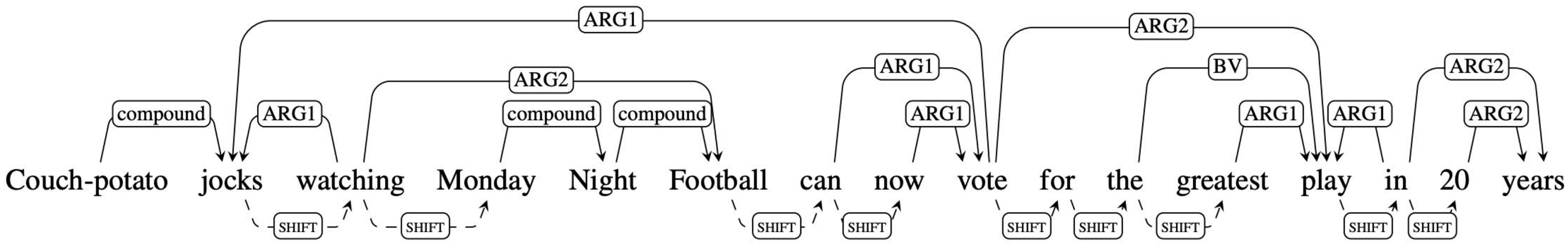
SDP Linearization (Connectivity)

- **Problem: No single root** from which to start linearization



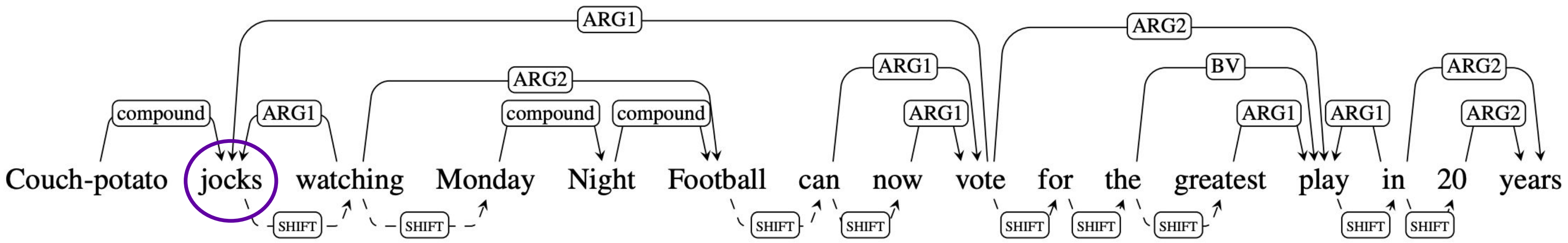
- **Solution**: Artificial SHIFT edges between non-connected adjacent words
  - All nodes are now reachable from the first word

SDP Linearization (Re-entrancies)



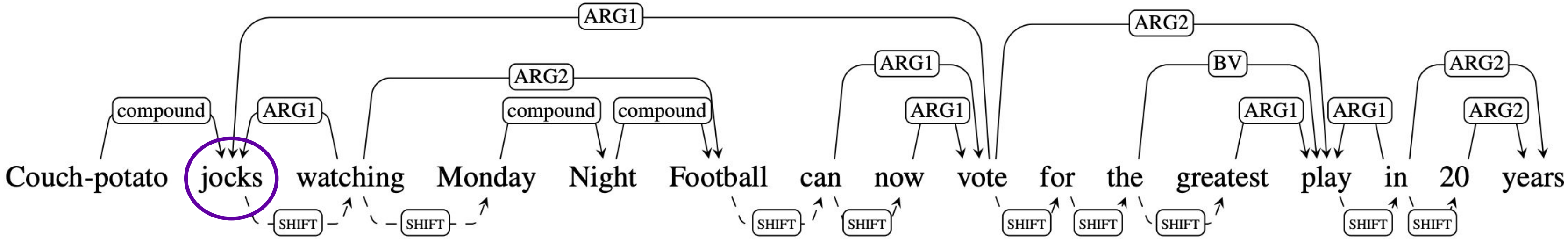
- Re-entrancies require a 1:1 node representation

