The Relative Clauses AMR Parsers Hate Most

Xiulin Yang Nathan Schneider

Georgetown University

May 19, 2024

How effectively do AMR parsers handle different types of English relative clauses (RCs)?

Relative Clauses

I know the person who you like .

Relative Clauses

I know the person who you like .

Long Distance Dependency (LDD)

dep

I know the person who you like __ .

Long Distance Dependency (LDD)

dep

I know the person who you think that Tom mentioned once that __ likes you .

Abstract Meaning Representation (AMR)

- AMR is a graph semantic representation that captures the core semantic roles and relations in a sentence.
  - Usually *who did what to whom, where and when*.
- Each AMR is a single rooted, directed graph, which can be represented with Penman Notation.
- *I know the person who likes you*

Abstract Meaning Representation (AMR)

- AMR is a graph semantic representation that captures the core semantic roles and relations in a sentence.
    - Usually *who did what to whom, where and when*.
- Each AMR is a single rooted, directed graph, which can be represented with Penman Notation.
- *I know the person who likes you*

```
(k / know-01
:ARG0 (i / i)
:ARG1 (p / person
    :ARG0-of (l / like-01
        :ARG1 (y / you))))
```

Relative Clause in AMRs

- *I know the person who likes you.*

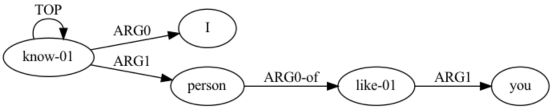

Figure: Canonical AMR graph. The `ARG0-of` edge corresponds to the relative clause.

Relative Clause in AMRs

- *I know the person who likes you.*

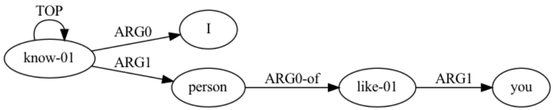

Figure: Canonical AMR graph. The `ARG0-of` edge corresponds to the relative clause.

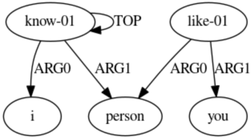

Figure: Normalized AMR graph. The `ARG0` edge from `like-01` to `person` corresponds to the relative clause.

Relative Clause Types

- **Subject RC**:

  acl:relcl nsubj

  He is the person **who** **__** **stole** **my** **book** .

- **Object RC**:

  acl:relcl obj

  He is the person **that** **you** **like** **__** .

- **Oblique RC**:

  obl acl:relcl

  He is the person **that** **I** **borrowed** **the** **book** **from** **__** .

- **Passive Subject RC**:

  acl:relcl nsubj:pass

  He is the person who __ is liked by you .

Relative Clause Types

- **Subject RC**:

  acl:relcl nsubj

  He is the person **who** __ **stole my book** .

- **Object RC**:

  acl:relcl obj

  He is the person **that you like** __ .

- **Oblique RC**:

  acl:relcl obl

  He is the person **that I borrowed the book from** __ .

- **Passive Subject RC**:

  acl:relcl nsubj:pass

  He is the person who __ is liked by you .

Reduced Relative Clause Types

- **Reduced Object RC**:

  acl:relcl obj

  He is the person **you like __** .

- **Reduced Oblique RC**:

  obl

  acl:relcl

  He is the person **I borrowed the book from __** .

"AMR parsing is far from solved" (Groschwitz et al., 2023)

- SOTA AMR Parser (Lee et al., 2022) achieved over 0.85 in Smatch (Cai and Knight, 2013).
- Relying solely on overall F-scores does not fully reveal a parser's performance across different linguistic phenomena (Groschwitz et al., 2023)
- Seq2seq models that simply take input as sequence string fail at structural generalization compared with models that explicitly encode structural information (Yao and Koller, 2022; Li et al., 2023; Shaw et al., 2021)
- Recovering reentrancy structures is a challenge for AMR parsers (Szubert et al., 2020; Damonte et al., 2017)

"AMR parsing is far from solved" (Groschwitz et al., 2023)

