Evaluation Framework for Layered Meaning Representation

Rémi de Vergnette, Maxime Amblard, Bruno Guillaume

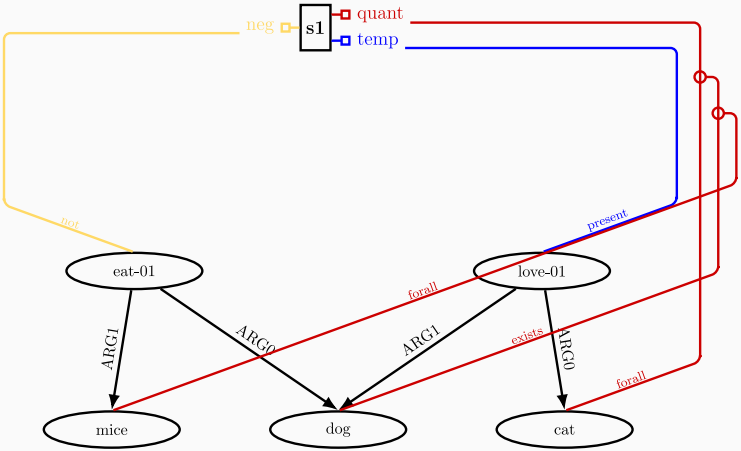
04/08/2025

Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

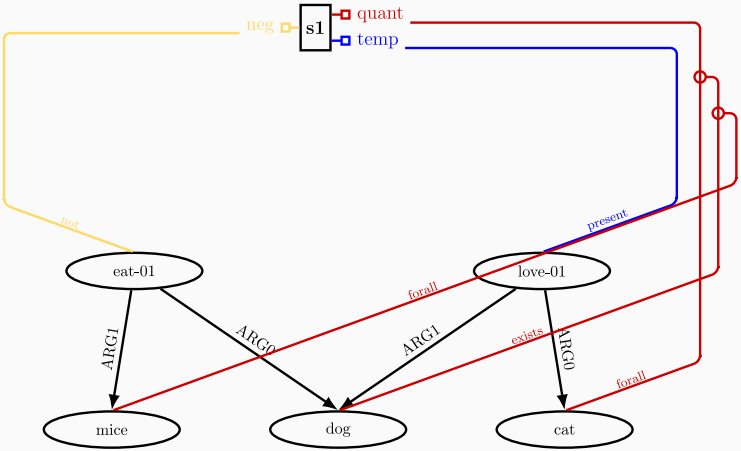
Introduction to YARN

- **YARN**[4]: laYered meAning RepresentatioN
- Extends AMR [1] with typed edges and vertices
- Richer structure : Some edges can connect to other edges (not just vertices)
- Modular framework for partial annotations
- Claims to be more expressive than AMR by handling quantification, modalities, aspect, scope

A YARN example (1/2)



A YARN example (1/2)



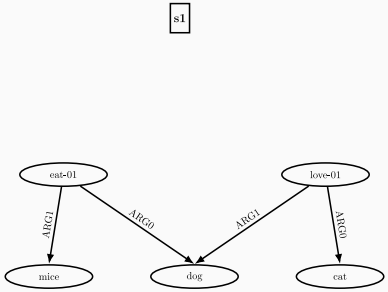
A YARN for : “ every cat loves a dog who doesn’t eat mice ”

YARN formal definition (1/4)

Definition

- **9-tuple definition** [4]:
  $Y = (S, V, F, D, E, C, L, H, I)$
- S: Elementary event nodes
- V: Vertices
- E: Edges
- (and more)

Illustration

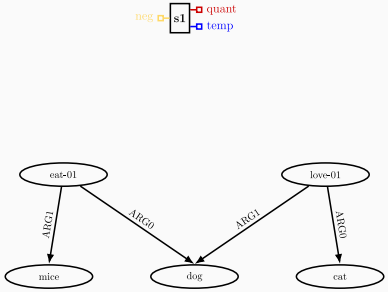


YARN formal definition (2/4)

Definition

- S: Elementary event nodes
- V: Vertices
- E: Edges
- F: Feature nodes associated with events
- (and more)

Illustration

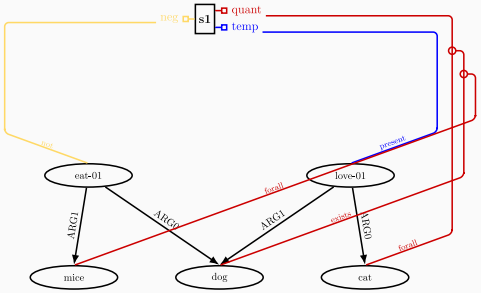


YARN formal definition (3/4)

Definition

- S: Elementary event nodes
- V: Vertices
- E: Edges
- F: Feature nodes associated with events
- L: Feature edges connecting features to V vertices
- H: Feature edges connecting feature edges to V vertices or E edges
- (and more)