SDP Linearization (Re-entrencies)



- Re-entrancies require a 1:1 node representation

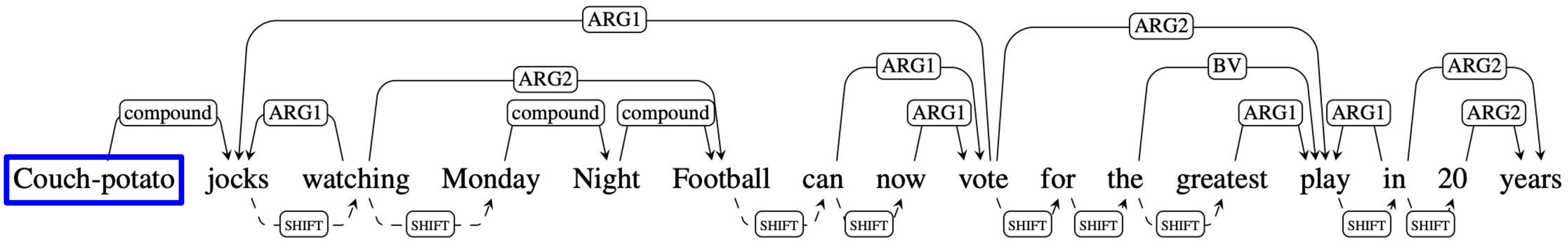
SDP Linearization (Re-entrencies)



- Re-entrancies require a 1:1 node representation

(relative index / surface form)

SDP Linearization (Re-entrencies)

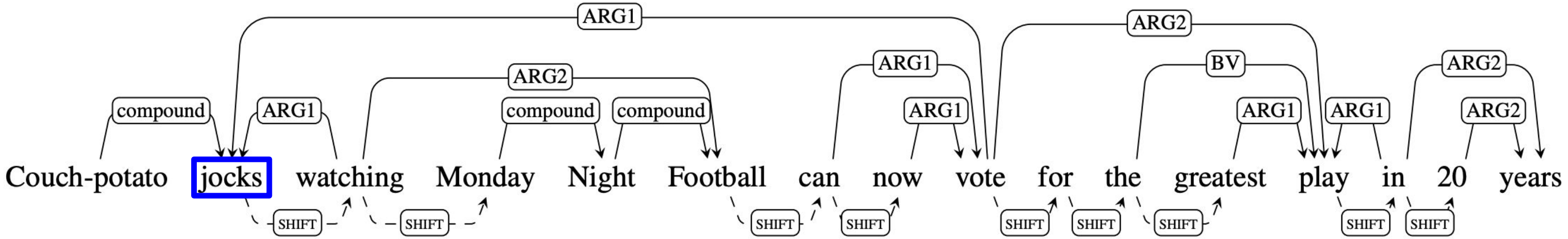


- Re-entrancies require a 1:1 node representation

(relative index / surface form)

0/couch-potato

SDP Linearization (Re-entrencies)

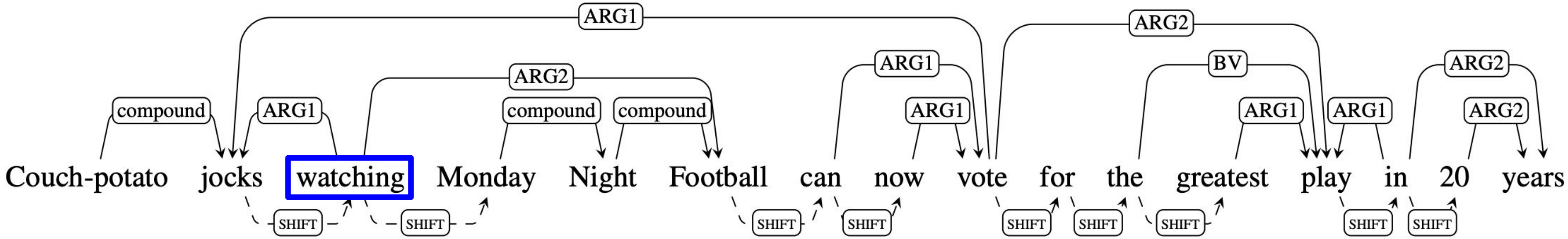


- Re-entrancies require a 1:1 node representation

(relative index / surface form)

