MRP 2020:

Cross-Framework and Cross-Lingual Meaning Representation Parsing

http://mrp.nlpl.eu

Stephan Oepen♣, Omri Abend♠, Lasha Abzianidze♡, Johan Bos♢,
Jan Hajič○, **Daniel Hershcovich**⋆, Bin Li•, Tim O'Gorman⋄,
Nianwen Xue∗, and Daniel Zeman○

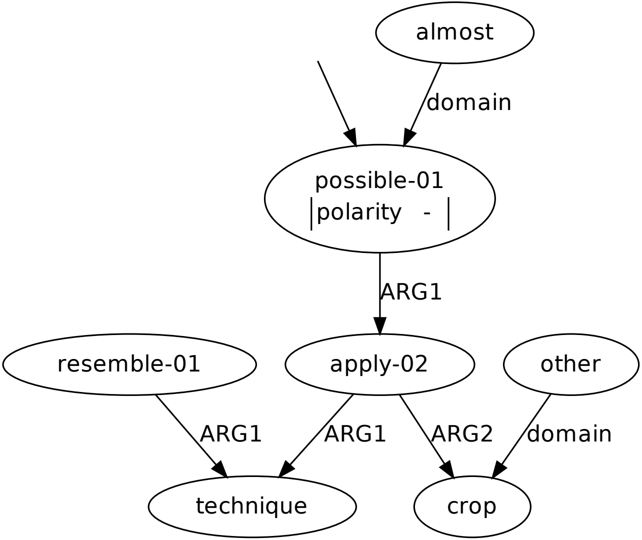
♣ University of Oslo, Department of Informatics
♠ The Hebrew University of Jerusalem, School of Computer Science and Engineering
♡ Utrecht University, Logic, Language, and Information
♢ University of Groningen, Center for Language and Cognition
○ Charles University, Prague, Institute of Formal and Applied Linguistics
⋆ University of Copenhagen, Department of Computer Science
• Nanjing Normal University, School of Chinese Language and Literature
⋄ University of Massachusetts at Amherst, College of Information and Computer Sciences
∗ Brandeis University, Department of Computer Science

mrp-organizers@nlpl.eu

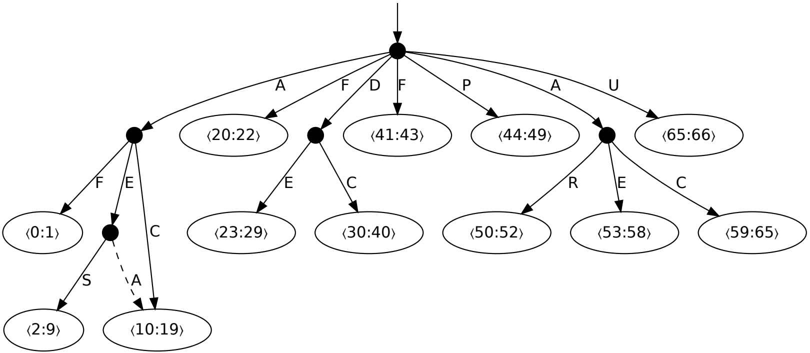
10,000-Meter Perspective: Parsing into Semantic Graphs

A similar technique is almost impossible to apply to other crops.

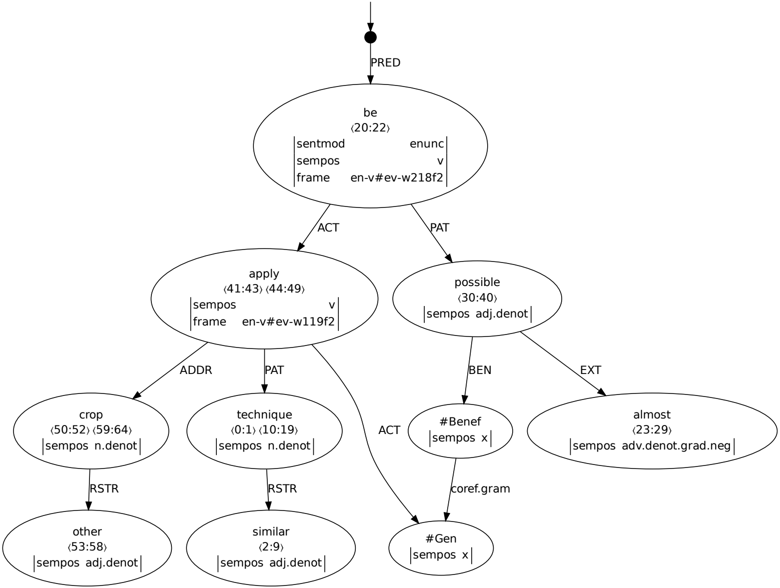
10,000-Meter Perspective: Parsing into Semantic Graphs



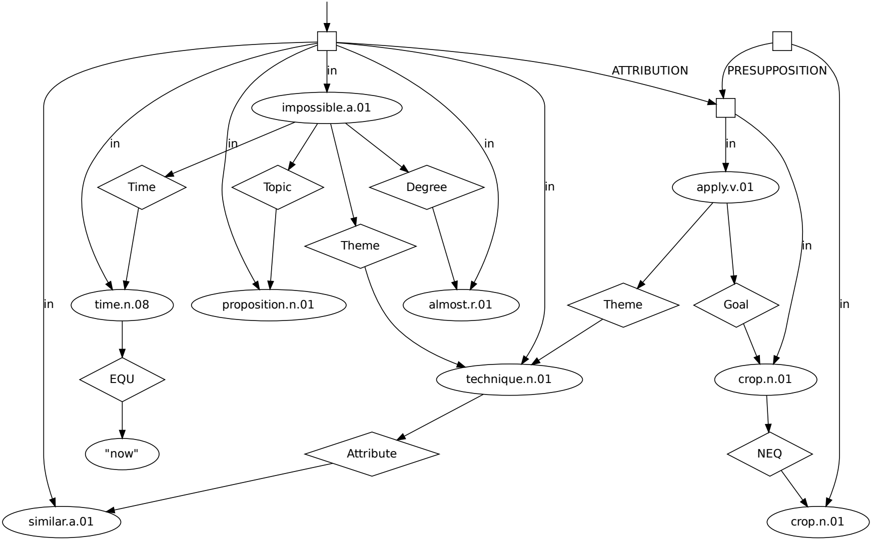
10,000-Meter Perspective: Parsing into Semantic Graphs



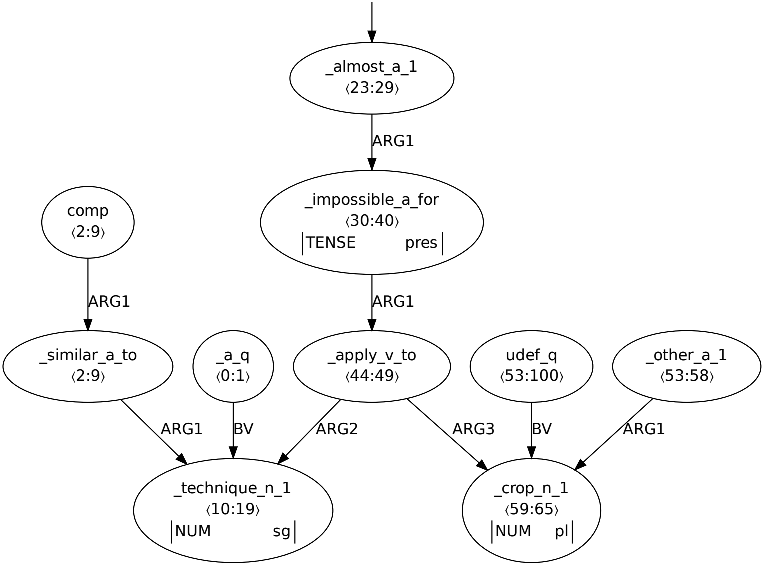
10,000-Meter Perspective: Parsing into Semantic Graphs



10,000-Meter Perspective: Parsing into Semantic Graphs



10,000-Meter Perspective: Parsing into Semantic Graphs

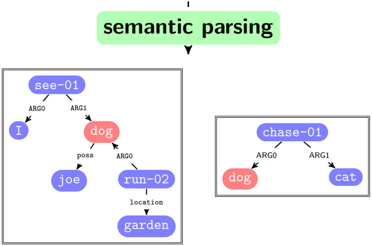


Why Graph-Based Meaning Representation?

I saw Joe's dog, which was running in the garden.
The dog was chasing a cat.

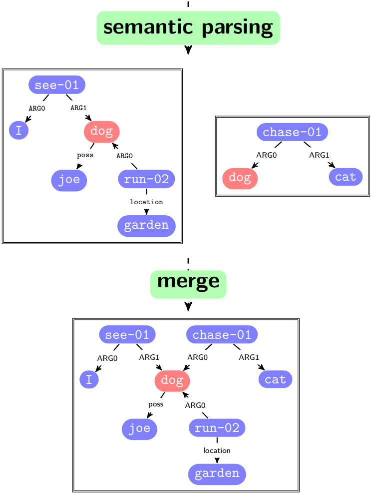
Why Graph-Based Meaning Representation?

I saw Joe's dog, which was running in the garden.
The dog was chasing a cat.



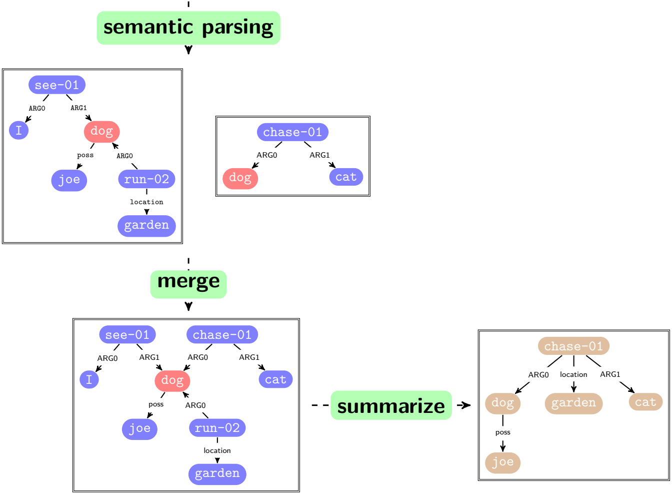
Why Graph-Based Meaning Representation?

I saw Joe's dog, which was running in the garden.
The dog was chasing a cat.



