Paris and Stanford at EPE 2017: Downstream Evaluation of Graph-based Dependency Representations

Sebastian Schuster, Éric Villemonte de la Clergerie, Marie Candito, Benoît Sagot, Christopher Manning, and Djamé Seddah

Stanford University/INRIA/Université Paris Diderot/Université Paris Sorbonne

September 20, 2017

Motivation

We developed **graph-based representations** that can be derived from Universal Dependency trees

Not clear whether these graph-based representations improve downstream task performance

Research questions

1. Do the enhancements improve downstream results?

2. How do the representations compare to other graph-based representations?

3. What is the best way of parsing to these representations?

Research questions

4. Is UD as good a representation for downstream tasks as SD?

5. Does higher parsing accuracy translate to better downstream performance?

Our setup

8 different representations

2 parsers and parsing strategies

2 data sets

➡ 23 runs

The representations

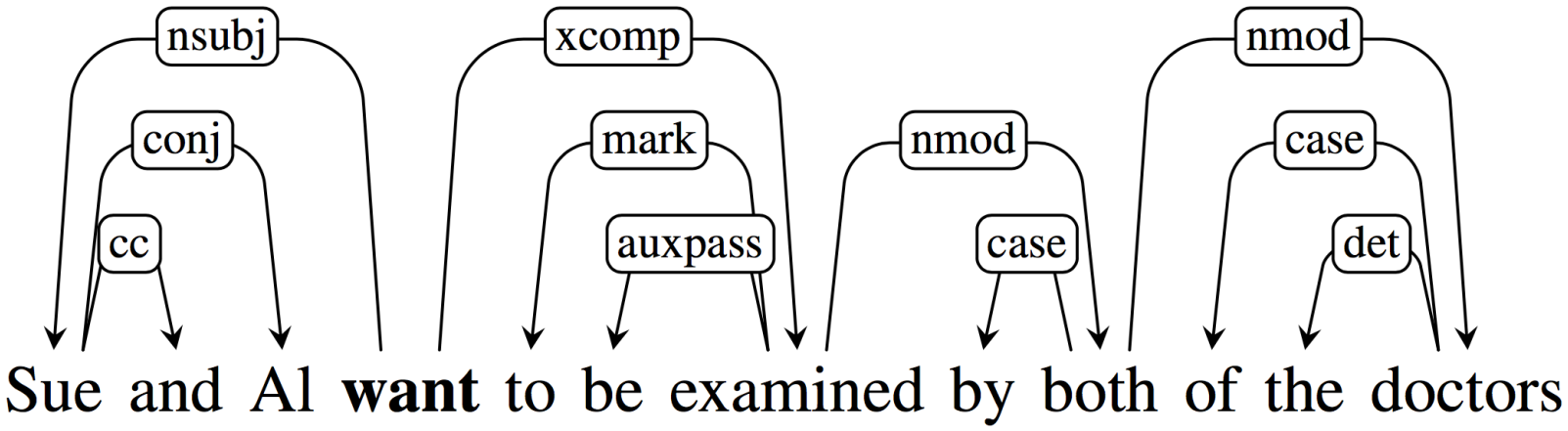
- 5 representations derived from Universal Dependencies:
  - UD basic
  - UD enhanced
  - UD enhanced++ (w/o empty nodes)
  - UD enhanced++diathesis
  - UD enhanced++diathesis --

The representations

- Stanford Dependencies basic
- DM
- Predicate Argument Structure (PAS)