0/couch-potato *compound* +1/jocks

SDP Linearization (Re-entrencies)

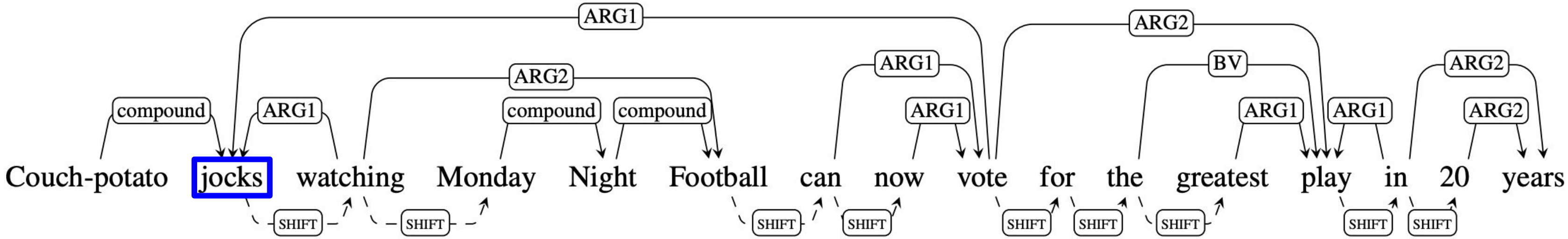


- Re-entrancies require a 1:1 node representation

(relative index / surface form)

0/couch-potato *compound* +1/jocks *shift* +1/watching

SDP Linearization (Re-entrencies)



- Re-entrancies require a 1:1 node representation

(relative index / surface form)

0/couch-potato *compound* +1/jocks *shift* +1/watching *ARG1* -1/jocks

Outline

- SDP as Machine Translation
  - Motivation: downstream tasks
  - Different formalisms as foreign languages
- **Model**
  - Linearization
  - Dual Encoder-Single decoder Seq2Seq
- **Results**
  - Raw text -> SDP (near state-of-the-art)
  - Novel inter-task analysis

Experimental Setup

- Train samples per task: 35,657 sentences (Oepen et al., 2015)
  - 9 translation tasks

Experimental Setup

- Train samples per task: 35,657 sentences (Oepen et al., 2015)
    - 9 translation tasks

- Total training samples: **320,913 source-target pairs**

Experimental Setup

- Train samples per task: 35,657 sentences (Oepen et al., 2015)
  - 9 translation tasks
- Total training samples: **320,913 source-target pairs**
- Trained in batches between the 9 different tasks

Evaluations:RAW $\longrightarrow$ SDP$_{(x)}$

	DM	**PAS**	**PSD**	**Avg.**
Peng et al. (2017a)	**90.4**	**92.7**	78.5	**87.2**
Single	70.1	73.6	63.6	69.1
MTL PRIMARY	82.4	87.2	71.4	80.3
MTL PRIMARY+AUX	87.5	90.9	**80.3**	86.2

Labeled F1 score

Evaluations:RAW $\longrightarrow$ SDP$_{(x)}$

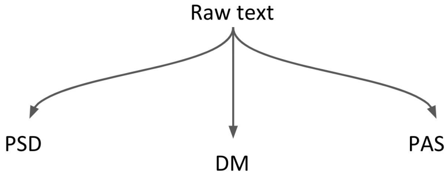
Raw text

DM

	DM	PAS	PSD	Avg.
Peng et al. (2017a)	**90.4**	**92.7**	78.5	**87.2**
Single	70.1	73.6	63.6	69.1
MTL PRIMARY	82.4	87.2	71.4	80.3
MTL PRIMARY+AUX	87.5	90.9	**80.3**	86.2

