Rule-based semantic interpretation for Universal Dependencies

Jamie Findlay, Saeedeh Salimifar, Ahmet Yıldırım, and Dag Haug

University of Oslo

UD Workshop, GURT/SyntaxFest
March 11, 2023

Goals

- The *Universal Natural Language Understanding* project aims to build a system which can
  1. create rich, logic-based representations (specifically, DRSs); and
  2. do this for as many languages as possible.
- Takes advantage of the universalism/cross-linguistic coverage of Universal Dependencies.
- Aims to avoid/postpone the need for language-specific (e.g. lexical) resources.

Semantic representations

Mary says John likes goats.

$$
b_1 : \begin{array}{|l|}
\hline
x \;\; e_1 \\
\hline
\mathit{person}(x) \\
\mathit{Name}(x, \text{“}\mathit{Mary}\text{”}) \\
\mathit{say}(e_1) \\
\mathit{Agent}(e_1, x) \\
\mathit{Topic}(e_1, b_2) \\
b_2 : \begin{array}{|l|}
\hline
e_2 \;\; y \;\; z \\
\hline
\mathit{person}(y) \\
\mathit{Name}(y, \text{“}\mathit{John}\text{”}) \\
\mathit{goat}(z) \\
\mathit{love}(e_2) \\
\mathit{Experiencer}(e_2, y) \\
\mathit{Source}(e_2, z) \\
\hline
\end{array} \\
\hline
\end{array}
$$

- **Discourse Representation Structures** (DRSs)
  - discourse referents + conditions on discourse referents
  - DRSs can be embedded inside other DRSs (also under negation, disjunction, modal operators, . . .)
- Translateable to first-order logic for use with theorem provers.

The Parallel Meaning Bank

- One good reason for choosing DRS representations is the Groningen Parallel Meaning Bank (Bos et al. 2017), which contains DRS representations for sentences in English, German, Dutch and Italian.

Manchester United defeated Fulham.

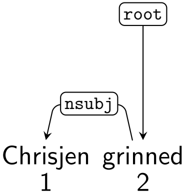
```
x1  x2  e1  t1
-----------------------------
team(x1)
   Name(x1, manchester~united)
time(t1)
   t1 < now
defeat(e1)
   Time(e1, t1)
   Co-Agent(e1, x2)
   Agent(e1, x1)
team(x2)
   Name(x2, fulham)
```

```
b1 REF x1                       % Manchester~United [0...17]
b1 Name x1 "manchester~united"  % Manchester~United [0...17]
b1 PRESUPPOSITION b3            % Manchester~United [0...17]
b1 team "n.01" x1               % Manchester~United [0...17]
b3 REF e1                       % defeated [18...26]
b3 REF t1                       % defeated [18...26]
b3 Agent e1 x1                  % defeated [18...26]
b3 Co-Agent e1 x2               % defeated [18...26]
b3 TPR t1 "now"                 % defeated [18...26]
b3 Time e1 t1                   % defeated [18...26]
b3 defeat "v.01" e1             % defeated [18...26]
b3 time "n.08" t1               % defeated [18...26]
b2 REF x2                       % Fulham [27...33]
b2 Name x2 "fulham"             % Fulham [27...33]
b2 PRESUPPOSITION b3            % Fulham [27...33]
b2 team "n.01" x2               % Fulham [27...33]
                                % . [33...34]
```

Glue semantics for composition

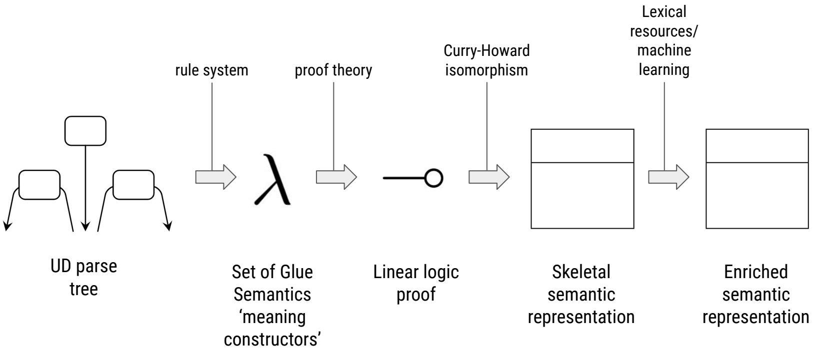
$$\textbf{grin}\quad \lambda y.\begin{array}{|l|}\hline e \\ \hline \mathit{grin}(e) \\ \mathit{Agent}(e,y) \\ \hline \end{array} : E(\hat{*}\ \texttt{nsubj}) \multimap T(\hat{*})$$