- SOTA AMR Parser (Lee et al., 2022) achieved over 0.85 in Smatch (Cai and Knight, 2013).
- Relying solely on overall F-scores does not fully reveal a parser's performance across different linguistic phenomena (Groschwitz et al., 2023)

structure unaware

- Seq2seq models that simply take input as sequence string fail at structural generalization compared with models that explicitly encode structural information (Yao and Koller, 2022; Li et al., 2023; Shaw et al., 2021)

structure-aware

- Recovering reentrancy structures is a challenge for AMR parsers (Szubert et al., 2020; Damonte et al., 2017)

Research Questions

- **How well can AMR parsers capture the long-distance predicate-argument dependencies in RCs?**
  - Does structure-awareness help the models to parse?
  - Which types of RC are most challenging and why?

Relative clause

- **Subject RC**:
  acl:relcl nsubj
  He is the person who __ stole my book .
- **Object RC**:
  acl:relcl obj
  He is the person that you like __ .
- **Oblique RC**:
  acl:relcl obl
  He is the person that I borrowed the book from __ .
- **Passive Subject RC**:
  acl:relcl nsubj:pass
  He is the person who __ is liked by you .
- **Reduced Object RC**:
  acl:relcl obj
  He is the person you like __ .
- **Reduced Oblique RC**:
  obl acl:relcl
  He is the person I borrowed the book from __ .

Method

- Datasets
- Models
- Evaluation Metric

Datasets

Dataset	# sents	# tokens
EWT (Silveira et al., 2014)	1,449	26.5
CRC (Prasad et al., 2019)	1,400	13.7
AMR 3.0 (Knight et al., 2021)	259	29.1

Table: Number of sentences containing RCs in the datasets and the mean sentence length

Datasets

Dataset	# sents	# tokens
EWT (Silveira et al., 2014)	**1,449**	**26.5**
CRC (Prasad et al., 2019)	1,400	13.7
AMR 3.0 (Knight et al., 2021)	259	29.1

Table: Number of sentences containing RCs in the datasets and the mean sentence length

Models

- Structure-aware models
  - `AM-Parser` (Groschwitz et al., 2018): compositional parser composed of a supertagger + dependency parser
  - `AMRBART` (Bai et al., 2022): structural pretraining + fine-tuning
- Structure-**un**aware models
  - `Spring` (Spring et al., 2021)
  - `amrlib-BART`[1]
  - `amrlib-T5`
- All models are fine-tuned on AMR 3.0.

[1] https://github.com/bjascob/amrlib

Evaluation Metric - Reentrancy recall

- Our evaluation assesses whether the relativized noun in a sentence is **reentrant**, with two incoming edges—one originating from **the main clause's predicate verb** and another from **the predicate within the RC**.
- To do so, we use LEAMR (Blodgett and Schneider, 2021), a probabilistic, fine-grained aligner optimized for English AMR.

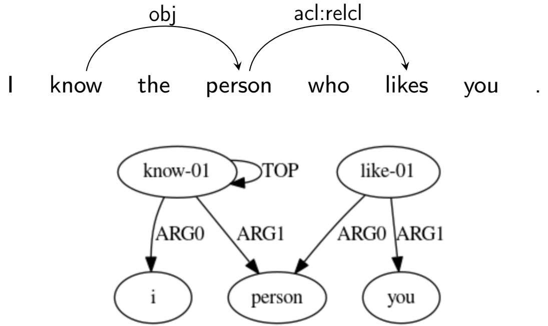

Figure: Normalized AMR graph for the sentence *I know the person who likes you.*.

I know the person who likes you.

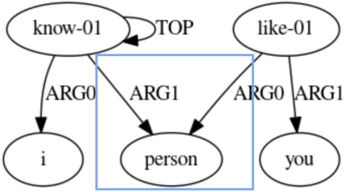

Figure: Correct prediction ✓

I know the person who likes you.