Labeled F1 score

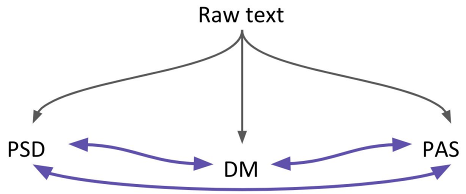
Evaluations:RAW $\longrightarrow$ SDP$_{(x)}$



	DM	PAS	PSD	Avg.
Peng et al. (2017a)	**90.4**	**92.7**	78.5	**87.2**
Single	70.1	73.6	63.6	69.1
MTL PRIMARY	82.4	87.2	71.4	80.3
MTL PRIMARY+AUX	87.5	90.9	**80.3**	86.2

Labeled F1 score

Evaluations:RAW $\longrightarrow$ SDP$_{(x)}$



	DM	PAS	PSD	Avg.
Peng et al. (2017a)	**90.4**	**92.7**	78.5	**87.2**
Single	70.1	73.6	63.6	69.1
MTL PRIMARY	82.4	87.2	71.4	80.3
MTL PRIMARY+AUX	87.5	90.9	**80.3**	86.2

Labeled F1 score

Evaluations:$\text{SDP}_{(a)} \longrightarrow \text{SDP}_{(b)}$

To \From	**DM**	**PAS**	**PSD**	**Avg.**
DM		96.1	92.4	94.3
PAS	95.7		91.7	93.7
PSD	89.5	87.6		88.6
Avg.	92.6	91.9	92.1	

Labeled F1 score

Evaluations:SDP$_{(a)} \longrightarrow$ SDP$_{(b)}$

To \From	DM	PAS	PSD	Avg.
DM		96.1	92.4	94.3
PAS	95.7		91.7	93.7
PSD	89.5	87.6		88.6
Avg.	92.6	91.9	92.1	

Labeled F1 score

- Translating **between representations** is easier than parsing from raw text

Evaluations:$SDP_{(a)} \longrightarrow SDP_{(b)}$

To \From	DM	PAS	PSD	Avg.
DM		$\boxed{96.1}$	92.4	94.3
PAS	$\boxed{95.7}$		91.7	93.7
PSD	89.5	87.6		88.6
Avg.	92.6	91.9	92.1	

Labeled F1 score

- Translating **between representations** is easier than parsing from raw text
- Easy to convert between PAS and DM

Evaluations: $SDP_{(a)} \longrightarrow SDP_{(b)}$

To \From	DM	PAS	PSD	Avg.
DM		96.1	92.4	94.3
PAS	95.7		91.7	93.7
PSD	89.5	87.6		88.6
Avg.	92.6	91.9	92.1	

Labeled F1 score

- Translating **between representations** is easier than parsing from raw text
- Easy to convert between PAS and DM
- PSD is a good input, but relatively hard output

Conclusions

Conclusions

- Effective graph linearization for SDP
  - Near state-of-the-art results

Conclusions

- Effective graph linearization for SDP
    - Near state-of-the-art results
- Inter-task analysis
    - Enabled by the generic seq2seq framework

Conclusions

- Effective graph linearization for SDP
    - Near state-of-the-art results
- Inter-task analysis
    - Enabled by the generic seq2seq framework
- Future work
    - Apply linearizations in downstream tasks (NMT)
    - Add more representations (AMR, UD)

Conclusions

- Effective graph linearization for SDP
  - Near state-of-the-art results
- Inter-task analysis
  - Enabled by the generic seq2seq framework
- Future work
  - Apply linearizations in downstream tasks (NMT)
  - Add more representations (AMR, UD)