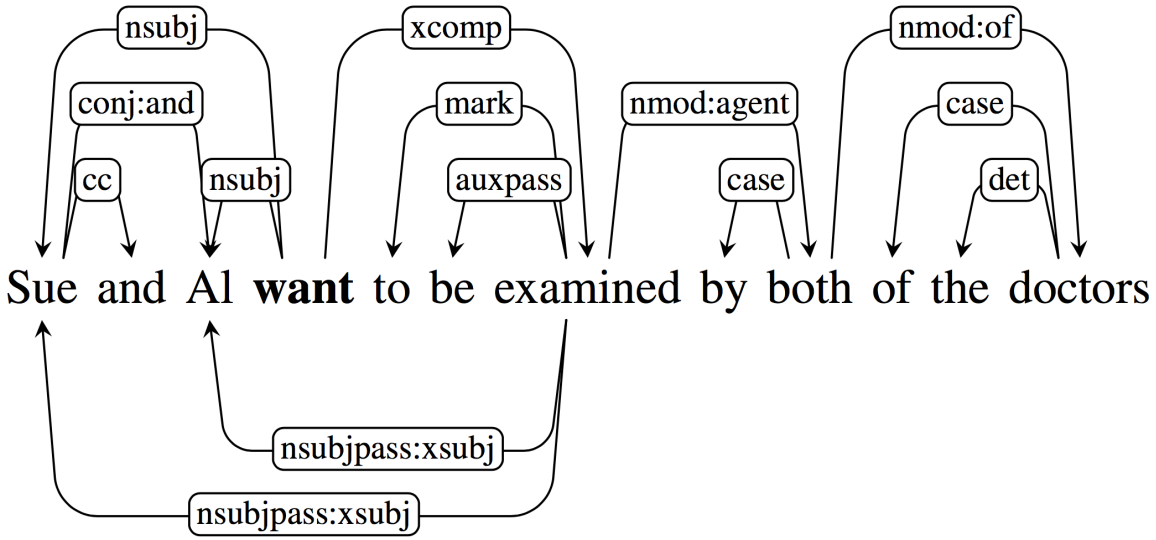
UD basic

- A dependency tree representation that
  - aims to allow **cross-linguistically consistent treebank** annotations
  - contains dependencies between **content words**



UD enhanced

- A graph-based dependency representation that
  - contains **additional edges** for phenomena such as control, raising, and coordination
  - **augments relation labels** with function words