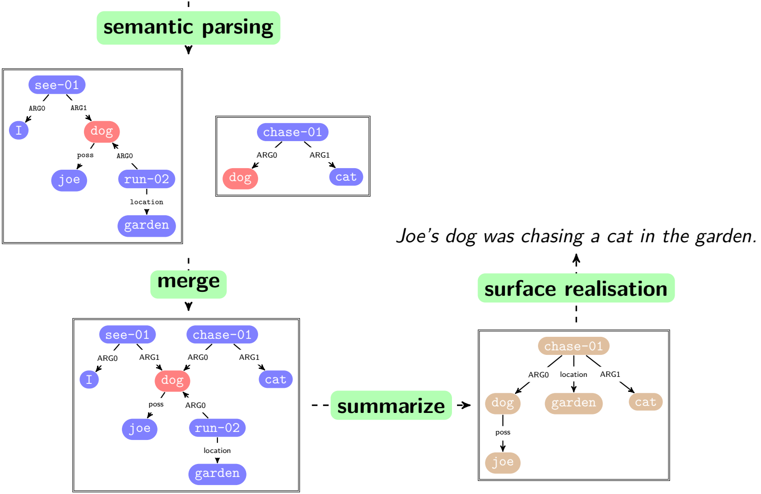
Why Graph-Based Meaning Representation?

I saw Joe's dog, which was running in the garden.
The dog was chasing a cat.

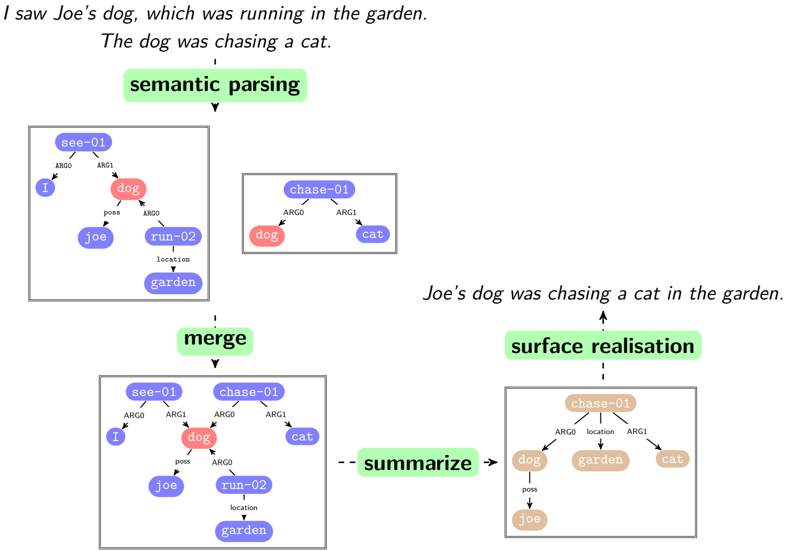


Why Graph-Based Meaning Representation?

I saw Joe's dog, which was running in the garden.
The dog was chasing a cat.

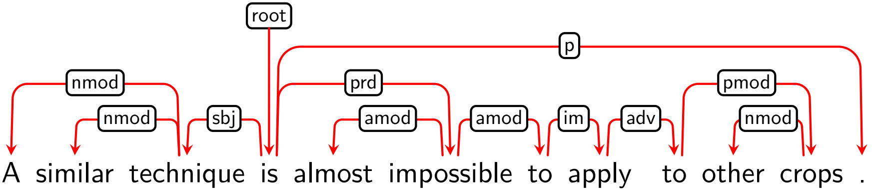


Why Graph-Based Meaning Representation?

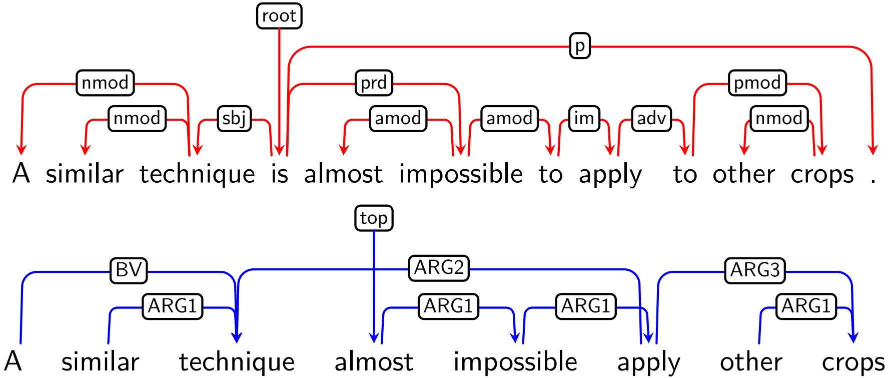


Hardy & Vlachos (2018): 2^+ ROUGE points over strong encoder–decoder.

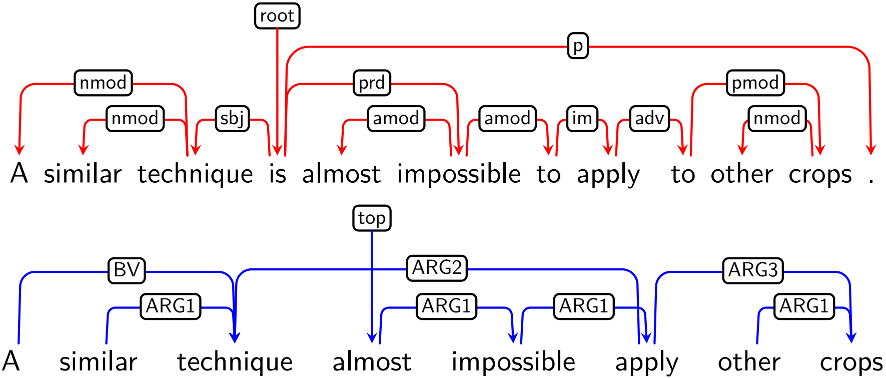
Syntactic Trees vs. Semantic Graphs



Syntactic Trees vs. Semantic Graphs



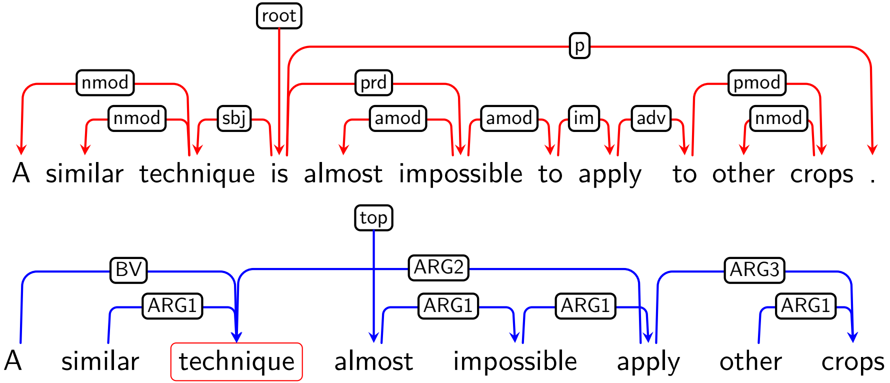
Syntactic Trees vs. Semantic Graphs



$$\exists x : \text{technique}'(x) \land \text{similar}'(x); \exists y : \text{crop}'(y) \land \text{other}'(y)$$

$$\rightarrow \text{almost}'(\neg\text{possible}'(\text{apply}'(e, x, y)))$$

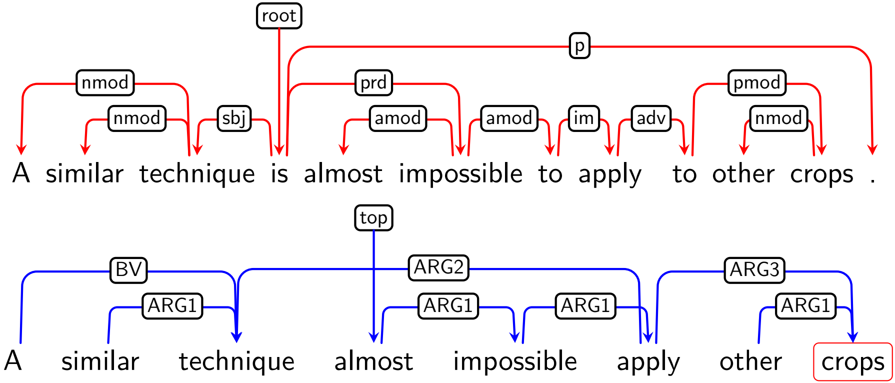
Syntactic Trees vs. Semantic Graphs



$$\exists {\color{red}x} : \text{technique}'({\color{red}x}) \wedge \text{similar}'({\color{red}x}); \exists y : \text{crop}'(y) \wedge \text{other}'(y)$$

$$\rightarrow \text{almost}'(\neg\text{possible}'(\text{apply}'(e, {\color{red}x}, y)))$$

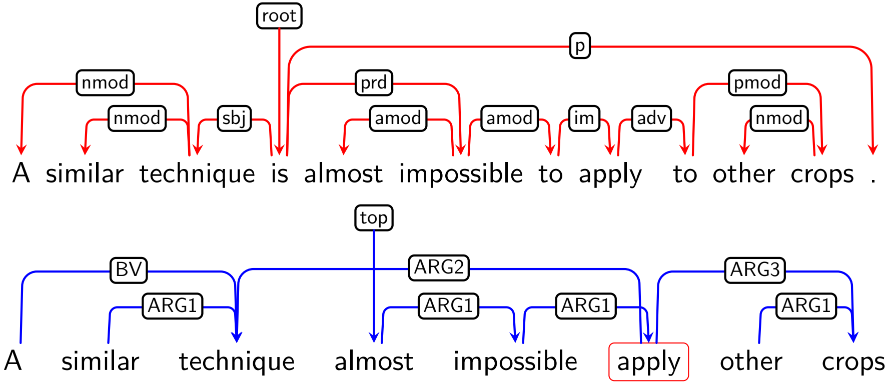
Syntactic Trees vs. Semantic Graphs



$$\exists x : \text{technique'}(x) \wedge \text{similar'}(x); \exists {\color{red}y} : \text{crop'}({\color{red}y}) \wedge \text{other'}({\color{red}y})$$

$$\rightarrow \text{almost'}(\neg\text{possible'}(\text{apply'}(e, x, {\color{red}y})))$$