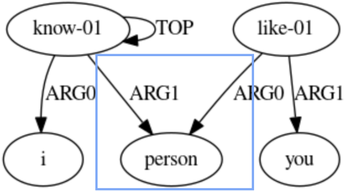

Figure: Correct prediction ✓

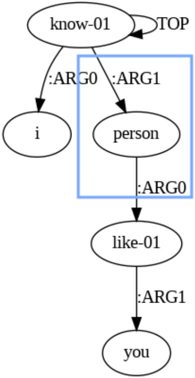

Figure: Incorrect prediction ✗

Structure-aware vs Structure-unaware

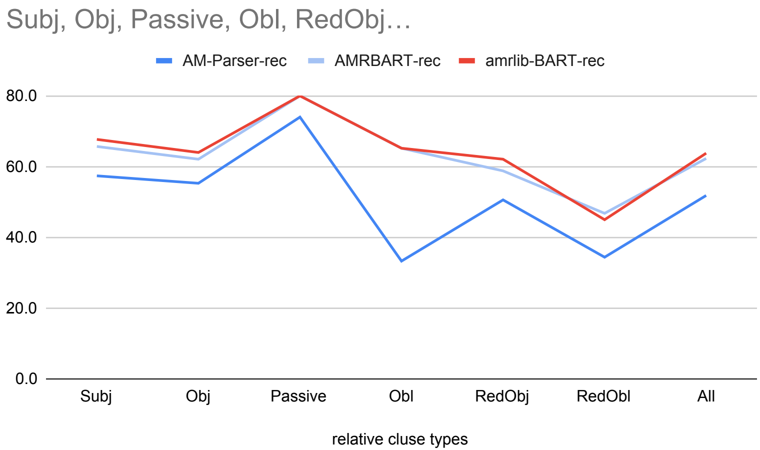

Figure: RC reentrancy recall of AM-Parser, amrlib-BART and AMRBART, by RC subtype and overall.

Relative Clause Types

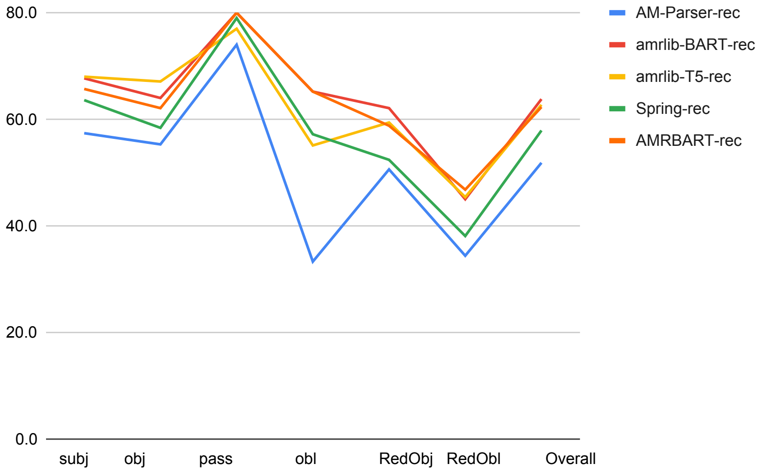

Figure: RC reentrancy recall of all parsers by RC subtype and overall.

Why contributes to such discrepancies?

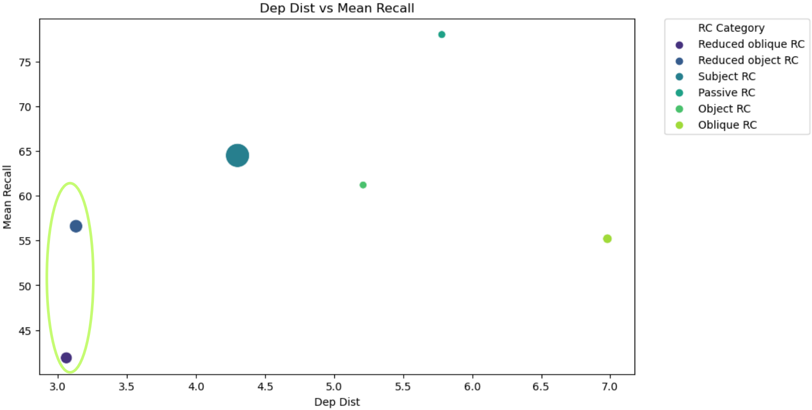

Figure: Average Dependency Distance vs Mean Recall across RC Types.