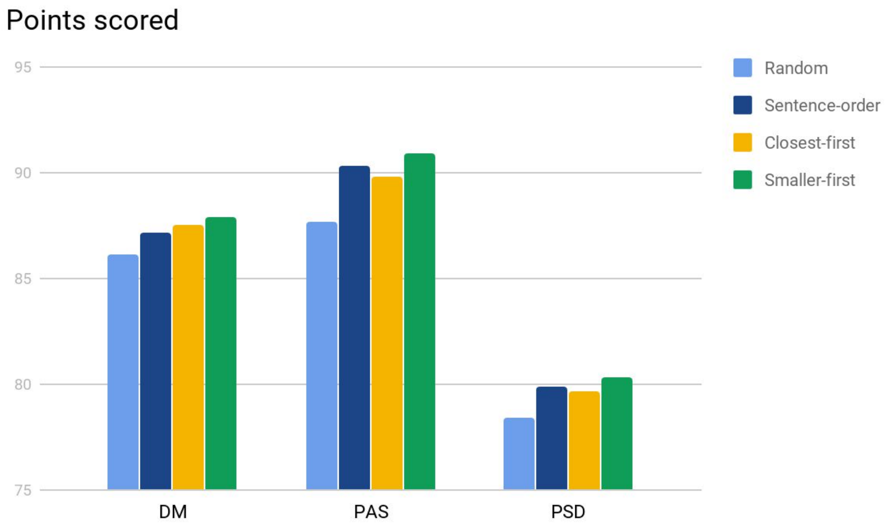
Thanks for listening!

BACKUP SLIDES

Evaluations: Node ordering

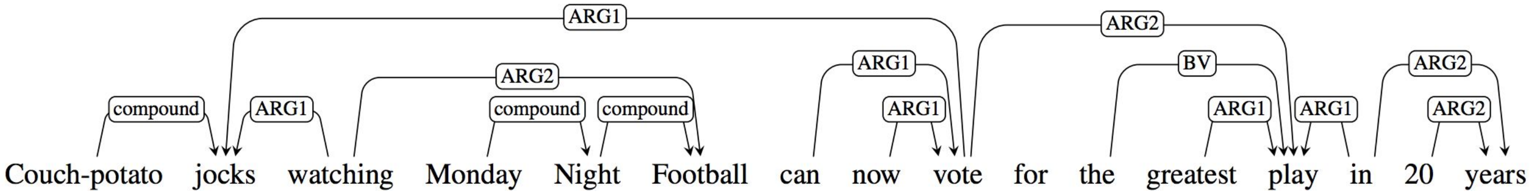
- Smaller-first ordering consistently does better across all representations

	DM	PAS	PSD	Avg.
Random	86.1	87.7	78.4	84.1
Sentence order	87.2	90.3	79.9	85.8
Closest words	87.5	89.8	79.7	85.8
Smaller-first	**87.9**	**90.9**	**80.3**	**86.2**



Semantic Formalisms

- Many formalisms try to represent the ***meaning*** of a sentence
  - MRS, AMR, PSD, SDP, etc…



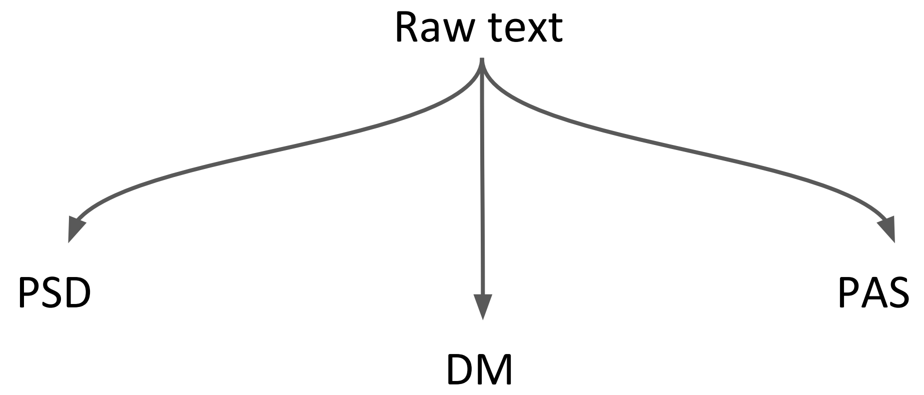
Semantic Dependencies as MT

- Syntactic parsing as MT ("Grammar as a foreign language", [Vinyals et. al, 2014])