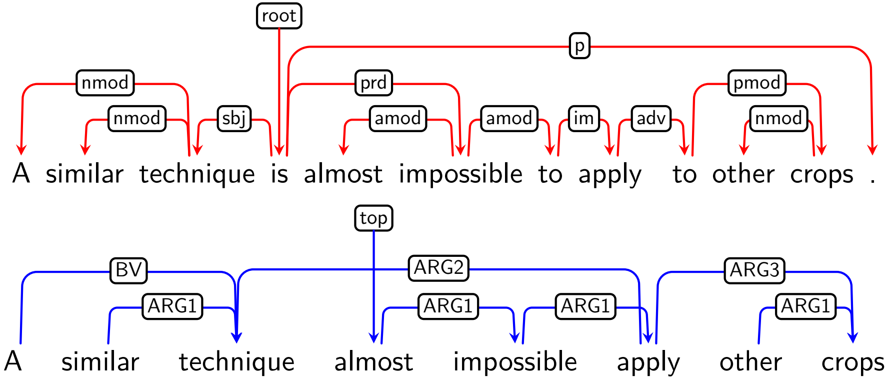
Syntactic Trees vs. Semantic Graphs



$$\exists x : \text{technique'}(x) \land \text{similar'}(x); \exists y : \text{crop'}(y) \land \text{other'}(y)$$

$$\rightarrow \text{almost'}(\neg\text{possible'}(\text{apply'}(e, x, y)))$$

Syntactic Trees vs. Semantic Graphs



$$\exists x : \text{technique'}(x) \wedge \text{similar'}(x); \exists y : \text{crop'}(y) \wedge \text{other'}(y)$$

$$\rightarrow \text{almost'}(\neg\text{possible'}(\text{apply'}(e, x, y)))$$

Different Desiderata and Levels of Abstraction

- Grammaticality (e.g. subject–verb agreement) vs. relational structure.

Structural Wellformedness Conditions on Trees

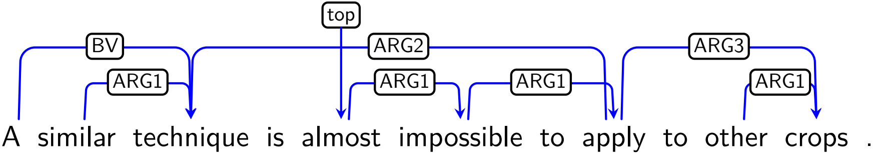
- Unique root, connected, single parent, free of cycles; maybe projective;

$\rightarrow$ all nodes (but the root) reachable by unique directed path from root.

Structural Wellformedness Conditions on Trees

- Unique root, connected, single parent, free of cycles; maybe projective;

$\rightarrow$ all nodes (but the root) reachable by unique directed path from root.

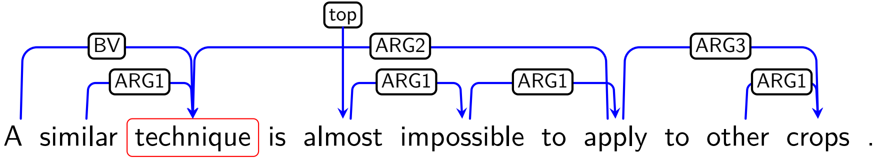


Semi-Formally: Trees vs. Graphs

Structural Wellformedness Conditions on Trees

- Unique root, connected, single parent, free of cycles; maybe projective;

$\rightarrow$ all nodes (but the root) reachable by unique directed path from root.



Beyond Trees: General Graphs

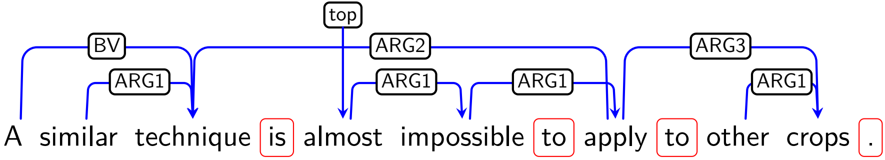
- Argument sharing: nodes with multiple incoming edges (***in***-degree > 1);

Semi-Formally: Trees vs. Graphs

Structural Wellformedness Conditions on Trees

- Unique root, connected, single parent, free of cycles; maybe projective;

$\rightarrow$ all nodes (but the root) reachable by unique directed path from root.



Beyond Trees: General Graphs

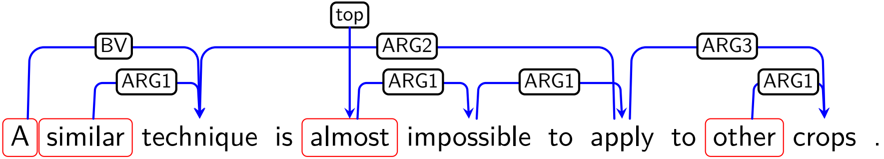
- Argument sharing: nodes with multiple incoming edges (***in***-degree > 1);
- some surface tokens do not contribute (as nodes; many function words);

Semi-Formally: Trees vs. Graphs

Structural Wellformedness Conditions on Trees

- Unique root, connected, single parent, free of cycles; maybe projective;

$\rightarrow$ all nodes (but the root) reachable by unique directed path from root.



Beyond Trees: General Graphs

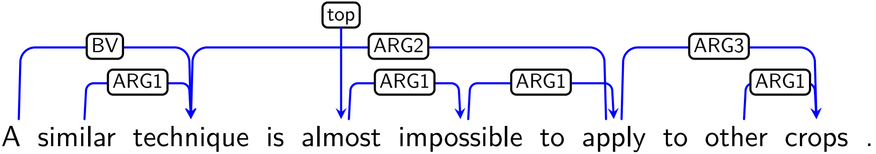
- Argument sharing: nodes with multiple incoming edges (***in***-degree > 1);
- some surface tokens do not contribute (as nodes; many function words);
- (structurally) multi-rooted: more than one node with zero ***in***-degree;

Semi-Formally: Trees vs. Graphs

Structural Wellformedness Conditions on Trees

- Unique root, connected, single parent, free of cycles; maybe projective;

$\rightarrow$ all nodes (but the root) reachable by unique directed path from root.



Beyond Trees: General Graphs

- Argument sharing: nodes with multiple incoming edges (***in***-degree > 1);
- some surface tokens do not contribute (as nodes; many function words);
- (structurally) multi-rooted: more than one node with zero ***in***-degree;

$\rightarrow$ massive growth in modeling and algorithmic complexity (NP-complete).

High-Level Goals of the Shared Task

Cross-Framework Comparability and Interoperability

- Vast, complex landscape of representing natural language meaning;
- diverse linguistic traditions, modeling assumptions, levels of ambition;

→ clarify concepts and terminology; unify representations and evaluation.

High-Level Goals of the Shared Task

Cross-Framework Comparability and Interoperability

- Vast, complex landscape of representing natural language meaning;
- diverse linguistic traditions, modeling assumptions, levels of ambition;

→ clarify concepts and terminology; unify representations and evaluation.

Parsing into Graph-Structured Representations

- Cottage industry of parsers with output structures beyond rooted trees;
- distinct techniques, e.g. based on transitions, composition, 'translation';

→ evaluate across frameworks; learning from complementary knowledge.

High-Level Goals of the Shared Task

Cross-Framework Comparability and Interoperability

- Vast, complex landscape of representing natural language meaning;
- diverse linguistic traditions, modeling assumptions, levels of ambition;

→ clarify concepts and terminology; unify representations and evaluation.

Parsing into Graph-Structured Representations

- Cottage industry of parsers with output structures beyond rooted trees;
- distinct techniques, e.g. based on transitions, composition, 'translation';

→ evaluate across frameworks; learning from complementary knowledge.

Two Distinct Tracks in MRP 2020

- Cross-Framework Perspective: Seek commonality and complementarity.
- Cross-Lingual Perspective: In-framework transfer to another language.

Graph Structure vs. Node (or Edge) Decorations

Zero-Arity Predicates vs. Constants

- Nodes and edges can be labeled (e.g. by relation and role identifiers);

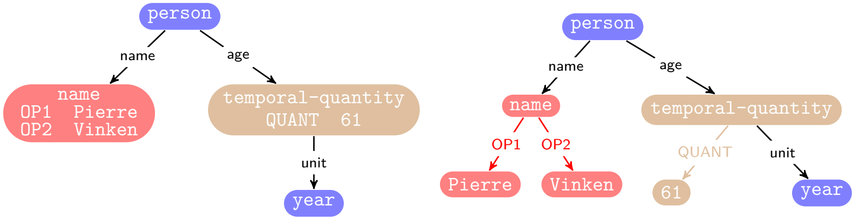
Graph Structure vs. Node (or Edge) Decorations

Zero-Arity Predicates vs. Constants

- Nodes and edges can be labeled (e.g. by relation and role identifiers);
- labels can be internally structured: node properties and edge attributes;
- properties (and attributes) are non-recursive attribute–value matrices;
- node (and edge) label is merely a distinguished property (or attribute);

Zero-Arity Predicates vs. Constants

- Nodes and edges can be labeled (e.g. by relation and role identifiers);
- labels can be internally structured: node properties and edge attributes;
- properties (and attributes) are non-recursive attribute–value matrices;
- node (and edge) label is merely a distinguished property (or attribute);
- distinction is not commonly discussed, but used by many frameworks.



Pierre Vinken is 61 years old.

Anchoring in the Surface String

Relating Pieces of Meaning to the Linguistic Signal

- Intuitively, sub-structures of meaning relate to sub-parts of the input;
- semantic frameworks vary in how much weight to put on this relation;

Anchoring in the Surface String

Relating Pieces of Meaning to the Linguistic Signal

- Intuitively, sub-structures of meaning relate to sub-parts of the input;
- semantic frameworks vary in how much weight to put on this relation;
- anchoring of graph elements in sub-strings of the underlying utterance;
- can be part of semantic annotations or not; can take different forms;

Anchoring in the Surface String

Relating Pieces of Meaning to the Linguistic Signal

- Intuitively, sub-structures of meaning relate to sub-parts of the input;
- semantic frameworks vary in how much weight to put on this relation;
- anchoring of graph elements in sub-strings of the underlying utterance;
- can be part of semantic annotations or not; can take different forms;
- hierarchy of anchoring types: Flavor (0)–(2); bilexical graphs strictest;

	Name	Example	Type of Anchoring
(0)	bilexical	~~DM, PSD~~	nodes are sub-set of surface tokens
(1)	anchored	EDS, PTG, UCCA	free node–sub-string correspondences
(2)	unanchored	AMR, DRG	no explicit sub-string correspondences

Anchoring in the Surface String

Relating Pieces of Meaning to the Linguistic Signal

- Intuitively, sub-structures of meaning relate to sub-parts of the input;
- semantic frameworks vary in how much weight to put on this relation;
- anchoring of graph elements in sub-strings of the underlying utterance;
- can be part of semantic annotations or not; can take different forms;
- hierarchy of anchoring types: Flavor (0)–(2); bilexical graphs strictest;
- anchoring central in parsing, explicit or latent; aka 'alignment' for AMR;
- relevant to at least some downstream tasks; should impact evaluation.