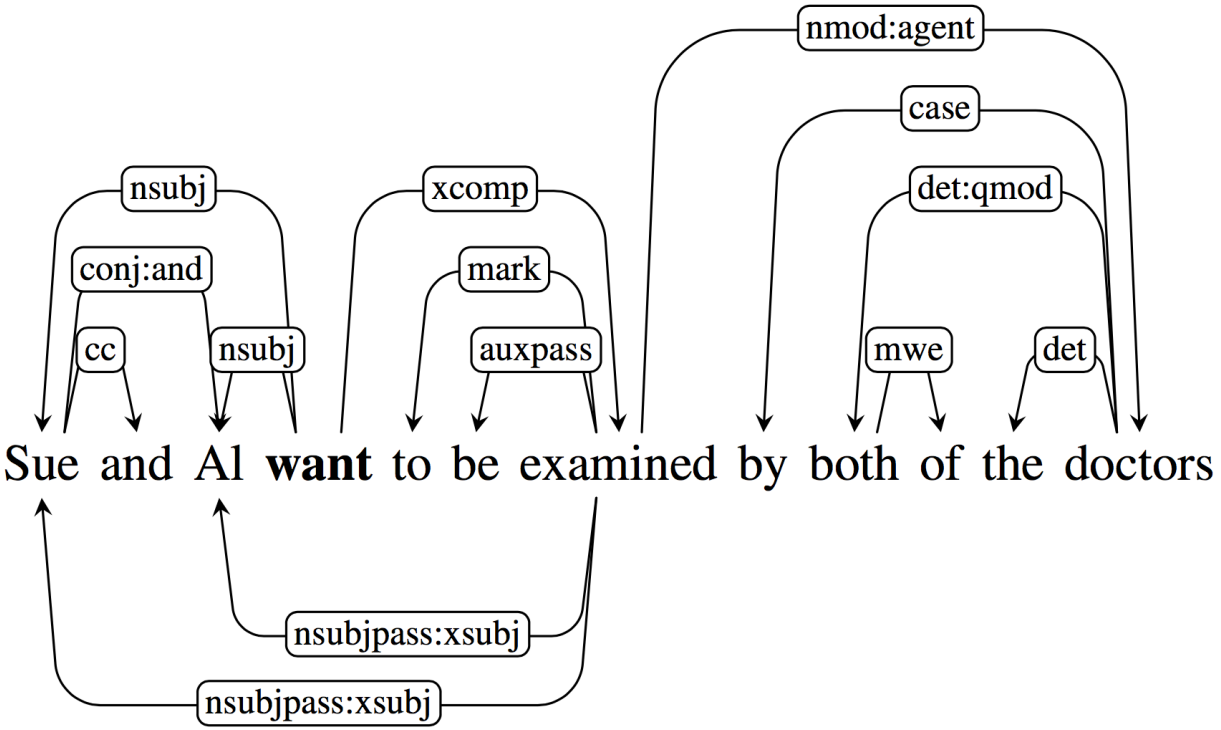


UD enhanced++

- A graph-based dependency representation that
  - is based on UD enhanced
  - **modifies the structure** such that there are more relations between content words

UD enhanced++

- A graph-based dependency representation that
  - is based on UD enhanced

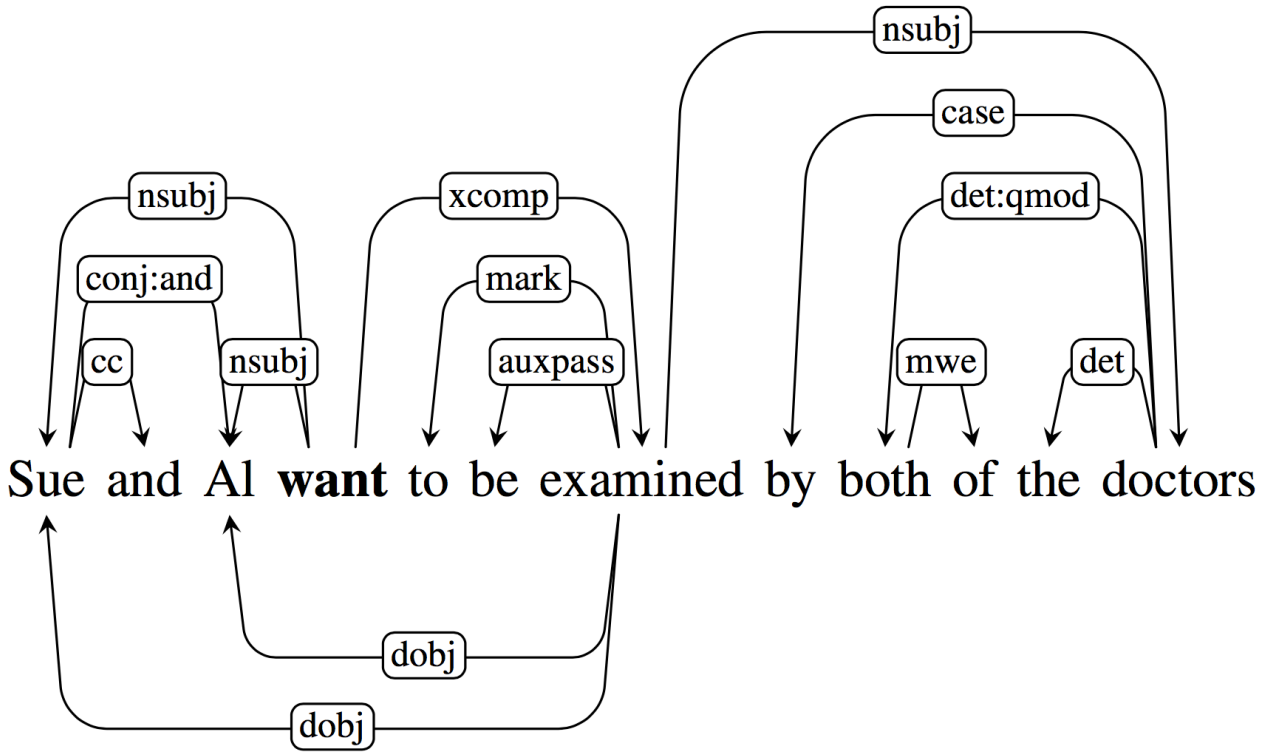


UD enhanced++ diathesis

- A graph-based dependency representation that
  - is based on UD enhanced++
  - **Neutralizes** some syntactic **alternations**
  - Introduces dependencies for **other forms of control**