  Jane had a cat $\rightarrow (_{ROOT}\ (_{S}\ (_{NP}\ \textbf{Jane}\ )_{NP}\ (_{VP}\ \textbf{had}\ (_{NP}\ \textbf{a cat}\ )_{NP}\ )_{VP}\ \textbf{.}\ )_{S}\ )_{ROOT}$

- We aim to do the same for SDP
  - The different formalisms as foreign languages

Semantic Dependencies as MT



Our Model

- Seq2Seq translation model:
  - Bi-LSTM encoder-decoder with attention

- Two shared encoders
  - From raw to SDP graphs
  - Between SDP graphs

- One global decoder for all samples

- Add “*<from:X> <to:Y>”* tags to input as preprocessing
  - Where X, Y in {RAW, PSD, PAS, DM}
  - Different than Google’s NMT, which **didn’t** have *<from:X>* tags
    - No “code-switching” is allowed

Motivation

- Linearization is an easy way to **plug-in** predicted structures in NNs
  - MT Target side syntax
    (Aharoni and Goldberg, 2017; Wang et al., 2018)

- Allows Inter-task analysis

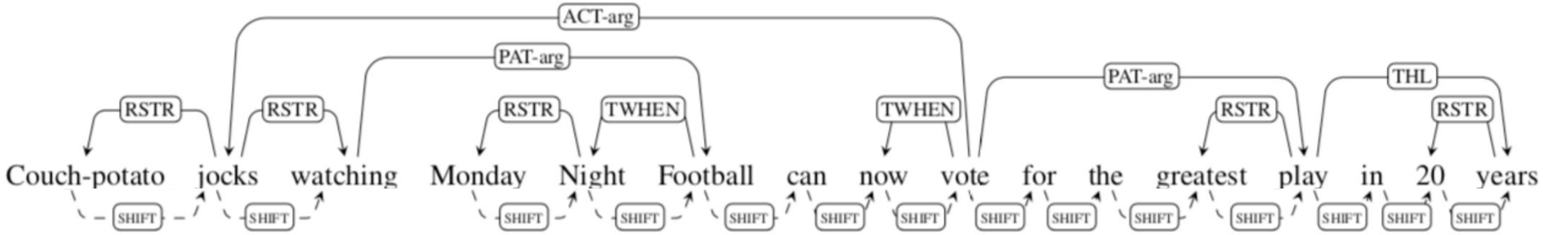
Motivation

- Linearization is an easy way to **plug-in** predicted structures in NNs
  - MT Target side syntax
    (Aharoni and Goldberg, 2017; Wang et al., 2018)

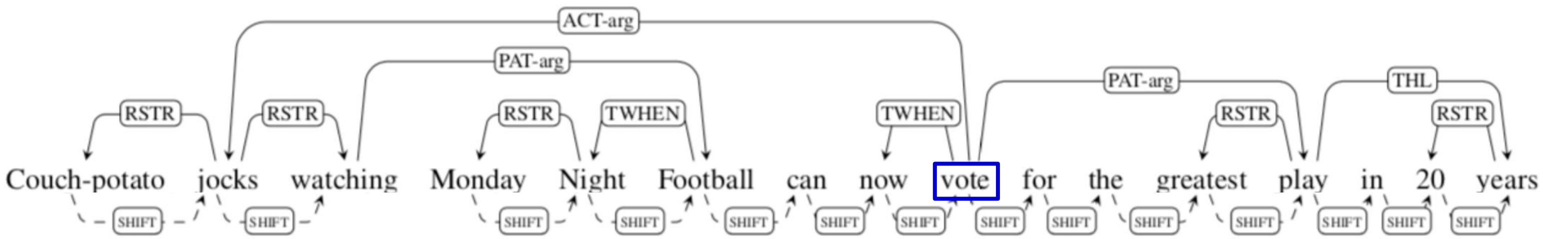
Motivation

- Linearization is an easy way to **plug-in** predicted structures in NNs
  - MT Target side syntax
    (Aharoni and Goldberg, 2017; Wang et al., 2018)
- Allows Inter-task analysis
- Easily extendable framework