	Name	Example	Type of Anchoring
(0)	bilexical	~~DM, PSD~~	nodes are sub-set of surface tokens
(1)	anchored	EDS, PTG, UCCA	free node–sub-string correspondences
(2)	unanchored	AMR, DRG	no explicit sub-string correspondences

A Selection of Semantic Graphbanks

Selection Criteria

- 'Full-sentence' semantics: all content-bearing units receive annotations;
- natively graph-based: meaning representation through (directed) graphs;
- large-scale, gold-standard annotations and parsers are publicly available;

A Selection of Semantic Graphbanks

Selection Criteria

- ‘Full-sentence’ semantics: all content-bearing units receive annotations;
- natively graph-based: meaning representation through (directed) graphs;
- large-scale, gold-standard annotations and parsers are publicly available;

→ five distinct frameworks, bi-lexical to unanchored; sadly, English only;

A Selection of Semantic Graphbanks

Selection Criteria

- ‘Full-sentence’ semantics: all content-bearing units receive annotations;
- natively graph-based: meaning representation through (directed) graphs;
- large-scale, gold-standard annotations and parsers are publicly available;

→ five distinct frameworks, bi-lexical to unanchored; sadly, English only;

A Selection of Semantic Graphbanks

Selection Criteria

- ‘Full-sentence’ semantics: all content-bearing units receive annotations;
- natively graph-based: meaning representation through (directed) graphs;
- large-scale, gold-standard annotations and parsers are publicly available;

→ five distinct frameworks, bi-lexical to unanchored; sadly, English only;

→ new in MRP 2020: one additional language for four of the frameworks.

A Selection of Semantic Graphbanks

Selection Criteria

- 'Full-sentence' semantics: all content-bearing units receive annotations;
- natively graph-based: meaning representation through (directed) graphs;
- large-scale, gold-standard annotations and parsers are publicly available;

→ five distinct frameworks, bi-lexical to unanchored; sadly, English only;

→ new in MRP 2020: one additional language for four of the frameworks.

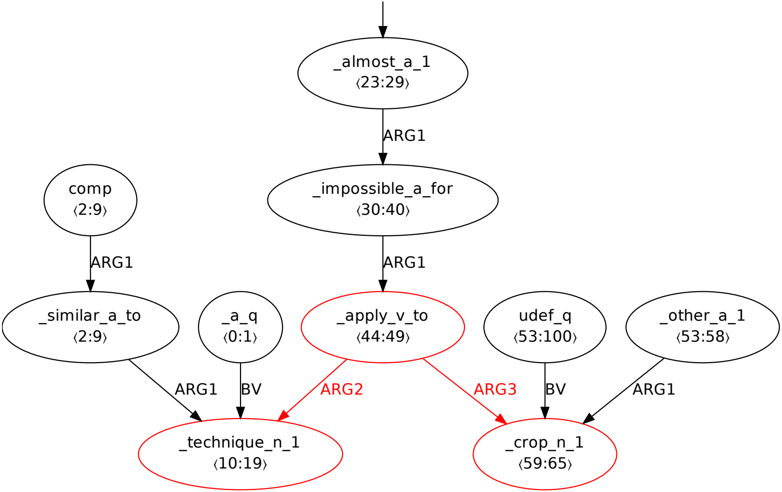
(With Apologies to) Non-Graph or Non-Meaning Banks

- PropBank (Palmer et al., 2005), Framenet (Baker et al., 1998), . . . ;
- Universal Decompositional Semantics (White et al., 2016);
- Enhanced Universal Dependencies (Schuster & Manning, 2016);
- . . .

(1) Elementary Dependency Structures (EDS)

Simplification of Underspecified Logical Forms (Oepen & Lønning, 2006)

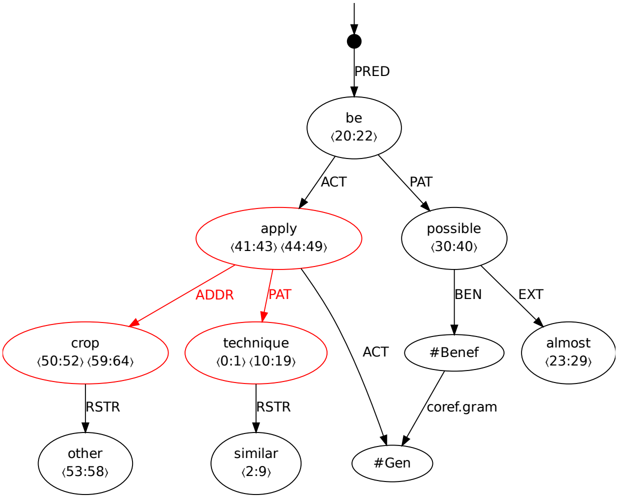
- Converted from LinGO Redwoods Treebank (Flickinger et al., 2017);
- decomposition or construction meaning; anchors: arbitrary sub-strings.



(1) Prague Tectogrammatical Graphs (PTG)

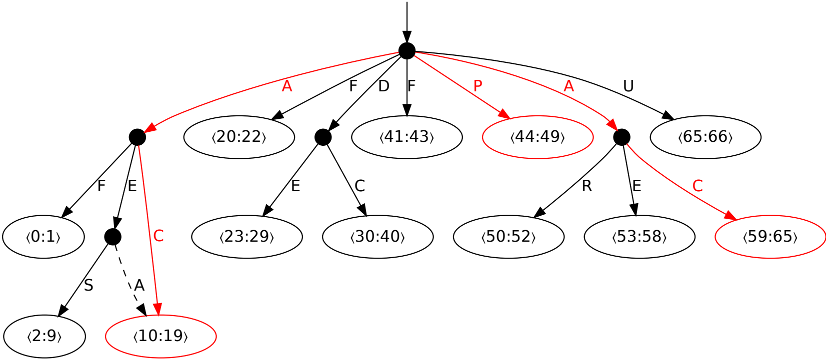
Simplification of FGD Tectogrammatical 'Trees' (Zeman & Hajič, 2020)

- Prague (Czech–English) Dependency Treebanks (Hajič et al., 2012);
- unanchored nodes for unexpressed arguments, e.g. `#Benef` and `#Gen`.



Multi-Layered Design (Abend & Rappoport, 2013); Foundational Layer

- Tree backbone: semantic 'constituents' are scenes ('clauses') and units;

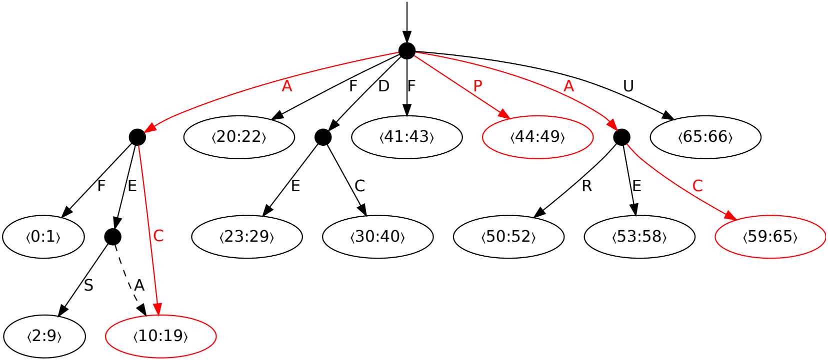


A similar technique is almost impossible to apply to other crops.

(1) Universal Conceptual Cognitive Annotation (UCCA)

Multi-Layered Design (Abend & Rappoport, 2013); Foundational Layer

- Tree backbone: semantic 'constituents' are scenes ('clauses') and units;
- scenes (Process or State): pArticipants and aDverbials (plus F and U);

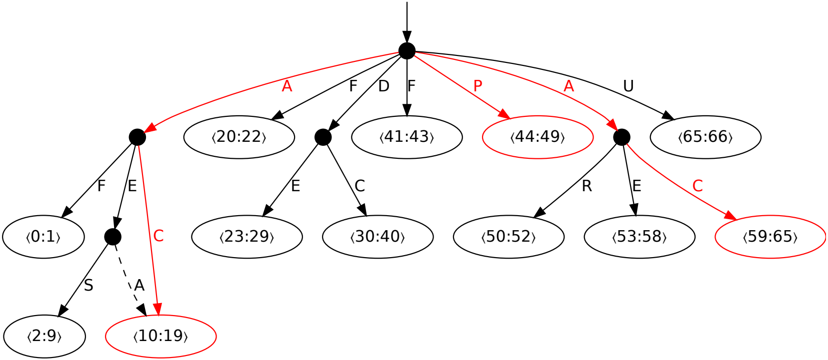


A similar technique is almost impossible to apply to other crops.

(1) Universal Conceptual Cognitive Annotation (UCCA)

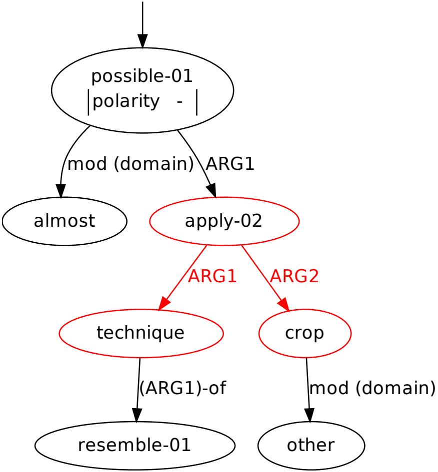
Multi-Layered Design (Abend & Rappoport, 2013); Foundational Layer

- Tree backbone: semantic 'constituents' are scenes ('clauses') and units;
- scenes (Process or State): pArticipants and aDverbials (plus F and U);
- complex units distinguish Center and Elaborator(s); allow remote edges.



A similar technique is almost impossible to apply to other crops.