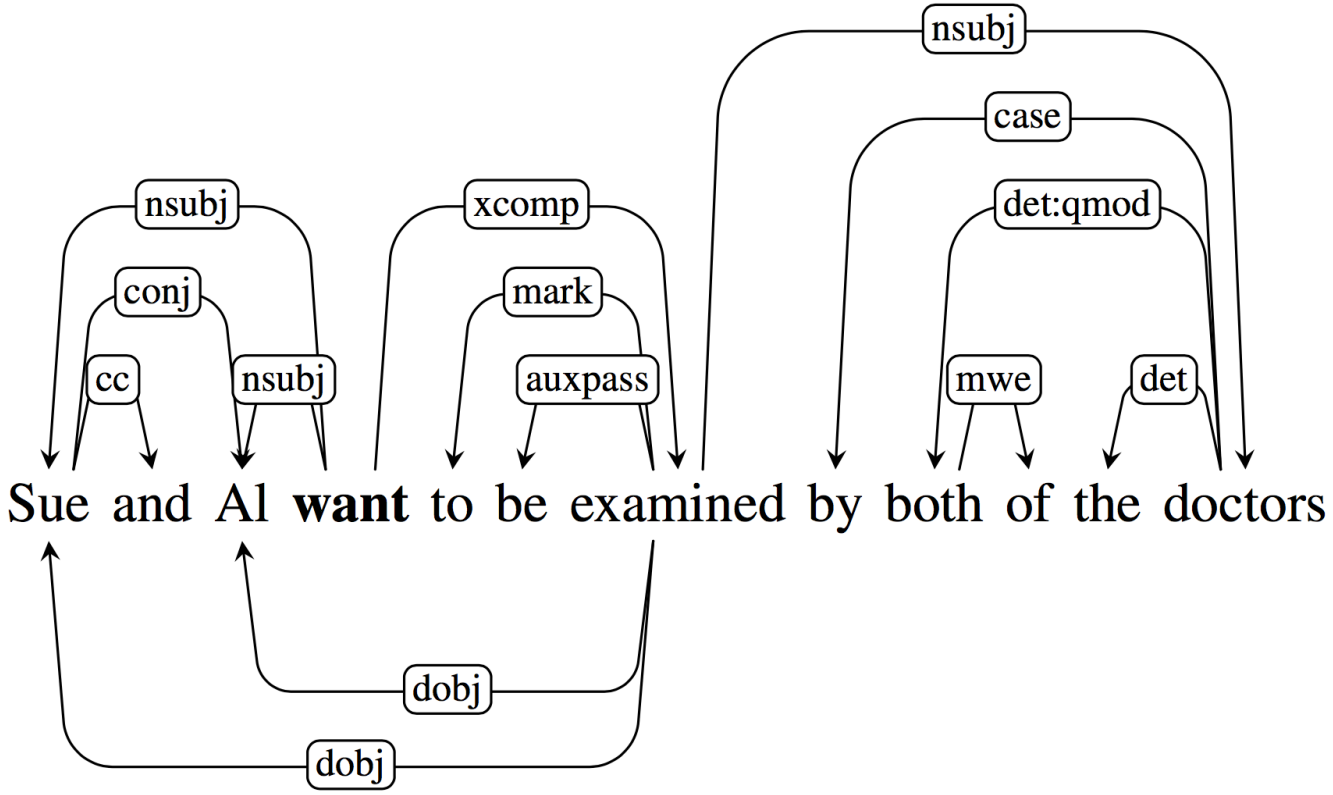
UD enhanced++ diathesis

- A graph-based dependency representation that
  - is based on UD enhanced++



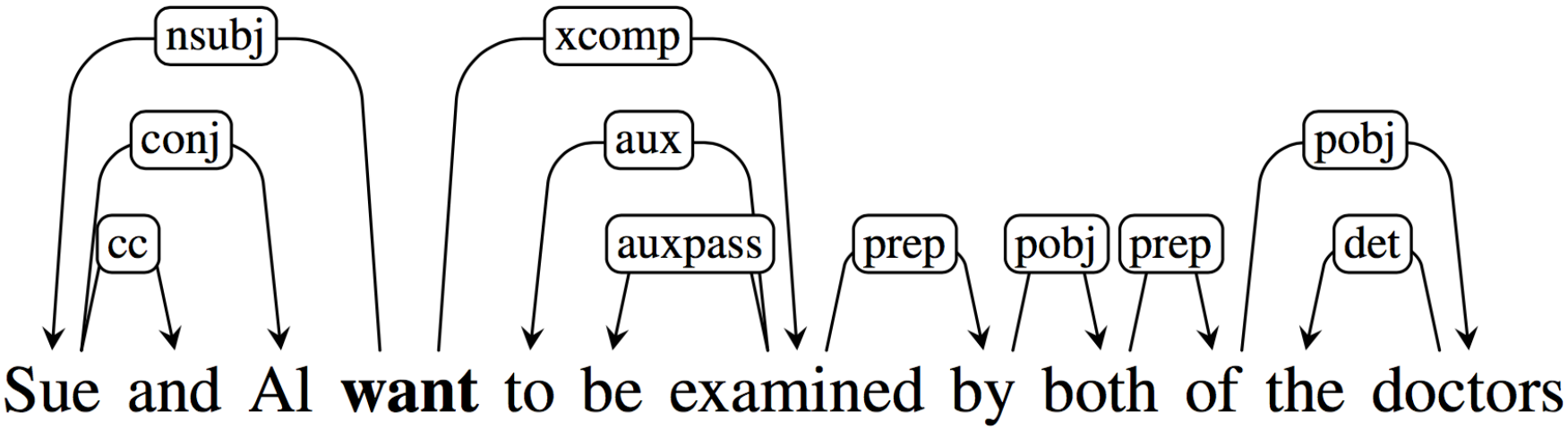
UD enhanced++ diathesis --

