UNIVERSITY OF COPENHAGEN
DEPARTMENT OF COMPUTER SCIENCE

Can AMR Assist Legal and Logical Reasoning?

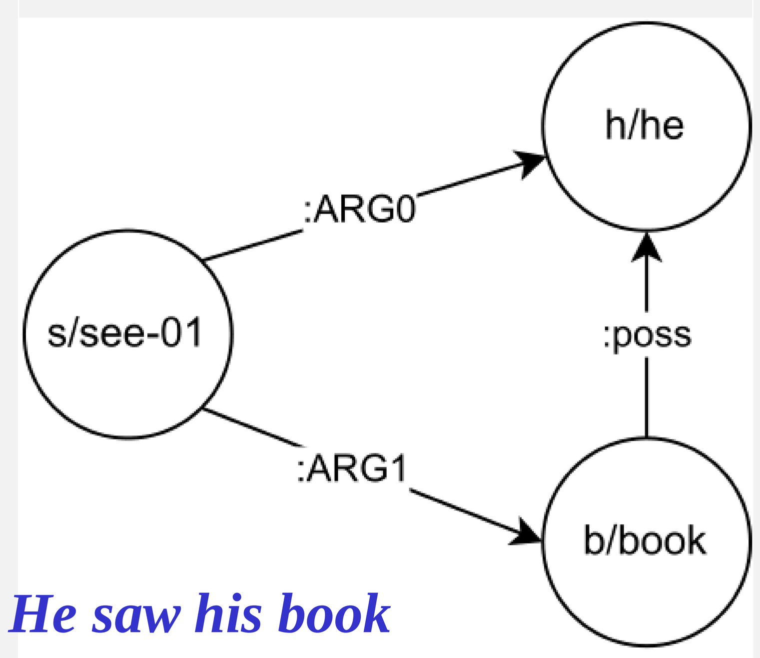
Nikolaus Schrack, Ruixiang Cui, Hugo A. López and Daniel Hershcovich

- Can **Abstract Meaning Representation (AMR)** help capture logical relationships on multiple choice question answering (MCQA) tasks?
- We propose neural architectures that utilize **linearised AMR graphs** in combination with pre-trained language models.
- While not able to outperform text-only baselines, they have **complementary abilities**.
- Error analysis further reveals that **AMR parsing quality** is the most prominent challenge, especially with multiple sentences.
- Theoretical analysis of logical relations in AMR concludes it might be helpful in **some logical statements** but not for others.

Understanding the logic in law is a major challenge in legal NLP. Use AMR instead of or in addition to the textual input may allow a system to better encode the document semantics.

What is AMR?

Graph-structured representation of sentence meaning (Banarescu et al., 2013).
Rooted, directed acyclic graph. Nodes represent concepts, edges encode relations.



Logical Relations in AMR

Propositional logic includes negation, conjunction, implication, and quantifiers.

- Conditional statements (*if*, *unless*, *in case of*, etc.) represented by `:condition`.
- Negations represented by `:polarity`.

```
(a / accident
  :polarity -
  :mod (t / traffic)
  :ARG1-of (n / major-02)
  :condition (c / close-01
    :polarity -
    :ARG1 (h / highway)))
```

No major traffic accidents will occur if the highway is not closed

Other operators do not follow such patterns. AMR for conjunctions (*and*, *however*, *moreover*, etc.) depends on specific surface form

AMR helps capture some logical statements but not others.

MCQA

CaseHOLD: legal reasoning (Zheng et al., 2021).
Common task for lawyers, identify legal holding of a case.

LogiQA: logical reasoning (Liu et al., 2020).
Sourced from the National Civil Servants Examination of China, professionally translated into English.