Why contributes to such discrepancies?

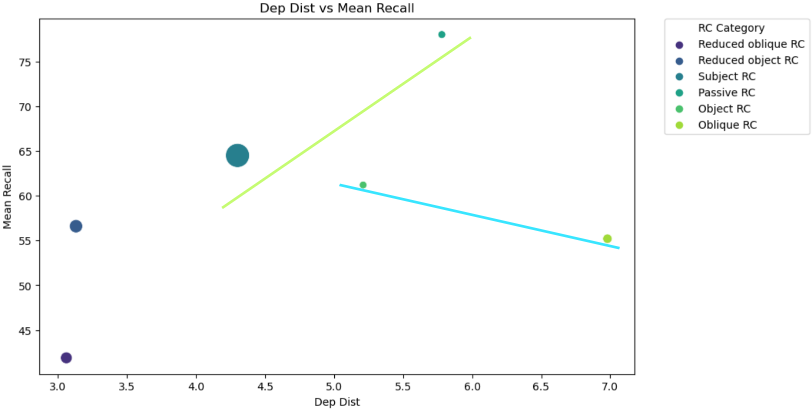

Figure: Average Dependency Distance vs Mean Recall across RC Types.

Why contributes to such discrepancies?

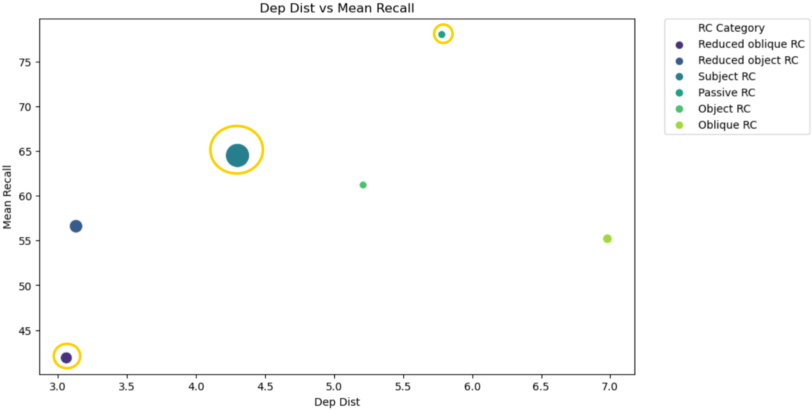

Figure: Average Dependency Distance vs Mean Recall across RC Types.

Takeaways

- **Does structure-awareness help the models to parse?**
- **Which types of RC are most challenging and why?**

Takeaways

- **Does structure-awareness help the models to parse?**
  - Seq2seq models, on the whole, outperform the compositional model
  - There is little difference in performance between seq2seq models that are aware of structure and those that are not.
- **Which types of RC are most challenging and why?**
  - Relative clauses are challenging for current parsers
  - Reduced RCs are the most challenging RC types.
  - The full RCs with shorter dependency distances are easier to parse
  - Linguistic cues?

Thank you for your attention!

References I

Xuefeng Bai, Yulong Chen, and Yue Zhang. 2022. Graph pre-training for AMR parsing and generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6001–6015, Dublin, Ireland. Association for Computational Linguistics.

Austin Blodgett and Nathan Schneider. 2021. Probabilistic, structure-aware algorithms for improved variety, accuracy, and coverage of AMR alignments. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3310–3321, Online. Association for Computational Linguistics.

Shu Cai and Kevin Knight. 2013. 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)*, pages 748–752, Sofia, Bulgaria. Association for Computational Linguistics.

Marco Damonte, Shay B. Cohen, and Giorgio Satta. 2017. An incremental parser for Abstract Meaning Representation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 536–546, Valencia, Spain. Association for Computational Linguistics.

Jonas Groschwitz, Shay Cohen, Lucia Donatelli, and Meaghan Fowlie. 2023. AMR parsing is far from solved: GrAPES, the granular AMR parsing evaluation suite. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 10728–10752, Singapore. Association for Computational Linguistics.