$$\textbf{chrisjen}\quad \lambda P.\begin{array}{|l|}\hline x \\ \hline \mathit{person}(x) \\ \mathit{Name}(x, \text{“}\mathit{Chrisjen}\text{”}) \\ \hline \end{array} + P(x) : [E(\hat{*}) \multimap T(\uparrow)] \multimap T(\uparrow)$$



$$\cfrac{\cfrac{\textbf{chrisjen} : [E(1) \multimap T(2)] \multimap T(2) \qquad \textbf{grin} : E(1) \multimap T(2)}{\textbf{chrisjen}(\textbf{grin}) : T(2)}}{\begin{array}{|l|}\hline x\ \ e \\ \hline \mathit{person}(x) \\ \mathit{Name}(x, \text{“}\mathit{Chrisjen}\text{”}) \\ \mathit{grin}(e) \\ \mathit{Agent}(e,x) \\ \hline \end{array} : T(2)} \Rightarrow_{\beta}$$

Pipeline


rule system
proof theory
Curry-Howard isomorphism
Lexical resources/ machine learning
λ
UD parse tree
Set of Glue Semantics ‘meaning constructors’
Linear logic proof
Skeletal semantic representation
Enriched semantic representation

Examples of rules

- Based on POS

```
coarsePos = PROPN -> \X.(([], [Name(X, ‘:LEMMA:‘)])) : e(!) -o t(!)
```

$$\lambda X.\begin{array}{|c|}\hline \\ \hline Name(X, \text{“}:LEMMA:\text{”}) \\ \hline\end{array} : E(\hat{*}) \multimap T(\hat{*})$$

- Based on deprel

```
relation = nsubj; ^ {coarsePos = VERB} ->
\Q.\V.\F.(Q(\X.(V(\E.(([], [nsubj(E,X)]) + F(E)))))) :
((e(!) -o t(^)) -o t(^)) -o (x(^) -o x(^))
```

$$\lambda Q \lambda V \lambda F.Q(\lambda X.(V(\lambda E.\begin{array}{|c|}\hline \\ \hline nsubj(E,X) \\ \hline\end{array} + F(E)))) : [[(E(\hat{*}) \multimap T(\uparrow)] \multimap T(\uparrow)] \multimap X(\uparrow) \multimap X(\uparrow)$$

$$(x \equiv \langle\langle v, t\rangle, t\rangle)$$

- Based on features

```
coarsePos = VERB; ~ aux.*; Tense = Pres ->
\V.\F.(V(\E.(([T], [time(T), EQ(T, ‘now‘), Time(E, T)]) + F(E)))) : x(!) -o x(!)
```

$$\lambda V \lambda F.V(\lambda E.\begin{array}{|l|}\hline \quad T \\ \hline time(T) \\ T = \text{‘}now\text{’} \\ Time(E,T) \\ \hline\end{array} + F(E)) : X(\hat{*}) \multimap X(\hat{*})$$

Language parametrisation

- Always aiming for universal rules, but ...
- Some structural differences emerge from typological parameters.
  - E.g. negative concord: one negation or two?
- Some structural differences emerge from lexical facts.
  - E.g. quantifiers: universal or existential? (*every* vs. *some*)