(2) Abstract Meaning Representation (AMR)



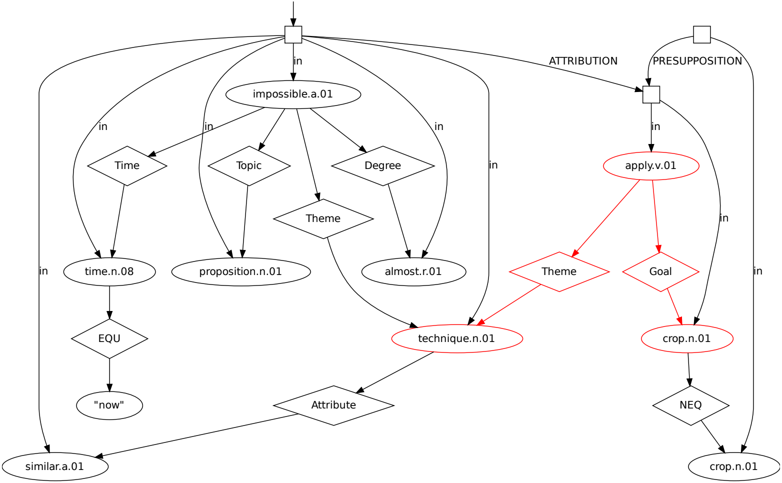
Banarescu et al. (2013)

- Abstractly (if not linguistically) similar to EDS, but unanchored;
- verbal senses from PropBank^{++};
- negation as node-local property;
- tree-like annotation: inversed edges normalized for evaluation;
- originally designed for (S)MT; various NLU applications to date.

A similar technique is almost impossible to apply to other crops.

Graph Encoding of DRS 'Nested Boxes' (Kamp & Reyle, 1993)

- From Groningen Parallel Meaning Bank (Abzianidze et al., 2017);
- explicit encoding of scope (boxes and `in` edges), using reified roles.



Cross-Framework Training and Evaluation Data

		EDS	PTG	UCCA	AMR	DRG
	Flavor	1	1	1	2	2
train	**Text Type**	newspaper	newspaper	mixed	mixed	mixed
	Sentences	37,192	42,024	6,872	57,885	6,606
	Tokens	861,831	1,026,033	171,838	1,049,083	44,692
validate	**Text Type**	mixed	mixed	mixed	mixed	mixed
	Sentences	3,302	1,664	1,585	3,560	885
	Tokens	65,564	40,770	25,982	61,722	5,541
test	**Text Type**	mixed	newspaper	mixed	mixed	mixed
	Sentences	4,040	2,507	600	2,457	898
	Tokens	68,280	59,191	18,633	49,760	5,991

Cross-Framework Training and Evaluation Data

		EDS	PTG	UCCA	AMR	DRG
	Flavor	1	1	1	2	2
train	**Text Type**	newspaper	newspaper	mixed	mixed	mixed
	Sentences	37,192	42,024	6,872	57,885	6,606
	Tokens	861,831	1,026,033	171,838	1,049,083	44,692
validate	**Text Type**	mixed	mixed	mixed	mixed	mixed
	Sentences	3,302	1,664	1,585	3,560	885
	Tokens	65,564	40,770	25,982	61,722	5,541
test	**Text Type**	mixed	newspaper	mixed	mixed	mixed
	Sentences	4,040	2,507	600	2,457	898
	Tokens	68,280	59,191	18,633	49,760	5,991

Cross-Framework Training and Evaluation Data

		EDS	PTG	UCCA	AMR	DRG
	Flavor	1	1	1	2	2
train	**Text Type**	newspaper	newspaper	mixed	mixed	mixed
	Sentences	37,192	42,024	6,872	57,885	6,606
	Tokens	861,831	1,026,033	171,838	1,049,083	44,692
validate	**Text Type**	mixed	mixed	mixed	mixed	mixed
	Sentences	3,302	1,664	1,585	3,560	885
	Tokens	65,564	40,770	25,982	61,722	5,541
test	**Text Type**	mixed	newspaper	mixed	mixed	mixed
	Sentences	4,040	2,507	600	2,457	898
	Tokens	68,280	59,191	18,633	49,760	5,991

- Validation split is MRP 2019 evaluation data; allowed for fine-tuning;

Cross-Framework Training and Evaluation Data

		EDS	PTG	UCCA	AMR	DRG
	Flavor	1	1	1	2	2
train	**Text Type**	newspaper	newspaper	mixed	mixed	mixed
	Sentences	37,192	42,024	6,872	57,885	6,606
	Tokens	861,831	1,026,033	171,838	1,049,083	44,692
validate	**Text Type**	mixed	mixed	mixed	mixed	mixed
	Sentences	3,302	1,664	1,585	3,560	885
	Tokens	65,564	40,770	25,982	61,722	5,541
test	**Text Type**	mixed	newspaper	mixed	mixed	mixed
	Sentences	4,040	2,507	600	2,457	898
	Tokens	68,280	59,191	18,633	49,760	5,991

- Validation split is MRP 2019 evaluation data; allowed for fine-tuning;
- linguistics: smallish WSJ sample in all frameworks publicly available;
- evaluation: subset of 100 sentences from *The Little Prince* is public.

Cross-Lingual Training and Evaluation Data

		PTG	UCCA	AMR	DRG
	Language	Czech	German	Chinese	German
	Flavor	1	1	1	2
train	**Text Type**	newspaper	mixed	mixed	mixed
	Sentences	43,955	4,125	18,365	1,575
	Tokens	740,466	95,634	428,054	9,088
test	**Text Type**	newspaper	mixed	mixed	mixed
	Sentences	5,476	444	1,713	403
	Tokens	92,643	10,585	39,228	2,384

Cross-Lingual Training and Evaluation Data

		PTG	UCCA	AMR	DRG
	Language	Czech	German	Chinese	German
	Flavor	1	1	1	2
train	**Text Type**	newspaper	mixed	mixed	mixed
	Sentences	43,955	4,125	18,365	1,575
	Tokens	740,466	95,634	428,054	9,088
test	**Text Type**	newspaper	mixed	mixed	mixed
	Sentences	5,476	444	1,713	403
	Tokens	92,643	10,585	39,228	2,384

Cross-Lingual Training and Evaluation Data

		PTG	UCCA	AMR	DRG
	Language	Czech	German	Chinese	German
	Flavor	1	1	1	2
train	**Text Type**	newspaper	mixed	mixed	mixed
	Sentences	43,955	4,125	18,365	1,575
	Tokens	740,466	95,634	428,054	9,088
test	**Text Type**	newspaper	mixed	mixed	mixed
	Sentences	5,476	444	1,713	403
	Tokens	92,643	10,585	39,228	2,384

Cross-Lingual Training and Evaluation Data

		PTG	UCCA	AMR	DRG
	Language	Czech	German	Chinese	German
	Flavor	1	1	1	2
train	**Text Type**	newspaper	mixed	mixed	mixed
	Sentences	43,955	4,125	18,365	1,575
	Tokens	740,466	95,634	428,054	9,088
test	**Text Type**	newspaper	mixed	mixed	mixed
	Sentences	5,476	444	1,713	403
	Tokens	92,643	10,585	39,228	2,384

- Gold-standard graphs for one additional language in four frameworks;

Cross-Lingual Training and Evaluation Data

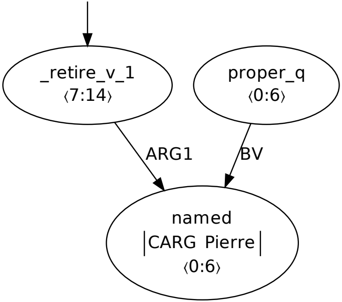
		PTG	UCCA	AMR	DRG
	Language	Czech	German	Chinese	German
	Flavor	1	1	1	2
train	**Text Type**	newspaper	mixed	mixed	mixed
	Sentences	43,955	4,125	18,365	1,575
	Tokens	740,466	95,634	428,054	9,088
test	**Text Type**	newspaper	mixed	mixed	mixed
	Sentences	5,476	444	1,713	403
	Tokens	92,643	10,585	39,228	2,384

- Gold-standard graphs for one additional language in four frameworks;
- ‘low-resource’ training setting for two frameworks: UCCA and DRG;

? explor opportunities for cross-lingual transfer learning (in-framework).

Cross-Framework Evaluation: MRP Graph Similarity

- Break down graphs into types of information: per-type and overall F_1;



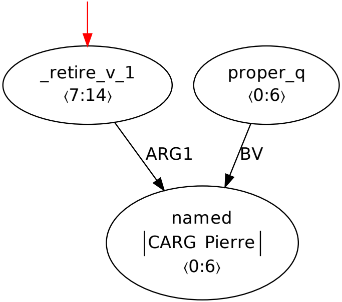
Pierre retired.

Different Types of Semantic Graph 'Atoms'

	EDS	PTG	UCCA	AMR	DRG
Top Nodes	✓	✓	✓	✓	✓
Labeled Edges	✓	✓	✓	✓	(✓)
Node Labels	✓	✓	✗	✓	✓
Node Properties	✓	✓	✗	✓	✗
Node Anchoring	✓	(✓)	(✓)	✗	✗
Edge Attributes	✗	✓	✓	✗	✗

Cross-Framework Evaluation: MRP Graph Similarity

- Break down graphs into types of information: per-type and overall F_1;
- tops



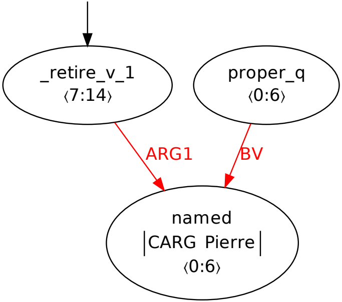
Pierre retired.

Different Types of Semantic Graph 'Atoms'

	EDS	PTG	UCCA	AMR	DRG
Top Nodes	✓	✓	✓	✓	✓
Labeled Edges	✓	✓	✓	✓	(✓)
Node Labels	✓	✓	✗	✓	✓
Node Properties	✓	✓	✗	✓	✗
Node Anchoring	✓	(✓)	(✓)	✗	✗
Edge Attributes	✗	✓	✓	✗	✗

Cross-Framework Evaluation: MRP Graph Similarity

- Break down graphs into types of information: per-type and overall F_1;
- tops and (labeled) edges;



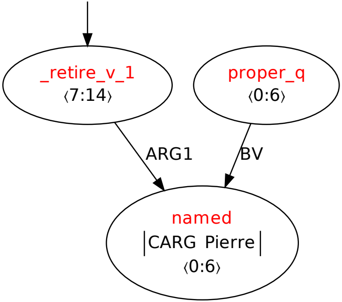
Pierre retired.