Jonas Groschwitz, Matthias Lindemann, Meaghan Fowlie, Mark Johnson, and Alexander Koller. 2018. AMR dependency parsing with a typed semantic algebra. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1831–1841, Melbourne, Australia. Association for Computational Linguistics.

Kevin Knight, Bianca Badarau, Laura Baranescu, Claire Bonial, Madalina Bardocz, Kira Griffitt, Ulf Hermjakob, Daniel Marcu, Martha Palmer, Tim O'Gorman, and Nathan Schneider. 2021. Abstract Meaning Representation (AMR) Annotation Release 3.0. Linguistic Data Consortium, LDC2020T02.

References II

Young-Suk Lee, Ramón Astudillo, Hoang Thanh Lam, Tahira Naseem, Radu Florian, and Salim Roukos. 2022. Maximum Bayes Smatch ensemble distillation for AMR parsing. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5379–5392, Seattle, United States. Association for Computational Linguistics.

Bingzhi Li, Lucia Donatelli, Alexander Koller, Tal Linzen, Yuekun Yao, and Najoung Kim. 2023. SLOG: A structural generalization benchmark for semantic parsing. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3213–3232, Singapore. Association for Computational Linguistics.

Grusha Prasad, Marten van Schijndel, and Tal Linzen. 2019. Using priming to uncover the organization of syntactic representations in neural language models. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 66–76, Hong Kong, China. Association for Computational Linguistics.

Peter Shaw, Ming-Wei Chang, Panupong Pasupat, and Kristina Toutanova. 2021. Compositional generalization and natural language variation: Can a semantic parsing approach handle both? In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 922–938, Online. Association for Computational Linguistics.

Natalia Silveira, Timothy Dozat, Marie-Catherine de Marneffe, Samuel R. Bowman, Miriam Connor, John Bauer, and Christopher D. Manning. 2014. A gold standard dependency corpus for English. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, pages 2897–2904, Reykjavík, Iceland.

Nicolas Spring, Annette Rios, and Sarah Ebling. 2021. Exploring German multi-level text simplification. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1339–1349, Held Online. INCOMA Ltd.

Ida Szubert, Marco Damonte, Shay B. Cohen, and Mark Steedman. 2020. The role of reentrancies in Abstract Meaning Representation parsing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2198–2207, Online. Association for Computational Linguistics.

References III

Yuekun Yao and 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*, pages 5048–5062, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Other RC Types

- Free relatives (e.g., *I heard what* ***you said***)
- Possessive RCs (e.g., *I like the girl* ***whose dress is blue***)
- Reduced subject RCs (e.g., *I met the person* ***you mentioned*** ___ ***finished all the work this week***)
- Adnominal participial clauses (e.g., *the sheep* ***eaten by wolves***)

Attainable Rate vs Recall

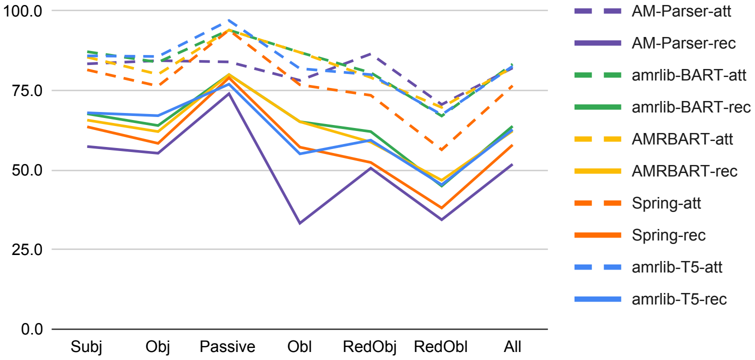

Figure: RC reentrancy recall (solid lines) and attainability rate (dashed) of all parsers, by RC subtype and overall.

Dependency Distances and Counts across RC Types

RC Category	Dep Dist	Count	Mean Recall
Reduced oblique RC	3.06	1,092	41.9
Reduced object RC	3.13	1,371	56.6
Subject RC	4.30	4,226	64.5
Passive Subject RC	5.78	534	78.0
Object RC	5.21	516	61.2
Oblique RC	6.98	729	55.2

Table: Mean dependency distance of 6 types of RCs in our experiments