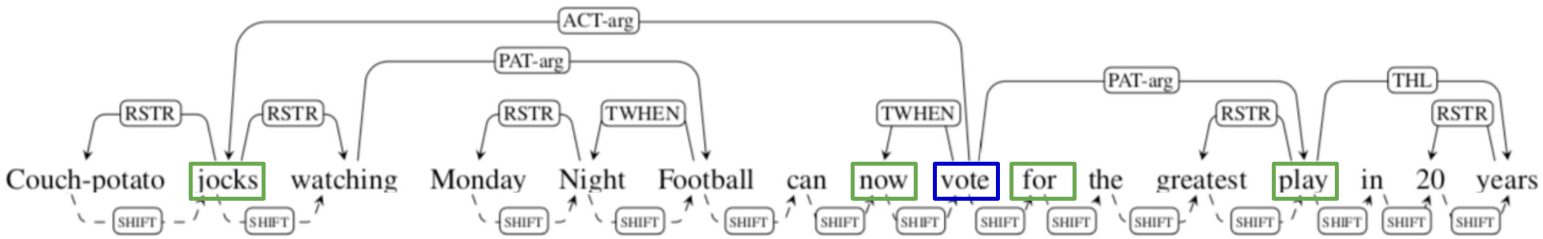
SDP Linearization (node ordering)



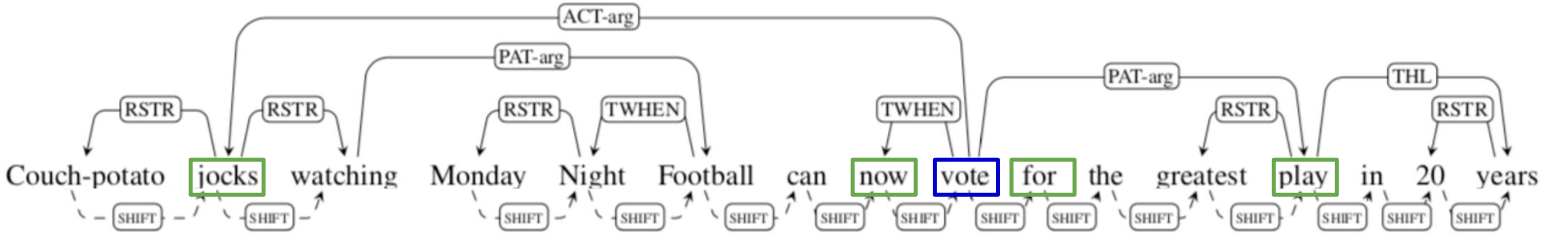
SDP Linearization (node ordering)



SDP Linearization (node ordering)



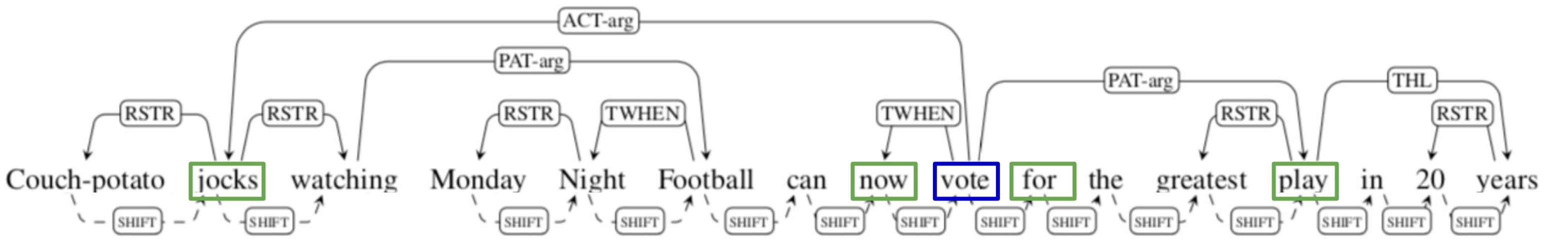
SDP Linearization (node ordering)