Court decision statement Drapeau's cohorts, the cohort would be a "victim" of making the bomb. Further, firebombs are inherently dangerous. There is no peaceful purpose for making a bomb. Felony offenses that involve explosives qualify as "violent crimes" for purposes of enhancing the sentences of career offenders. See 18 U.S.C. §924(e)(2)(B)(ii) (defining a "violent felony" as: "any crime punishable by imprisonment for a term exceeding one year ... that ... involves use of explosives"). Courts have found possession of a bomb to be a crime of violence based on the lack of a nonviolent purpose for a bomb and the fact that, by its very nature, there is a substantial risk that the bomb would be used against the person or property of another. See United States v. Newman, 125 F.3d 863 (10th Cir.1997) (unpublished) (<HOLDING>); United States v. Dodge, 846 F.Supp. 181,
Holding Statement 1 (correct) holding that possession of a pipe bomb is a crime of violence for purposes of 18 usc 3142f1
Holding Statement 2 (not correct) holding that bank robbery by force and violence or intimidation under 18 usc 2113a is a crime of violence
Holding Statement 3 (not correct) holding that sexual assault of a child qualified as crime of violence under 18 usc 16
Holding Statement 4 (not correct) holding for the purposes of 18 usc 924e that being a felon in possession of a firearm is not a violent felony as defined in 18 usc 924e2b
Holding Statement 5 (not correct) holding that a court must only look to the statutory definition not the underlying circumstances of the crime to determine whether a given offense is by its nature a crime of violence for purposes of 18 usc 16

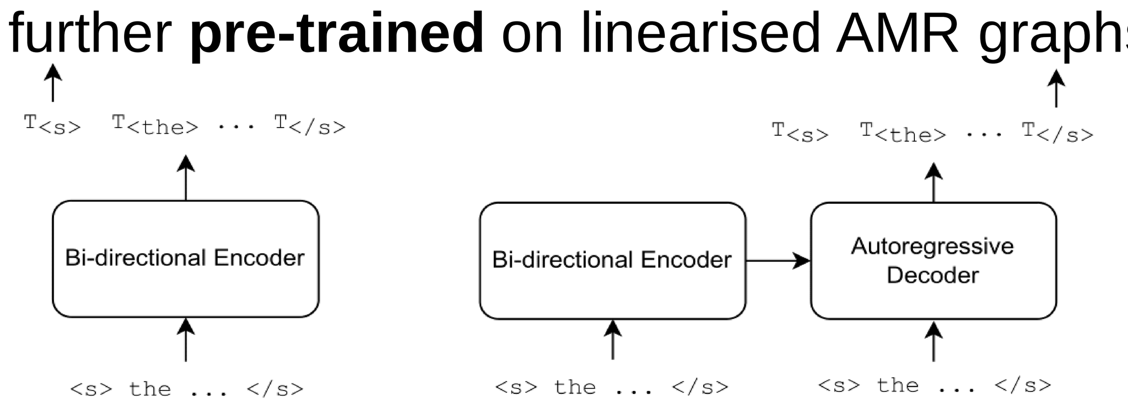
Example of CaseHOLD. A court decision statement and five holding statements are given.