- Does not use augmented relation labels



Stanford Dependencies

- A dependency tree representation that
  - is less content-word centric than UD



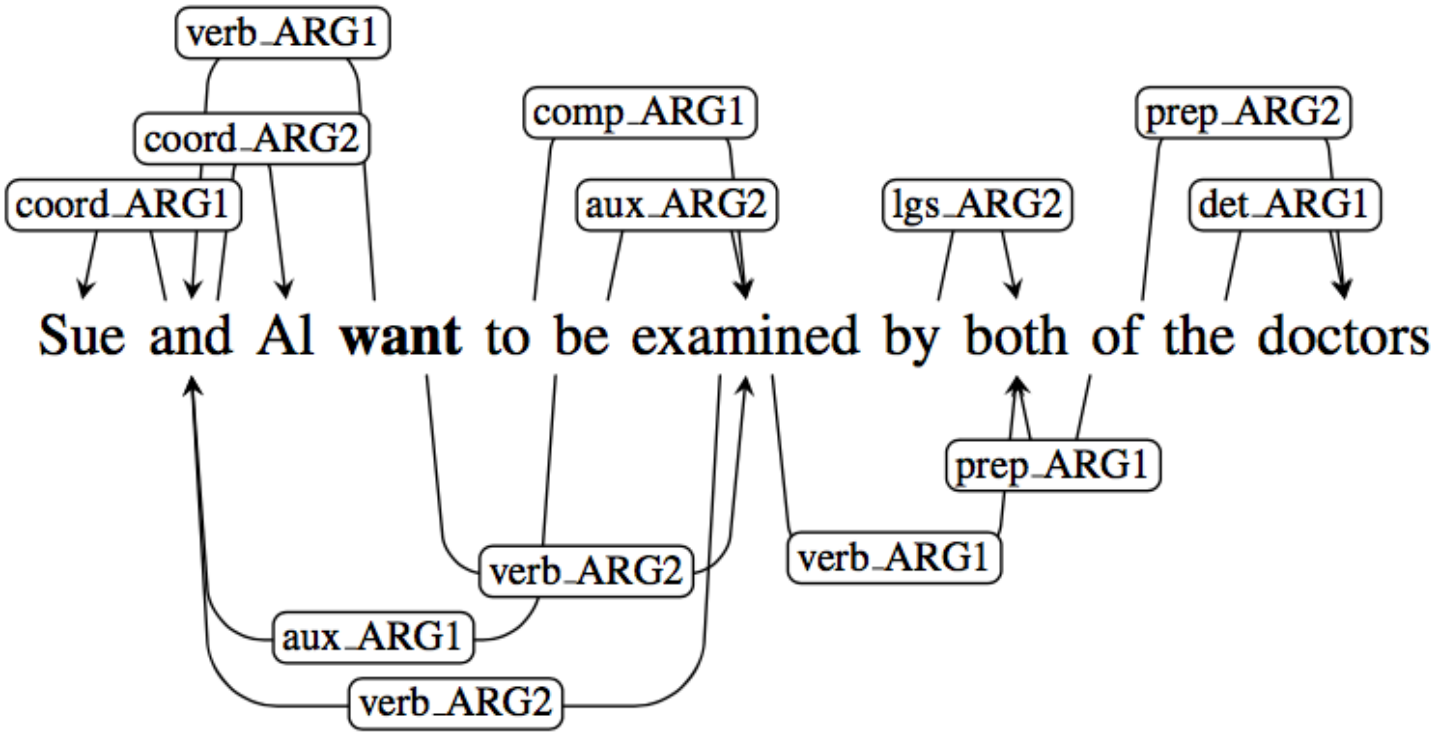
Predicate Argument Structure (PAS)