Illustration



$$\forall c, \text{cat}(c) \Rightarrow (\exists d, \text{dog}(d) \wedge$$

$$(\forall m, \text{mice}(m) \implies \neg \text{eat}(m, d)) \wedge$$

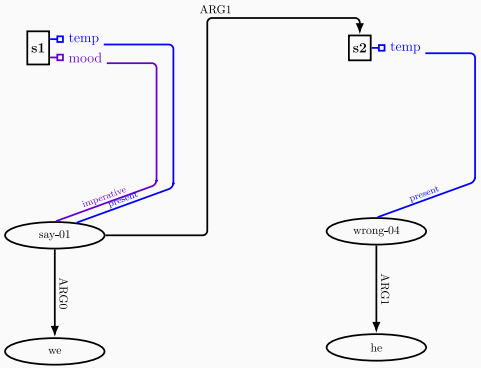
$$(\text{loves.01}(c, d)))$$

YARN formal definition (4/4)

Definition

- D: Discourse relation edges (From S nodes to S nodes)
- C: Clause-linking edges (From V nodes to S nodes)
- I: Edges imposing restrictions on interpretation (between V nodes)

Illustration



Metrics for YARN

- SMATCH [2] is well know for AMR graphs
- YARN's complex structures cannot be directly evaluated with existing metrics
- Need for modular evaluation matching YARN's modular nature
- Requirement to evaluate specific linguistic phenomena separately

Extending SMATCH to YARN

- Encode YARN structures as sets of clauses (triples and quadruples)
- Add variables corresponding to edges

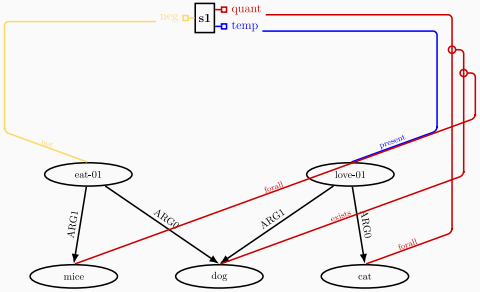
Smatch : an edge x - [rel] - y is encoded as a triple rel(x, y) where:

SmatchY : an edge x - [rel] - y is encoded as a quadruple $a :=$ rel(x, y) where:

- x and y are vertices variables, rel is the relation label
- a is the edge variable (used to reference the edge in other clauses)

Triples and Quadruples

Structure



Clauses

$l_6 := \texttt{forall\_l}(quant, c)$

$l_7 := \texttt{not\_l}(neg, e)$

$l_8 := \texttt{present\_l}(temp, l)$

$h_2 := \texttt{forall\_h}(h_1, m)$

$h_1 := \texttt{exists\_h}(l_6, d)$

$\texttt{instance\_v}(\texttt{cat}, c)$

$e_1 := \texttt{ARG0\_e}(l, c)$

...

ILP Formulation

- Use Integer Linear Programming (ILP) to find optimal variable alignment [3][2]
  Objective: Find optimal variable alignment between two YARN structures

Variables:

- $v : V_1 \times V_2 \to \{0, 1\}$ (variable assignment)
- $t : C_1 \times C_2 \to \{0, 1\}$ (clause matching)

Constraints:

- Partial one-to-one variable alignment: $\sum_{i=1}^{n} v_{ij} \leq 1$, $\sum_{j=1}^{m} v_{ij} \leq 1$
- For two comparable clauses (same label and type) $t_{c_i c_j} \leq v_{xa}, v_{yb}, v_{zc}$

Optimization: $\max_{(t,v)\in\Lambda} \sum_{c_i \in C_1, c_j \in C_2} t_{c_i c_j}$

Problem with Base Approach

- All YARN elements treated equally
- High baseline scores (0.45 for random pairs)
- Nearly empty graph scores 0.55 against real annotations
- No focus on specific semantic phenomena
- Solution : filter the clauses considered according to a set of features $\mathcal{F}$ and a set of types T

SMATCHY Variants

SMATCHY-GENERAL

- $T = \{S, V, D, E, C, L, H, I\}$ (excludes only F)
- General structure similarity
- Baseline score drops to 0.20

SMATCHY-PA

- $T = \{V, E\}$ (predicate-argument structure)
- Similar to original SMATCH

SMATCHY-FOL

- $T = \{S, V, E, H, L\}$, $\mathcal{F} = \{\text{quant}, \text{neg}\}$
- Evaluates first-order logic capabilities

YARNBLEU

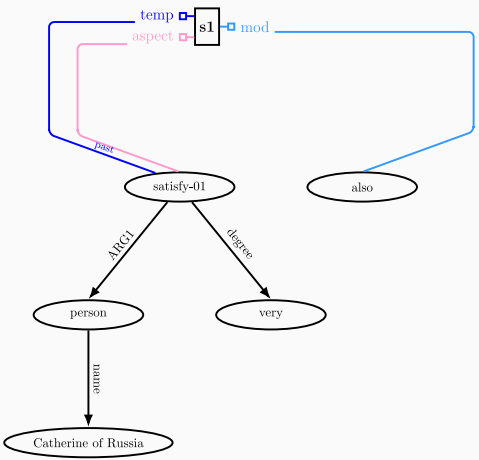
- Adapt SemBLEU [5]
- Reify every edge
- Apply breadth-first traversal for k-grams extraction
- Use BLEU formula