Language parametrisation

```
"eng": {
    "stanza_lang_code":"en",
    "stanza_processors":"tokenize,mwt,pos,lemma,depparse",
    "rules":["rules.dat"],
    "typological_features": {
        "negative_concord" : "no",
        "sequence_of_tense": "yes",
        "grammatical_gender": "no"
      },
    "lexical_items": {
        "future_aux": "will",
        "definite_det": "(the|this|that)",
        "indefinite_det": "(a|some)",
        "universal_quantifier": "(every|each)",
        "infinitive_marker": "to",
        "conjunction": "and",
        "disjunction": "or"
      },
    "chop_rules":["chopRules.dat"],
    "templates":["templates.dat"]
  }
```

Post-processing

- Output of rule-system is underspecified; language-specific information added when available, such as
  - Translation of UD relations to thematic roles

Jim loves Naomi.

$$
\begin{array}{|l|}
\hline
e \; x \; y \\
\hline
person(x) \\
Name(x, \text{‘}Jim\text{’}) \\
person(y) \\
Name(x, \text{‘}Naomi\text{’}) \\
love(e) \\
\mathbf{nsubj(e, x)} \\
\mathbf{obj(e, y)} \\
\hline
\end{array}
\Rightarrow
\begin{array}{|l|}
\hline
e \; x \; y \\
\hline
person(x) \\
Name(x, \text{‘}Jim\text{’}) \\
person(y) \\
Name(x, \text{‘}Naomi\text{’}) \\
love(e) \\
\mathbf{Experiencer(e, x)} \\
\mathbf{Source(e, y)} \\
\hline
\end{array}
$$

- Also (to come):
  - Idioms/multiword expressions
  - Anaphora resolution

Comparison

Neural network DRS parsing

Number of parameters	Encoder	Decoder	PMB 2.2.0 dev	PMB 2.2.0 test	PMB 3.0.0 dev	PMB 3.0.0 test	PMB 4.0.0 dev	PMB 4.0.0 test	PMB 4.0.0 eval
139,636,648	No-PT, 6x768-6	6x768-6	86.65	87.65	89.78	89.16	89.05	89.48	87.32
139,636,648	No-PT, 6x768-12	6x768-12	86.45	87.8	89.64	89.48	89.03	89.45	87.51
102,389,160	No-PT, 8x512-8	8x512-8	86.87	87.26	89.46	89.64	89.06	89.61	87.38
55,775,144	bert_base_uncased	8x512-8	87.17	88.45	89.69	89.78	89.1	89.79	87.2
55,775,144	bert_base_cased	8x512-8	87.51	88.23	**89.96**	89.89	89.19	89.9	**88.18**
133,633,960	bert_base_uncased	12x768-12	87.41	88.18	89.57	89.66	**89.4**	**90.26**	87.36
133,633,960	bert_base_cased	12x768-12	**87.53**	**89.23**	89.78	**90.32**	88.07	89.04	86.9
134,421,160	bert_large_uncased	12x768-12	86.93	88.56	89.08	88.65	88.71	89.6	87.29
134,421,160	bert_large_cased	12x768-12	86.9	88.27	89.39	90.03	88.81	90.12	87.42
≈106 million	Prior work		86.1	88.7	88.4	89.3			

Is rule-based parsing hopeless?

Language	Raw comparison F1	Rec	Prec	Structural only F1	Rec	Prec	Covered sentences	Total sentences	Proportion covered
German (de)	30.78	30.85	31.21	59.58	57.39	63.48	434	547	0.79
English (en)	46.69	48.07	46.28	63.42	63.92	64.32	874	1048	0.83
Italian (it)	30.68	30.55	31.40	58.88	57.25	61.92	429	461	0.93
Dutch (nl)	28.63	29.07	28.84	58.41	56.84	61.59	399	491	0.81