Different Types of Semantic Graph 'Atoms'

	EDS	PTG	UCCA	AMR	DRG
Top Nodes	✓	✓	✓	✓	✓
Labeled Edges	✓	✓	✓	✓	(✓)
Node Labels	✓	✓	✗	✓	✓
Node Properties	✓	✓	✗	✓	✗
Node Anchoring	✓	(✓)	(✓)	✗	✗
Edge Attributes	✗	✓	✓	✗	✗

Cross-Framework Evaluation: MRP Graph Similarity

- Break down graphs into types of information: per-type and overall F_1;
- tops and (labeled) edges; labels,



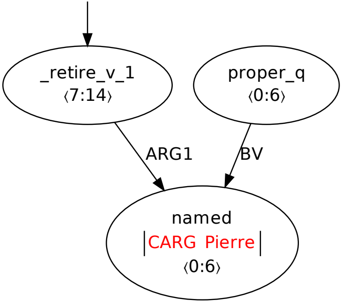
Pierre retired.

Different Types of Semantic Graph 'Atoms'

	EDS	PTG	UCCA	AMR	DRG
Top Nodes	✓	✓	✓	✓	✓
Labeled Edges	✓	✓	✓	✓	(✓)
Node Labels	✓	✓	✗	✓	✓
Node Properties	✓	✓	✗	✓	✗
Node Anchoring	✓	(✓)	(✓)	✗	✗
Edge Attributes	✗	✓	✓	✗	✗

Cross-Framework Evaluation: MRP Graph Similarity

- Break down graphs into types of information: per-type and overall F_1;
- tops and (labeled) edges; labels, properties,



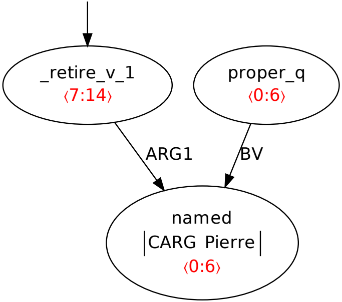
Pierre retired.

Different Types of Semantic Graph 'Atoms'

	EDS	PTG	UCCA	AMR	DRG
Top Nodes	✓	✓	✓	✓	✓
Labeled Edges	✓	✓	✓	✓	(✓)
Node Labels	✓	✓	✗	✓	✓
Node Properties	✓	✓	✗	✓	✗
Node Anchoring	✓	(✓)	(✓)	✗	✗
Edge Attributes	✗	✓	✓	✗	✗

Cross-Framework Evaluation: MRP Graph Similarity

- Break down graphs into types of information: per-type and overall F_1;
- tops and (labeled) edges; labels, properties, anchors,



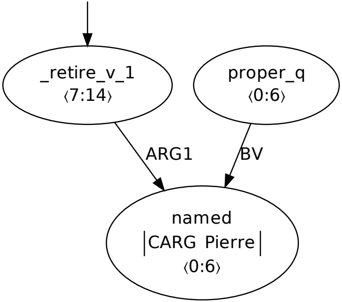
Pierre retired.

Different Types of Semantic Graph 'Atoms'

	EDS	PTG	UCCA	AMR	DRG
Top Nodes	✓	✓	✓	✓	✓
Labeled Edges	✓	✓	✓	✓	(✓)
Node Labels	✓	✓	✗	✓	✓
Node Properties	✓	✓	✗	✓	✗
Node Anchoring	✓	(✓)	(✓)	✗	✗
Edge Attributes	✗	✓	✓	✗	✗

Cross-Framework Evaluation: MRP Graph Similarity

- Break down graphs into types of information: per-type and overall F_1;
- tops and (labeled) edges; labels, properties, anchors, and attributes;



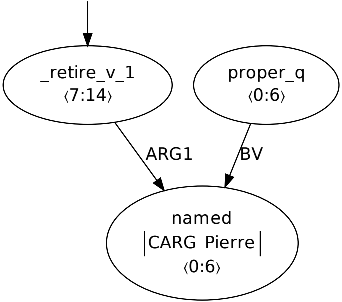
Pierre retired.

Different Types of Semantic Graph 'Atoms'

	EDS	PTG	UCCA	AMR	DRG
Top Nodes	✓	✓	✓	✓	✓
Labeled Edges	✓	✓	✓	✓	(✓)
Node Labels	✓	✓	✗	✓	✓
Node Properties	✓	✓	✗	✓	✗
Node Anchoring	✓	(✓)	(✓)	✗	✗
Edge Attributes	✗	✓	✓	✗	✗

Cross-Framework Evaluation: MRP Graph Similarity

- Break down graphs into types of information: per-type and overall F_1;
- tops and (labeled) edges; labels, properties, anchors, and attributes;
- requires node–node correspondences; search for overall maximum score;
- maximum common edge subgraph isomorphism (MCES) is NP-hard;



Pierre retired.

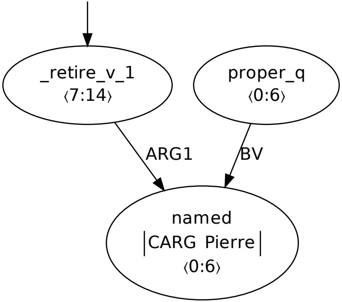
Different Types of Semantic Graph 'Atoms'

	EDS	PTG	UCCA	AMR	DRG
Top Nodes	✓	✓	✓	✓	✓
Labeled Edges	✓	✓	✓	✓	(✓)
Node Labels	✓	✓	✗	✓	✓
Node Properties	✓	✓	✗	✓	✗
Node Anchoring	✓	(✓)	(✓)	✗	✗
Edge Attributes	✗	✓	✓	✗	✗

Cross-Framework Evaluation: MRP Graph Similarity

- Break down graphs into types of information: per-type and overall F_1;
- tops and (labeled) edges; labels, properties, anchors, and attributes;
- requires node–node correspondences; search for overall maximum score;
- maximum common edge subgraph isomorphism (MCES) is NP-hard;

→ smart initialization, scheduling, and pruning yield strong approximation.



Pierre retired.

Different Types of Semantic Graph 'Atoms'

	EDS	PTG	UCCA	AMR	DRG
Top Nodes	✓	✓	✓	✓	✓
Labeled Edges	✓	✓	✓	✓	(✓)
Node Labels	✓	✓	✗	✓	✓
Node Properties	✓	✓	✗	✓	✗
Node Anchoring	✓	(✓)	(✓)	✗	✗
Edge Attributes	✗	✓	✓	✗	✗

Graphbank Statistics (Kuhlmann & Oepen, 2016)

			EDS	PTG	UCCA	AMR^{-1}	DRG
	(02)	Average Tokens per Graph	22.17	24.42	25.01	18.12	6.77
	(03)	Average Nodes per Token	1.26	0.74	1.33	0.64	2.09
	(04)	Distinct Edge Labels	10	72	15	101	16
proportions	(05)	Percentage of top nodes	0.99	1.27	1.66	3.77	3.40
	(06)	Percentage of node labels	29.02	21.61	–	43.91	39.81
	(07)	Percentage of node properties	12.54	26.22	–	7.63	–
	(08)	Percentage of node anchors	29.02	19.63	38.80	–	–
	(09)	Percentage of (labeled) edges	28.43	26.10	56.88	44.69	56.79
	(10)	Percentage of edge attributes	–	5.17	2.66	–	–
treeness	(11)	$\%_g$ Rooted Trees	0.09	22.63	28.19	22.05	0.35
	(12)	$\%_g$ Treewidth One	68.60	22.67	34.17	49.91	0.35
	(13)	Average Treewidth	1.317	2.067	1.691	1.561	2.131
	(14)	Maximal Treewidth	3	7	4	5	5
	(15)	Average Edge Density	1.015	1.177	1.055	1.092	1.265
	(16)	$\%_n$ Reentrant	32.77	16.23	4.90	19.89	25.92
	(17)	$\%_g$ Cyclic	0.27	33.97	0.00	0.38	0.27
	(18)	$\%_g$ Not Connected	1.90	0.00	0.00	0.00	0.00
	(19)	$\%_g$ Multi-Rooted	99.93	0.00	0.00	71.64	32.32

Graphbank Statistics (Kuhlmann & Oepen, 2016)

			EDS	PTG	UCCA	AMR^{-1}	DRG
	(02)	Average Tokens per Graph	22.17	24.42	25.01	18.12	6.77
	(03)	Average Nodes per Token	1.26	0.74	1.33	0.64	2.09
	(04)	Distinct Edge Labels	10	72	15	101	16
proportions	(05)	Percentage of top nodes	0.99	1.27	1.66	3.77	3.40
	(06)	Percentage of node labels	29.02	21.61	–	43.91	39.81
	(07)	Percentage of node properties	12.54	26.22	–	7.63	–
	(08)	Percentage of node anchors	29.02	19.63	38.80	–	–
	(09)	Percentage of (labeled) edges	28.43	26.10	56.88	44.69	56.79
	(10)	Percentage of edge attributes	–	5.17	2.66	–	–
treeness	(11)	$\%_g$ Rooted Trees	0.09	22.63	28.19	22.05	0.35
	(12)	$\%_g$ Treewidth One	68.60	22.67	34.17	49.91	0.35
	(13)	Average Treewidth	1.317	2.067	1.691	1.561	2.131
	(14)	Maximal Treewidth	3	7	4	5	5
	(15)	Average Edge Density	1.015	1.177	1.055	1.092	1.265
	(16)	$\%_n$ Reentrant	32.77	16.23	4.90	19.89	25.92
	(17)	$\%_g$ Cyclic	0.27	33.97	0.00	0.38	0.27
	(18)	$\%_g$ Not Connected	1.90	0.00	0.00	0.00	0.00
	(19)	$\%_g$ Multi-Rooted	99.93	0.00	0.00	71.64	32.32

Graphbank Statistics (Kuhlmann & Oepen, 2016)