Same filtering approach as SMATCHY:

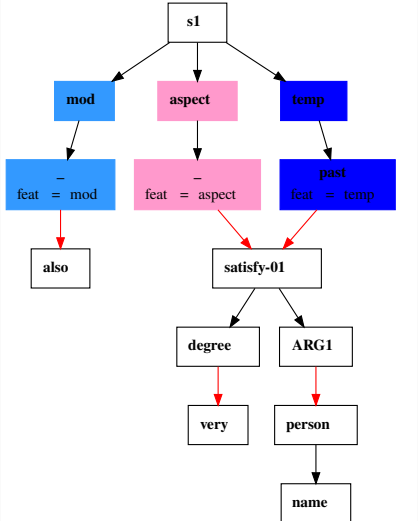
- **YARNBLEU-GENERAL**: General similarity
- **YARNBLEU-PA**: Predicate-argument focus
- **YARNBLEU-FOL**: First-order logic focus

YARN as graphs

Original structure



Graph translation



Evaluation Protocol

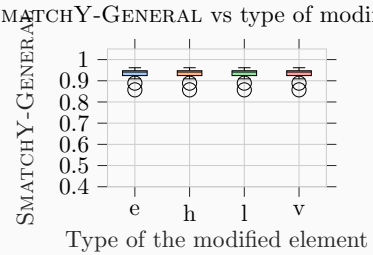
- Small dataset of 100 annotated YARN structures
- Apply random modifications simulating annotation errors
- Maintain valid YARN structures at each step
- Observe score degradation

Modification Types:

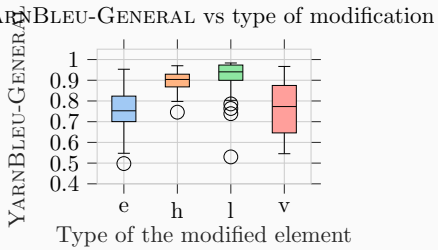
- Label changes
- Add or remove edges (L, H, E)

Results (1 modification)

SmatchY

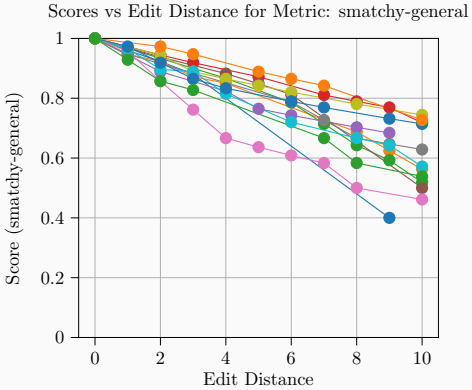


YarnBleu

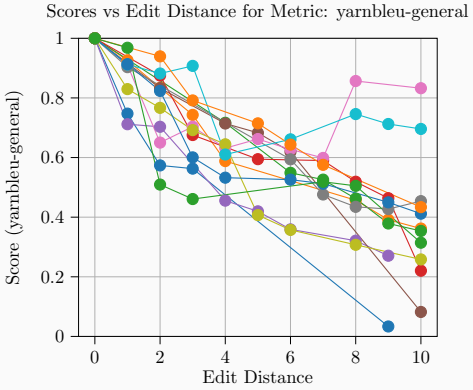


Results (Several modifications)

SmatchY



YARNBleu



Conclusion

- **YARN** introduces a complex structure requiring new evaluation metrics
- **SMATCHY** and **YARNBLEU** provide modular evaluation frameworks
- Both metrics can be tailored to specific linguistic phenomena
- **SMATCHY** shows better behavior with respect to our evaluation protocol
- Browse through YARN online !

[1] Laura Banarescu et al. **"Abstract Meaning Representation for Sembanking."** In: *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*. Sofia, Bulgaria: Association for Computational Linguistics, Aug. 2013, pp. 178–186. URL: https://aclanthology.org/W13-2322.

[2] Shu Cai and Kevin Knight. **"Smatch: an Evaluation Metric for Semantic Feature Structures."** In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Sofia, Bulgaria: Association for Computational Linguistics, Aug. 2013, pp. 748–752. URL: https://aclanthology.org/P13-2131.

[3] Juri Opitz. **"SMATCH++: Standardized and Extended Evaluation of Semantic Graphs."** In: *Findings of the Association for Computational Linguistics: EACL 2023*. Ed. by Andreas Vlachos and Isabelle Augenstein. Dubrovnik, Croatia: Association for Computational Linguistics, May 2023, pp. 1595–1607. DOI: 10.18653/v1/2023.findings-eacl.118. URL: https://aclanthology.org/2023.findings-eacl.118/.

[4] Siyana Pavlova. **"Tools and methods for semantically annotated corpora."** PhD thesis. Université de Lorraine, 2025.

[5] Linfeng Song and Daniel Gildea. **"SemBleu: A Robust Metric for AMR Parsing Evaluation."** In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 4547–4552. DOI: 10.18653/v1/P19-1446. URL: https://aclanthology.org/P19-1446.