- A graph-based representation derived from an **automatic HPSG-style** re-annotation of the **Penn Treebank**

- Relation names encode the **index of the arguments** and the **POS tag of the head**

Predicate Argument Structure (PAS)

- A graph-based representation derived from an **automatic HPSG-style** re-annotation of the **Penn Treebank**

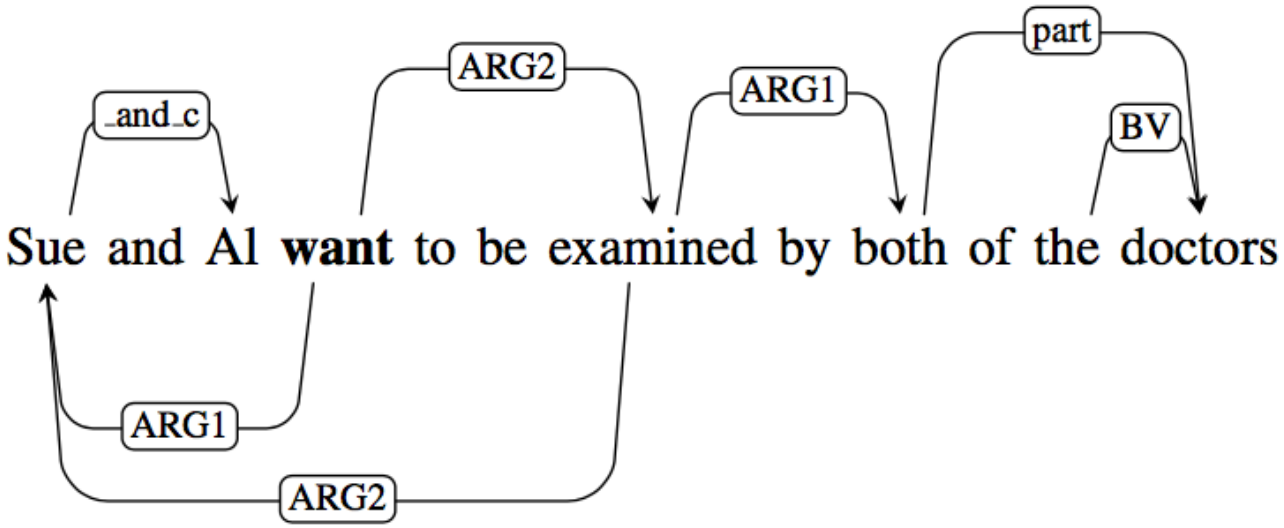


DM

- A graph-based representation derived from the **DeepBank** HPSG annotations
- Most dependency labels **encode the index** of the argument
- Special relations for some phenomena such as **bound variables**, **coordination**, and **partitives**

DM

- A graph-based representation derived from the **DeepBank** HPSG annotations
- Most dependency labels **encode the index** of the argument