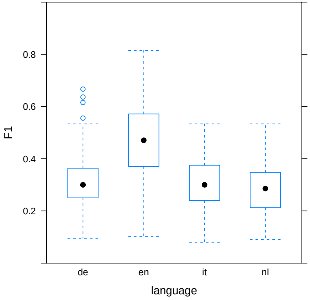


Figure: Raw comparison

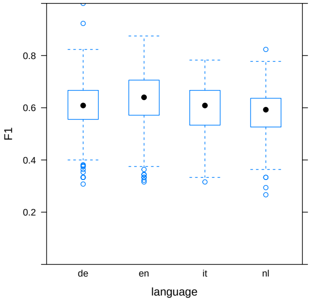
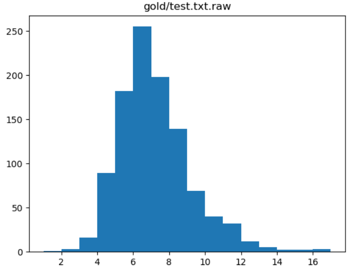
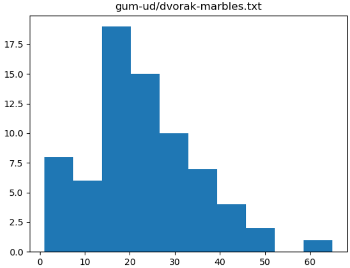


Figure: Structural only

(Dis)graceful degradation

- Neural nets do well on the PMB data, but the sentences in these data are all very short.
- We annotated two Wikipedia texts from the GUM-UD corpus in the PMB. These have much longer sentences.





(Dis)graceful degradation

- Now the advantage of neural approaches drops off almost entirely.
  (cf. Yao & Koller 2022; Donatelli & Koller 2023)

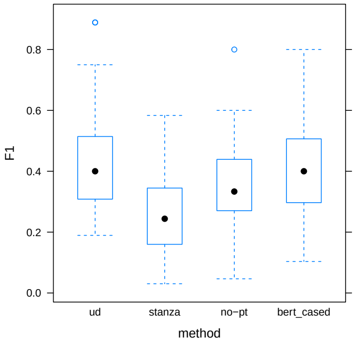


Figure: F1 scores on the GUM data

- As our rule coverage improves, we can only expect the UD scores to improve.

Conclusion

- Ongoing project to develop a rule-based semantic interpretation system for UD.
- Universal base with highly modular additions.
- Compositional system, with concomitant advantages for scaling with sentence length.

References

Bos, Johan, Valerio Basile, Kilian Evang, Noortje J Venhuizen & Johannes Bjerva. 2017. The Groningen meaning bank. In *Handbook of linguistic annotation*, 463–496. Springer.

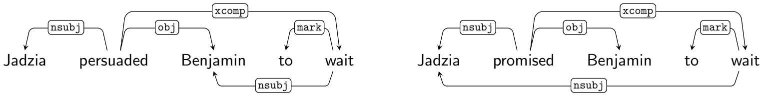
Donatelli, Lucia & Alexander Koller. 2023. Compositionality in computational linguistics. *Annual Review of Linguistics* 9. 463–481. https://doi.org/10.1146/annurev-linguistics-030521-044439.

Findlay, Jamie Y. & Dag T. T. Haug. 2021. How useful are Enhanced Universal Dependencies for semantic interpretation? In *Proceedings of the Sixth International Conference on Dependency Linguistics (Depling, SyntaxFest 2021)*, 22–34. Sofia: Association for Computational Linguistics. https://aclanthology.org/2021.depling-1.3.

Yao, Yuekun & Alexander Koller. 2022. Structural generalization is hard for sequence-to-sequence models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 5048–5062. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. https://aclanthology.org/2022.emnlp-main.337.

What about Enhanced UD?

- EUD provides semantically richer starting point.



- But cross-linguistic coverage much sparser.
  - Only (usefully) present in 31/213 UD treebanks (Findlay & Haug 2021).

Features: morphosyntactic or semantic?

- UD feature annotations scrupulously kept to the word level: problematic when features of *phrases* emerge non-compositionally.
- But e.g. `Tense` given a semantic definition:

Tense is a feature that specifies the time when the action took/takes/will take place, in relation to a reference point.

- And yet:

(1) I was singing
`Tense=Past, VerbForm=Fin` `Tense=`$\boxed{\texttt{Pres}}$`, VerbForm=Part`

(2) They will be defeated
`Tense=`$\boxed{\texttt{Past}}$`, VerbForm=Part, Voice=Pass`