			EDS	PTG	UCCA	$\mathbf{AMR}^{-1}$	DRG
	(02)	Average Tokens per Graph	22.17	24.42	25.01	18.12	6.77
	(03)	Average Nodes per Token	1.26	0.74	1.33	0.64	2.09
	(04)	Distinct Edge Labels	10	72	15	101	16
proportions	(05)	Percentage of top nodes	0.99	1.27	1.66	3.77	3.40
	(06)	Percentage of node labels	29.02	21.61	–	43.91	39.81
	(07)	Percentage of node properties	12.54	26.22	–	7.63	–
	(08)	Percentage of node anchors	29.02	19.63	38.80	–	–
	(09)	Percentage of (labeled) edges	28.43	26.10	56.88	44.69	56.79
	(10)	Percentage of edge attributes	–	5.17	2.66	–	–
treeness	(11)	$\%_g$ Rooted Trees	0.09	22.63	28.19	22.05	0.35
	(12)	$\%_g$ Treewidth One	68.60	22.67	34.17	49.91	0.35
	(13)	Average Treewidth	1.317	2.067	1.691	1.561	2.131
	(14)	Maximal Treewidth	3	7	4	5	5
	(15)	Average Edge Density	1.015	1.177	1.055	1.092	1.265
	(16)	$\%_n$ Reentrant	32.77	16.23	4.90	19.89	25.92
	(17)	$\%_g$ Cyclic	0.27	33.97	0.00	0.38	0.27
	(18)	$\%_g$ Not Connected	1.90	0.00	0.00	0.00	0.00
	(19)	$\%_g$ Multi-Rooted	99.93	0.00	0.00	71.64	32.32

Graphbank Statistics (Kuhlmann & Oepen, 2016)

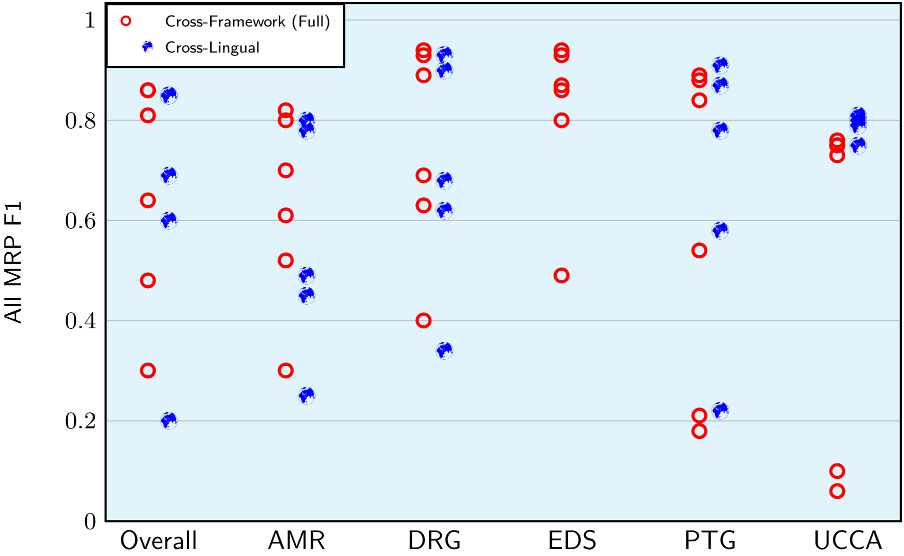
			EDS	PTG	UCCA	AMR^{-1}	DRG
	(02)	Average Tokens per Graph	22.17	24.42	25.01	18.12	6.77
	(03)	Average Nodes per Token	1.26	0.74	1.33	0.64	2.09
	(04)	Distinct Edge Labels	10	72	15	101	16
proportions	(05)	Percentage of top nodes	0.99	1.27	1.66	3.77	3.40
	(06)	Percentage of node labels	29.02	21.61	–	43.91	39.81
	(07)	Percentage of node properties	12.54	26.22	–	7.63	–
	(08)	Percentage of node anchors	29.02	19.63	38.80	–	–
	(09)	Percentage of (labeled) edges	28.43	26.10	56.88	44.69	56.79
	(10)	Percentage of edge attributes	–	5.17	2.66	–	–
treeness	(11)	$\%_g$ Rooted Trees	0.09	22.63	28.19	22.05	0.35
	(12)	$\%_g$ Treewidth One	68.60	22.67	34.17	49.91	0.35
	(13)	Average Treewidth	1.317	2.067	1.691	1.561	2.131
	(14)	Maximal Treewidth	3	7	4	5	5
	(15)	Average Edge Density	1.015	1.177	1.055	1.092	1.265
	(16)	$\%_n$ Reentrant	32.77	16.23	4.90	19.89	25.92
	(17)	$\%_g$ Cyclic	0.27	33.97	0.00	0.38	0.27
	(18)	$\%_g$ Not Connected	1.90	0.00	0.00	0.00	0.00
	(19)	$\%_g$ Multi-Rooted	99.93	0.00	0.00	71.64	32.32

High-Level Overview of Submissions

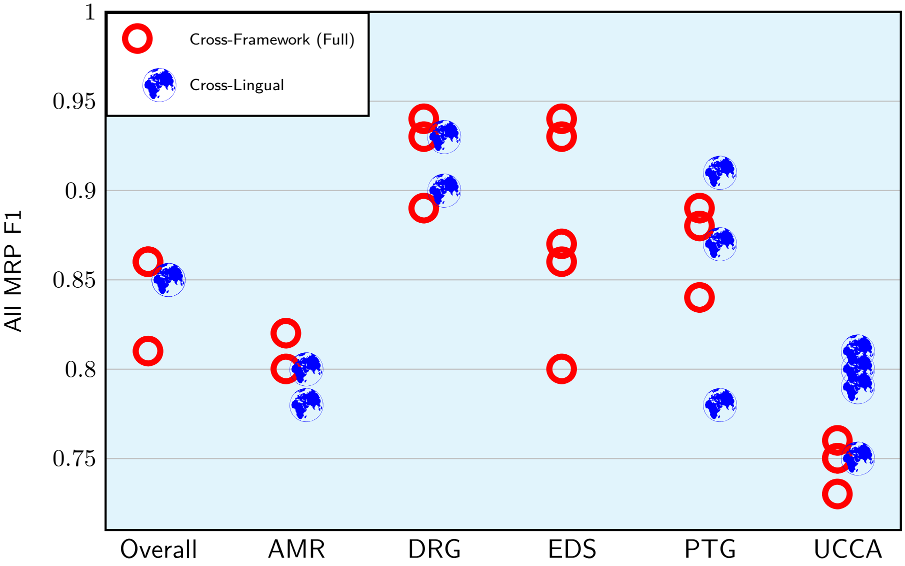
○ Cross-Framework 🌍 Cross-Lingual

Teams	AMR	DRG	EDS	PTG	UCCA
Hitachi	○🌍	○🌍	○	○🌍	○🌍
ÚFAL	○🌍	○🌍	○	○🌍	○🌍
HIT-SCIR	○🌍	○🌍	○	○🌍	○🌍
HUJI-KU	○🌍	○🌍	○	○🌍	○🌍
ISCAS	○	○	○	○	○
TJU-BLCU	○🌍	○🌍	○	○🌍	○
JBNU	○				
ÚFAL	○🌍	○🌍	○	○🌍	○🌍
ERG			○		

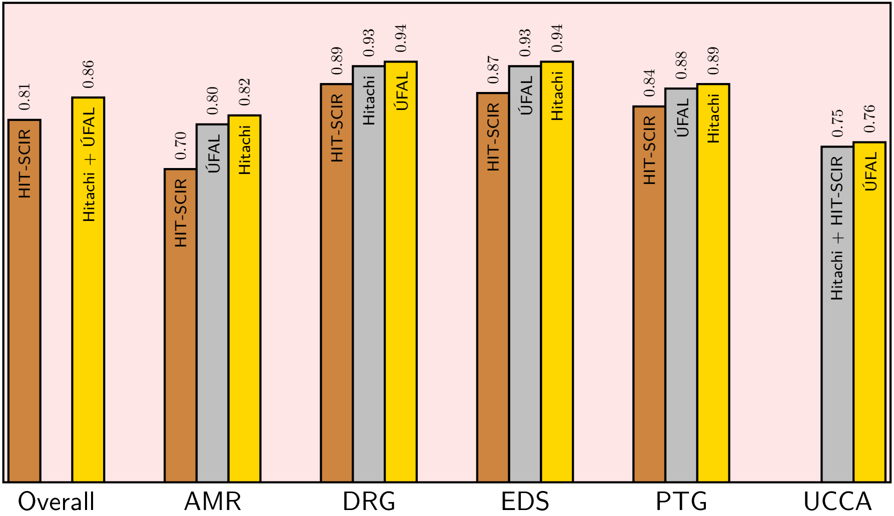
Score Distribution



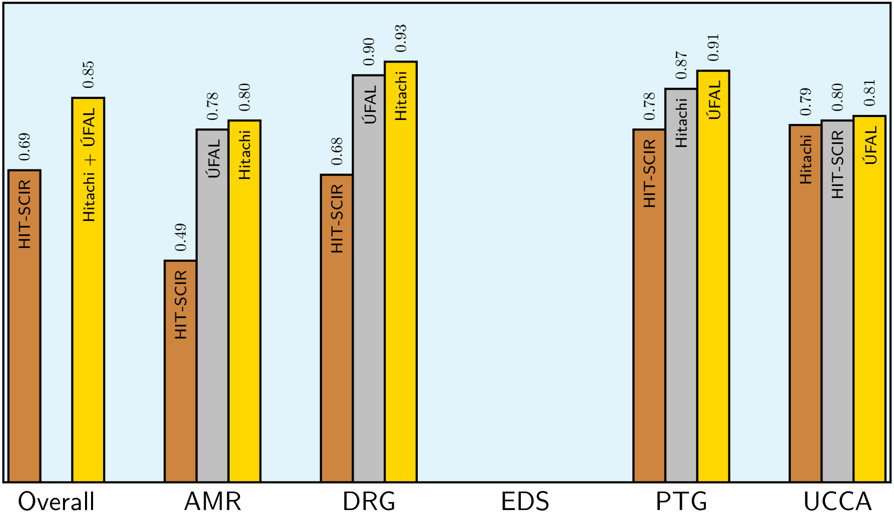
Score Distribution: Zoom In



○Cross-Framework Track: Full Evaluation



Cross-Lingual Track: Full Evaluation

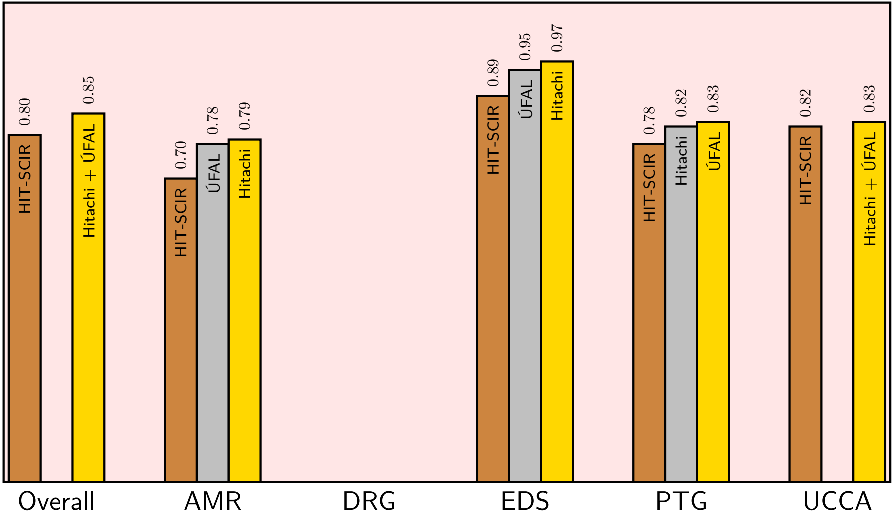


Medal Ceremony!