Parsing strategies

- **Directly parsing to graphs** with the dyalog-SRNN parser (Ribeyre et al., 2013; de la Clergerie et al., 2017)

- **Parsing to dependency** trees with the Dozat and Manning (2017) parser and applying **rule-based augmentations**

Data: DM Split

- WSJ data from SemEval 2014 Semantic Dependency Parsing Shared Task
- PAS and DM data from SDP Shared Task
- UD and SD representations converted from PTB constituency trees

Data: Full

- WSJ + Brown + GENIA
- not available for DM and PAS
- UD and SD representations converted from PTB constituency trees

Overview of our runs

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	yes	yes	yes	yes	yes	no	yes	yes
	FULL	yes	yes	yes	yes	yes	no	no	no
Dep parser + conv.	DM	yes	yes	yes	yes	yes	no	no	no
	FULL	yes	yes	yes	yes	yes	yes	no	no

Research questions

1. **Do the enhancements improve downstream results?**
2. How do the representations compare to other graph-based representations?
3. What is the best way of parsing to these representations?
4. Is UD as good a representation for downstream tasks as SD?
5. Does higher parsing accuracy translate to better downstream performance?

Graph > surface syntax representations?

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	yes	yes	yes	yes	yes	no	yes	yes
	FULL	yes	yes	yes	yes	yes	no	no	no
Dep parser + conv.	DM	yes	yes	yes	yes	yes	no	no	no
	FULL	yes	yes	yes	yes	yes	yes	no	no

Graph > surface syntax representations?

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	2	1	4	3	5	no	yes	yes
	FULL	3	1	2	5	4	no	no	no
Dep parser + conv.	DM	4	2	1	3	5	no	no	no
	FULL	5	1	3	2	4	yes	no	no

Graph > surface syntax representations?

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	**-0.1**	**56.44**	**-1.06**	**-0.26**	**-1.19**	no	yes	yes
	FULL	**-0.55**	**56.81**	**-0.42**	**-1.95**	**-1.11**	no	no	no
Dep parser + conv.	DM	**-0.74**	**-0.51**	**59.08**	**-0.66**	**-1.06**	no	no	no
	FULL	**-0.97**	**60.51**	**-0.91**	**-0.64**	**-0.95**	yes	no	no

Graph > surface syntax representations?

- **UD enhanced**, on average, consistently lead to **better** downstream results than **UD basic**
- **UD enhanced++** and **enhanced++ diathesis** also good representations for downstream tasks, but higher variance

Task-specific findings: Event extraction and opinion analysis

- Representations that **worked well**:
  - UD enhanced
  - UD enhanced++
  - UD enhanced++ diathesis
- Representations that **worked less well**:
  - basic UD
  - UD diathesis --
- **Augmented relation labels** seem to be **useful** for this task!

Task-specific findings: Negation scope resolution

- Representations that **worked well**
  - enhanced UD
- Much more variance in results
- Augmented relation labels don’t seem to add anything

Research questions

1. Do the enhancements improve downstream results?
2. **How do the representations compare to other graph-based representations?**
3. What is the best way of parsing to these representations?
4. Is UD as good a representation for downstream tasks as SD?
5. Does higher parsing accuracy translate to better downstream performance?

UD representations > other graph representations?

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	yes	yes	yes	yes	yes	no	yes	yes
	FULL	yes	yes	yes	yes	yes	no	no	no
Dep parser + conv.	DM	yes	yes	yes	yes	yes	no	no	no
	FULL	yes	yes	yes	yes	yes	yes	no	no

UD representations > other graph representations?

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	**2**	**1**	**4**	**3**	**5**	no	**6**	**7**
	FULL	yes	yes	yes	yes	yes	no	no	no
Dep parser + conv.	DM	yes	yes	yes	yes	yes	no	no	no
	FULL	yes	yes	yes	yes	yes	yes	no	no

UD representations > other graph representations?

- No evidence that DM/PAS are better representations for downstream tasks than more surface-syntax aligned UD representations
- Especially true for event extraction and opinion analysis tasks
  - Suggests again that **rich label sets** are **important** for these tasks
- Gap widens much more if one uses more data, which is not available for DM and PAS!