AMR for MCQA

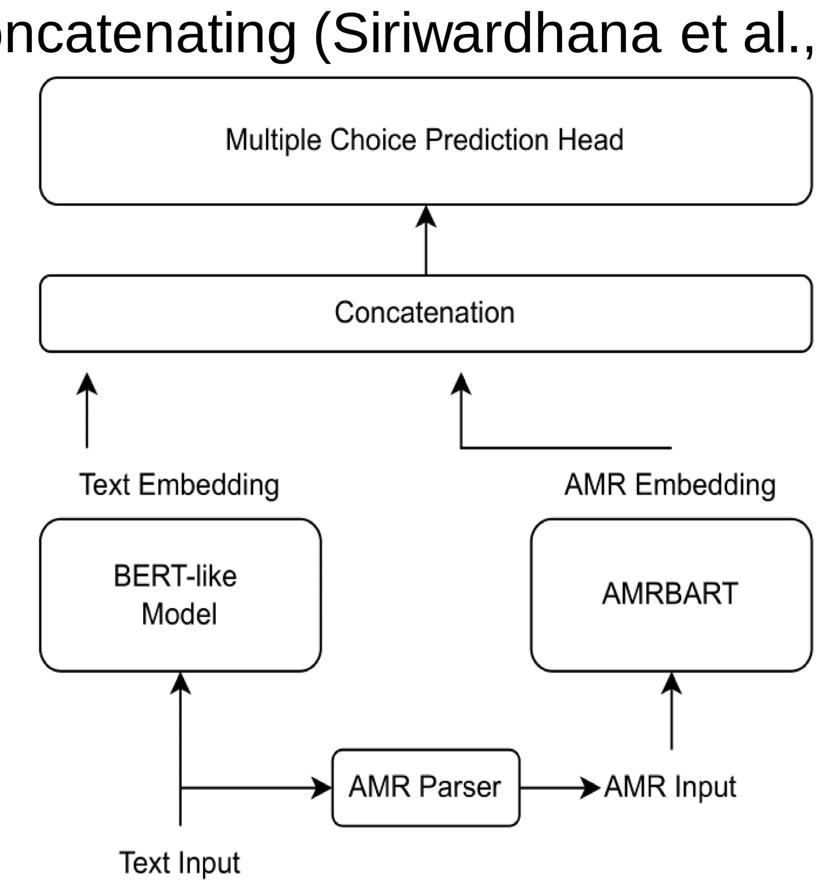
Similarity-Based Baseline (Bonial et al., 2020): Smatch (Cai and Knight, 2013) to measure overlap between two AMRs.

Encoding Linearised AMR with a text PLM (Mager et al., 2020): LegalBERT with AMR linearisation and simplification by Konstas et al. (2017), and adapters (Ribeiro et al., 2021).

AMRBART (Bai et al., 2022): based on BART (Lewis et al., 2020), further **pre-trained** on linearised AMR graphs.



Fusion Model: combine AMR and text input by concatenating (Siriwardhana et al., 2020).



Results

CaseHOLD: Fusion model performs similarly to the text-only BART model. Other AMR models perform worse than the text baselines.

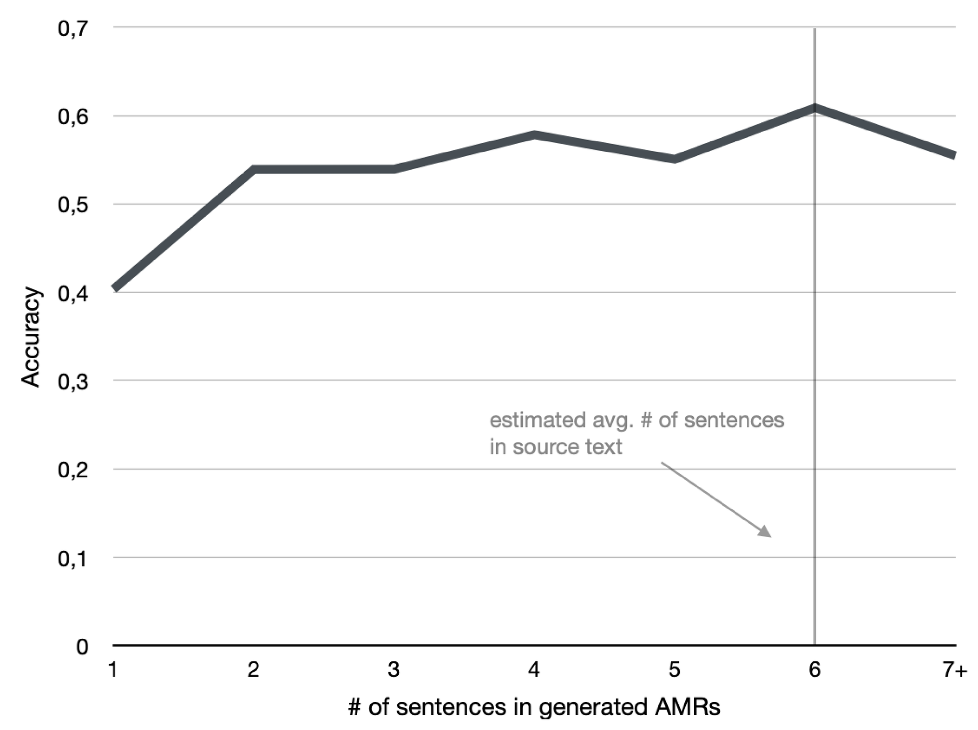
LogiQA: AMR models underperform.