○ Cross-Framework 🌍 Cross-Lingual

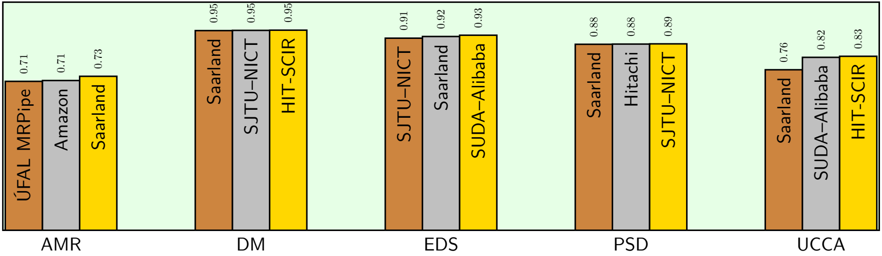
Teams		AMR	DRG	EDS	PTG	UCCA
Hitachi	○ 🌍	○ 🌍	○ 🌍	○	○ 🌍	○ 🌍
ÚFAL	○ 🌍	○ 🌍	○ 🌍	○	○ 🌍	○ 🌍
HIT-SCIR	○ 🌍	○ 🌍	○ 🌍	○	○ 🌍	○ 🌍
HUJI-KU		○ 🌍	○ 🌍	○	○ 🌍	○ 🌍
ISCAS		○	○	○	○	○
TJU-BLCU		○ 🌍	○ 🌍	○	○ 🌍	○

Cross-Framework Track: The Little Prince

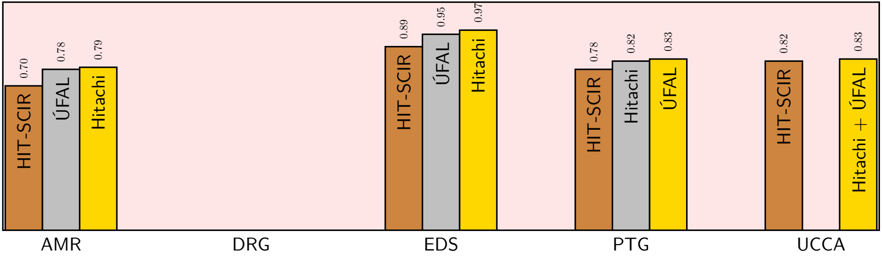


State of the Art: The Little Prince

MRP 2019:



MRP 2020:



Interim Conclusions & Outlook

Lessons Learned (from Two Consecutive Shared Tasks)

- Good community interest: 180 subscribers; 19 data licenses (via LDC);
- technical barriers and ‘competitive selection’: $6 + 2$ teams submitted;

Interim Conclusions & Outlook

Lessons Learned (from Two Consecutive Shared Tasks)

- Good community interest: 180 subscribers; 19 data licenses (via LDC);
- technical barriers and 'competitive selection': $6 + 2$ teams submitted;

$\rightarrow$ advanced state of the art on four frameworks (but possibly not AMR);

Interim Conclusions & Outlook

Lessons Learned (from Two Consecutive Shared Tasks)

- Good community interest: 180 subscribers; 19 data licenses (via LDC);
- technical barriers and 'competitive selection': $6 + 2$ teams submitted;

$\rightarrow$ advanced state of the art on four frameworks (but possibly not AMR);

$\rightarrow$ greatly increased cross-framework uniformity; but limited (M)TL so far.

Interim Conclusions & Outlook

Lessons Learned (from Two Consecutive Shared Tasks)

- Good community interest: 180 subscribers; 19 data licenses (via LDC);
- technical barriers and ‘competitive selection’: $6+2$ teams submitted;

→ advanced state of the art on four frameworks (but possibly not AMR);

→ greatly increased cross-framework uniformity; but limited (M)TL so far.

Outlook: Beyond MRP 2020

- High-quality, robust meaning representation parsers generally available;
- MRP 2020 data, metrics, submissions, and scores as stable benchmark;

Interim Conclusions & Outlook

Lessons Learned (from Two Consecutive Shared Tasks)

- Good community interest: 180 subscribers; 19 data licenses (via LDC);
- technical barriers and 'competitive selection': $6+2$ teams submitted;

→ advanced state of the art on four frameworks (but possibly not AMR);

→ greatly increased cross-framework uniformity; but limited (M)TL so far.

Outlook: Beyond MRP 2020

- High-quality, robust meaning representation parsers generally available;
- MRP 2020 data, metrics, submissions, and scores as stable benchmark;

? post-mortem contrastive analysis of architectures (Buljan et al., 2020);

Interim Conclusions & Outlook

Lessons Learned (from Two Consecutive Shared Tasks)

- Good community interest: 180 subscribers; 19 data licenses (via LDC);
- technical barriers and 'competitive selection': $6+2$ teams submitted;

→ advanced state of the art on four frameworks (but possibly not AMR);

→ greatly increased cross-framework uniformity; but limited (M)TL so far.

Outlook: Beyond MRP 2020

- High-quality, robust meaning representation parsers generally available;
- MRP 2020 data, metrics, submissions, and scores as stable benchmark;

? post-mortem contrastive analysis of architectures (Buljan et al., 2020);

? increased focus on evaluation metrics: score 'larger pieces'; SEMBLEU;

Interim Conclusions & Outlook

Lessons Learned (from Two Consecutive Shared Tasks)

- ► Good community interest: 180 subscribers; 19 data licenses (via LDC);
- ► technical barriers and 'competitive selection': $6 + 2$ teams submitted;
- → advanced state of the art on four frameworks (but possibly not AMR);
- → greatly increased cross-framework uniformity; but limited (M)TL so far.

Outlook: Beyond MRP 2020

- ► High-quality, robust meaning representation parsers generally available;
- ► MRP 2020 data, metrics, submissions, and scores as stable benchmark;
- ? post-mortem contrastive analysis of architectures (Buljan et al., 2020);
- ? increased focus on evaluation metrics: score 'larger pieces'; SEMBLEU;
- → open discussion with 2020 participants towards the end of this session.

Jayeol Chun, Dotan Dvir, Dan Flickinger,

Jiří Mírovský, Anna Nedoluzhko, Sebastian Schuster,

Milan Straka, and Zdeňka Urešová

Linguistic Data Consortium,

Nordic Language Processing Laboratory

Omri Abend & Ari Rappoport. 2013. Universal Conceptual Cognitive Annotation (UCCA). In Proceedings of the 51th Meeting of the Association for Computational Linguistics, pages 228–238, Sofia, Bulgaria.

Lasha Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik van Noord, Pierre Ludmann, Duc-Duy Nguyen, & Johan Bos. 2017. The Parallel Meaning Bank. Towards a multilingual corpus of translations annotated with compositional meaning representations. In Proceedings of the 15th Meeting of the European Chapter of the Association for Computational Linguistics, pages 242–247, Valencia, Spain.

Collin F. Baker, Charles J. Fillmore, & John B. Lowe. 1998. The Berkeley FrameNet project. In Proceedings of the 17th International Conference on Computational Linguistics and the 36th Meeting of the Association for Computational Linguistics, pages 86–90, Stroudsburg, PA, USA.

References II

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, & Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse, pages 178 – 186, Sofia, Bulgaria.

Maja Buljan, Joakim Nivre, Stephan Oepen, & Lilja Øvrelid. 2020. A tale of three parsers. Towards diagnostic evaluation for meaning representation parsing. In Proceedings of the 12th Language Resources and Evaluation Conference, pages 1902–1909, Marseille, France. European Language Resources Association.

Dan Flickinger, Stephan Oepen, & Emily M. Bender. 2017. Sustainable development and refinement of complex linguistic annotations at scale. In Nacy Ide & James Pustejovsky, editors, Handbook of Linguistic Annotation, pages 353 – 377. Springer, Dordrecht, The Netherlands.

Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, & Zdeněk Žabokrtský. 2012. Announcing Prague Czech-English Dependency Treebank 2.0. In Proceedings of the 8th International Conference on Language Resources and Evaluation, pages 3153–3160, Istanbul, Turkey.

Hardy & Andreas Vlachos. 2018. Guided neural language generation for abstractive summarization using abstract meaning representation. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium.

Hans Kamp & Uwe Reyle. 1993. From Discourse to Logic. An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and DRT. Kluwer, Dordrecht, The Netherlands.

Marco Kuhlmann & Stephan Oepen. 2016. Towards a catalogue of linguistic graph banks. Computational Linguistics, 42(4):819–827.

References IV

Stephan Oepen & Jan Tore Lønning. 2006. Discriminant-based MRS banking. In Proceedings of the 5th International Conference on Language Resources and Evaluation, pages 1250–1255, Genoa, Italy.

Martha Palmer, Dan Gildea, & Paul Kingsbury. 2005. The Proposition Bank. A corpus annotated with semantic roles. Computational Linguistics, 31(1):71–106.

Sebastian Schuster & Christopher D. Manning. 2016. Enhanced English Universal Dependencies. An improved representation for natural language understanding tasks. In Proceedings of the 10th International Conference on Language Resources and Evaluation, Portorož, Slovenia.

Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, & Benjamin Van Durme. 2016. Universal Decompositional Semantics on Universal Dependencies. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1713–1723, Austin, TX, USA.

Daniel Zeman & Jan Hajič. 2020. FGD at MRP 2020: Prague Tectogrammatical Graphs. In Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing, pages 33–39, Online.