Research questions

1. Do the enhancements improve downstream results?
2. How do the representations compare to other graph-based representations?
3. **What is the best way of parsing to these representations?**
4. Is UD as good a representation for downstream tasks as SD?
5. Does higher parsing accuracy translate to better downstream performance?

Parsing method

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	yes	yes	yes	yes	yes	no	yes	yes
	FULL	yes	yes	yes	yes	yes	no	no	no
Dep parser + conv.	DM	yes	yes	yes	yes	yes	no	no	no
	FULL	yes	yes	yes	yes	yes	yes	no	no

Parsing method

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	yes	yes	yes	yes	yes	no	yes	yes
	FULL	yes	yes	yes	yes	yes	no	no	no
Dep parser + conv.	DM	yes	yes	yes	yes	yes	no	no	no
	FULL	yes	yes	yes	yes	yes	yes	no	no

Parsing method

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	**2**	**2**	**2**	**2**	**2**	no	yes	yes
	FULL	yes	yes	yes	yes	yes	no	no	no
Dep parser + conv.	DM	**1**	**1**	**1**	**1**	**1**	no	no	no
	FULL	yes	yes	yes	yes	yes	yes	no	no

Parsing method

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	yes	yes	yes	yes	yes	no	yes	yes
	FULL	2	2	2	2	2	no	no	no
Dep parser + conv.	DM	yes	yes	yes	yes	yes	no	no	no
	FULL	1	1	1	1	1	yes	no	no

Parsing method

- **Two-step parsing consistently outperformed** direct graph parser
- In particular true for negation scope task (up 8 points difference)
- Very **small difference** for **event extraction** and small difference for **opinion analysis tasks**

Research questions

1. Do the enhancements improve downstream results?
2. How do the representations compare to other graph-based representations?
3. What is the best way of parsing to these representations?
4. **Is UD as good a representation for downstream tasks as SD?**
5. Does higher parsing accuracy translate to better downstream performance?

SD vs. UD

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	yes	yes	yes	yes	yes	no	yes	yes
	FULL	yes	yes	yes	yes	yes	no	no	no
Dep parser + conv.	DM	yes	yes	yes	yes	yes	no	no	no
	FULL	yes	yes	yes	yes	yes	yes	no	no

SD vs. UD

		UD basic	UD enh.	UD enh.++	UD enh.++ diat	UD enh.++ diat --	SD basic	DM	PAS
Graph parser	DM	yes	yes	yes	yes	yes	no	yes	yes
	FULL	yes	yes	yes	yes	yes	no	no	no
Dep parser + conv.	DM	yes	yes	yes	yes	yes	no	no	no
	FULL	**59.5**	yes	yes	yes	yes	**59.7**	no	no

SD vs. UD

- Both seem on average **similarly good representations** for downstream tasks
- SD slightly better for event extraction, UD better for opinion analysis
- No evidence that striving for cross-linguistic consistency hurts downstream performance

Research questions

1. Do the enhancements improve downstream results?
2. How do the representations compare to other graph-based representations?
3. What is the best way of parsing to these representations?
4. Is UD as good a representation for downstream tasks as SD?
5. **Does higher parsing accuracy translate to better downstream performance?**

Correlation between parsing and downstream performance

	LAS	UAS	Task F1
Graph parser	88.99	90.43	56.26
Dep. parser	91.13 (+ 2.14)	93.26 (+ 2.83)	59.54 (+ 3.28)

Conclusions

- **Adding explicit dependency relations** for long distance dependencies and **augmenting relation labels** seems to be useful for downstream tasks
- No evidence that representations that explicitly encode predicate-argument structures are better than representations derived from surface syntax trees
- **Two-step parsing** (currently) seems to be the best parsing approach
- **UD as good a representation** as SD for downstream tasks

Sponsored slide

- The UD representations seem to be good representations for downstream tasks because
  - they have **expressive labels**
  - **high-performing parsers** and **accurate converters** exist
  - **lots of data** can be obtained through conversion
  - enhanced variants **recover predicate-argument** structures in many cases

Thank you!