- Neighbor orderings:

  a. *Random - (play, for, jocks, now)*
  b. *Closest-first*
  c. *Sentence-order*
  d. *Smaller-first*

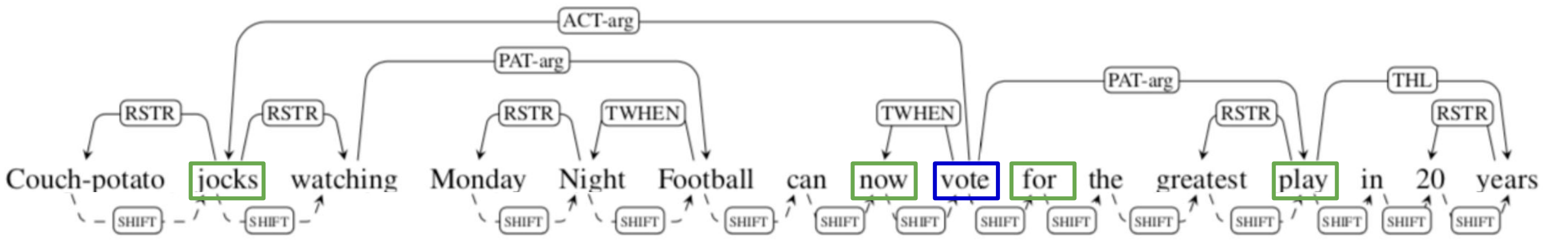
SDP Linearization



- Neighbor orderings:

  a. *Random*
  b. *Closest-first - (now, for, play, jocks)*
  c. *Sentence-order*
  d. *Smaller-first*

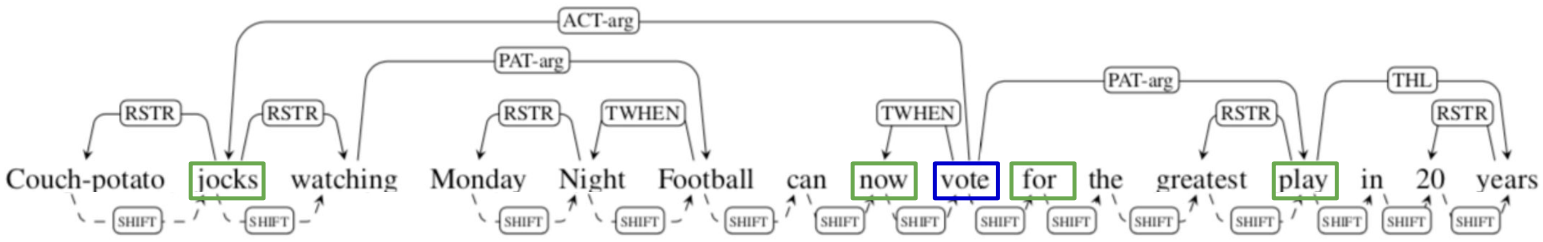
SDP Linearization



- Neighbor orderings:

  a. *Random*
  b. *Closest-first*
  c. *Sentence-order - (jocks, now, for, play)*
  d. *Smaller-first*

SDP Linearization



- Neighbor orderings:

  a. *Random*
  b. *Closest-first*
  c. *Sentence-order*
  d. *Smaller-first - (now, play, for, jocks)*

Evaluations: Node ordering

- **Smaller-first** ordering consistently does better across all representations

	DM	PAS	PSD	Avg.
Random	86.1	87.7	78.4	84.1
Sentence order	87.2	90.3	79.9	85.8
Closest words	87.5	89.8	79.7	85.8
Smaller-first	**87.9**	**90.9**	**80.3**	**86.2**

Labeled F1 score