Model	Input	Model Size	Accuracy
LegalBERT$_{\texttt{SMALL}}$	Text	35M	0.72[12]
LegalBERT$_{\texttt{SMALL}}$ + adapter	Text	35M	0.73
BART$_{\texttt{BASE}}$	Text	139M	0.74
LegalBERT$_{\texttt{SMALL}}$ + adapter	AMR (linearised and simplified)	35M	0.53
Smatch Model	AMR (Penman)	-	0.34
AMRBART$_{\texttt{BASE}}$	AMR (Spring prepr.)	142M	0.51
Fusion Model	Text and AMR (Spring prepr.)	252M	0.74

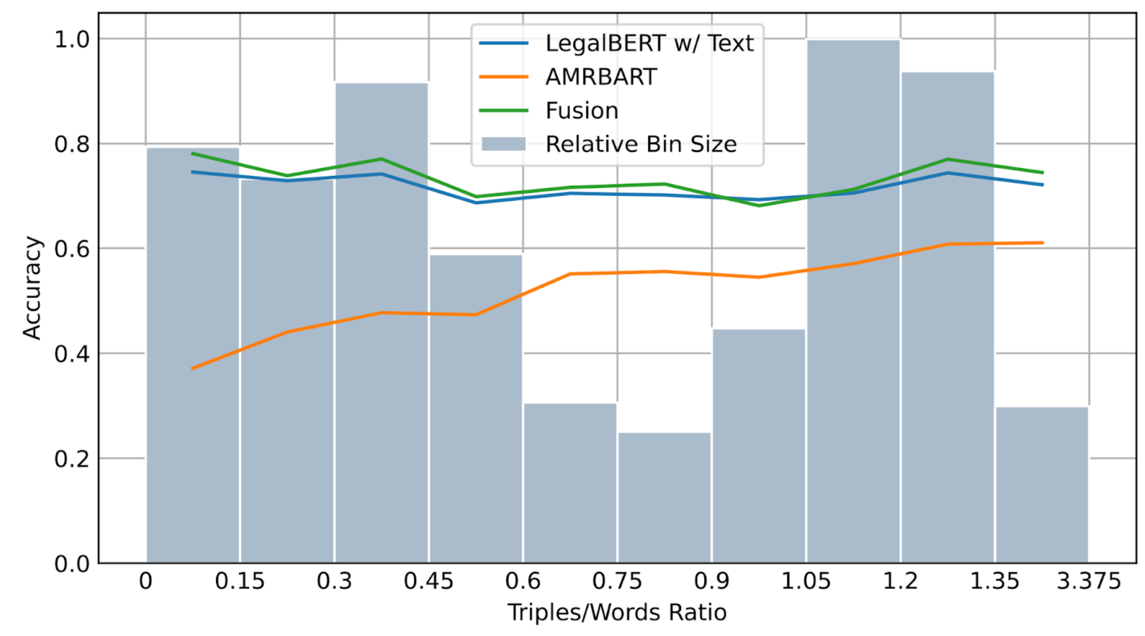
CaseHOLD performance. The Fusion Model uses LegalBERT to encode the text and AMRBART to encode the linearised AMR.

Error Analysis: Parser Quality

*N*early 50% of sentences missing in the generated AMR graphs.



Accuracy increases with the triples/words ratio: parser quality has a great impact on performance.



Missing information in AMR. Ratio between triples in parsed AMR and words in the text and their average accuracy for AM- RBART. As comparison the performance of the same instances for LegalBERT w/Text and Fusion model is shown. The data is taken from CaseHOLD test results.

Complementary abilities

BERT solves 65 instances consistently.
AMRBART solves 76 instances consistently.
The overlap is 13 instances.
The AMR model has learnt different knowledge about logical relations compared to the text-only models.