AMR Beyond the Sentence: the Multi-sentence AMR corpus

Tim O'Gorman[1]   Michael Regan[3]   Kira Griffitt[2]   Ulf Hermjakob[4]   Kevin Knight[4]   Martha Palmer[1]

[1]University of Colorado Boulder   [2]Linguistic Data Consortium
[3]University of New Mexico   [4]ISI-USC

Adding Coreference and Implicit Roles to AMR

The Abstract Meaning Representation format[1] already captures some pragmatic reference information, such as within-sentence coreference and entity linking.

e.g. *Madonna said she would sing*:
```
(s / say-01
    :ARG0 (p / person :wiki "Madonna_(performer)" :name (n / name :op1 "Madonna"))
    :ARG1 (s2 / sing-01
        :ARG0 p))
```
The output is also a computationally useful **directed acyclic graph format**. But for many purposes, we want a graph of a **document**, not just a sentence.

Goals and Decisions

- **Implicit roles:** Many important semantic roles aren't explicitly stated. Languages with more zero-anaphora shouldn't have less connected documents.
- **Linking to single variables:** Unlike normal coreference, there are no spans -- just equivalence classes of AMR variables
- **Bridging:** add links like set/member and part/whole, capture things like discourse anaphora.
- **All Coreferent Mentions:** Do both event and entity coreference, specific mentions and relevant ideas.

Implicit roles across sentences

Many semantic roles aren't explicit in a sentence, but **are implicit in context**.

Using the Propbank lexicon and AMR, we know which possible numbered arguments are not explicit. These are temporarily added to graphs during annotation (see example). **These can then be annotated as normal coreference relations.**

Lots of Tricky Implicit Role Situations

- **Reduced conversational forms:** ``He's trying''. ``And $\emptyset_{he}$ failing $\emptyset_{action}$.''
- **Dropped subjects:** ``Been to Cyprus recently?''
- **Arguments of adjectives:** ``Very true.'', ``damn right''
- **Arguments of nominals/nominalizations:** ``I will stick around until the end $\emptyset$'', ``my answer $\emptyset_{question}$ is no one''

Kinds of Annotation Issues

- **Granularity:** Annotators label different, related antecedents.
- **Specific vs generic experiencers:** E.g. experiencer of ``that's annoying'' -- when to link to prior mentions.
- **Causal Inference:** Absent arg0 labels can pull in all the issues you run into with casuation annotation.
- **Recoverable low-salience mentions:** E.g. ``He seems $\emptyset_{to\_me}$ happy'' or ``I went to the store $\emptyset_{from\_home}$''

Bridging Relations

Annotators label set/member and part/whole relations on the data as well. These are primarily cleanly anaphoric bridging relations ``one even punched him'' or ``the door slammed shut''.

We also allowed annotation of more discourse-oriented set-member relations, e.g. the referents of ``**that** got me thinking'', which can encompass multiple AMR heads.

Corpus Profile and Release

The data is primarily **English Discussion Forum data**. Will be released with the **2018 AMR public release** through LDC.

	Train(Single)	Analysis(Double)	Test(Single)
Files	284	43	9
AMRs	7826	588	201
Tokens	122000	8200	3700
Coreference Chains	3810	381	87
Implicit Roles	2386	371	67
Bridging Relations	1792	160	54

Table 1: Basic corpus statistics

Annotation Methodology

AMRs were represented in a separate coreference toolkit (Anafora; [3]) with possible implicit roles added (in turquoise) and color coding added. As AMR nodes are already ``spans'', adding a mention to a coreference chain is a single click:

bolt-eng-DF-170-181118-8874533_0003.23 ::: He expressed his condolences to President Barack Obama.

```
(e / express-01
    :ARG0 (h / he)
    :ARG1 (c / condolences
        :poss h)
    :ARG2 (p / person :wiki "Barack_Obama" :name (n / name :op1 "Barack" :op2 "Obama")
        :ARG0-of (h2 / have-org-role-91
            :ARG2 (p2 / president)
            :ARG1 (i / implicit-role :op1 "organization")
            :ARG3 (i2 / implicit-role :op1 "description_of_responsibility"))))
```

A Big Example

1. So was wondering what's the best thing to do?
```
(w / wonder-01
    :arg0 i
    :arg1 (t / thing          ← set
        :ARG1-of (d / do-02
            :ARG0 (i / implicit-role )
            :ARG2 (i2 / implicit-role ))
        :ARG1-of (h / have-degree-91
            :ARG2 (g / good-02 )
            :ARG3 (m / most )))
    :mod (s / so ))
```

3. See if I can speak to a counsellor?
```
(s / see-01          ← member
    :ARG0 i
    :ARG1 (s2 / speak-01
        :ARG0 (i / i )
        :ARG2 (p / person
            :ARG0-of (c / counsel-01
                :ARG1 i ))
        :ARG1 (i2 / implicit-role ))
    :ARG1-of (p2 / possible-01 )
    :polarity (a / amr-unknown ))
```

2. Speak to a doctor about it and see if anti-anxiety pills could take the edge off of things?
```
(a / and          ← member
    :op1 (s / speak-01
        :ARG0 (y / you )
        :ARG1 (i / it )
        :ARG2 (d / doctor ))
    :op2 (s2 / see-01
        :ARG0 y
        :ARG1 (t / take-off-07
            :ARG0 (p / pill
                :ARG0-of (c / counter-01
                    :ARG1 (a2 / anxiety )))
            :ARG1 (e / edge )
            :ARG1-of (p2 / possible-01 )
            :source (t2 / thing ))
    :polarity (a3 / amr-unknown ))
```

2

Annotation Quality and Discussion

- **Coreference Annotation** are roughly similar to IAA for other annotations -- 69.86 CoNLL F1
  - Not comparable to other tasks -- doing event coreference, omits easy within-sentence coreference
  - See paper for details -- most double-annotation used identical within-sentence AMRs.

- **Implicit role annotation** low agreement (κ=0.59), similar to simpler focused implicit role annotation tasks (κ=0.64 for [4])

- **Low agreement for bridging annotation.**

You should train systems on this!

This has the advantage of resulting in a **full graph of document**.

Normal AMR Evaluation (review of SMATCH)

AMR is evaluated by evaluating the ``triples'' against a gold AMR. E.g. the top left AMR can be treated as a list:

- (a , instance-of, arrive-01)
- (a , ARG1, h)
- (h , instance-of, he)
- (a , time, d)
- etc.

As variable labels are arbitrary, SMATCH metric [2] assumes you find the **optimal mapping of predicted variables to gold variables**, and then score you on the F1 of those mapped triples.

Document-level SMATCH

We propose that MS-AMR outputs can be scored using *document-level SMATCH*; merge all AMRs in a document, and apply the same SMATCH metric!

This requires a set of assumptions (details in the paper) for how to merge non-identical mentions. In general, we **removed clearly redundant information**.

Baselines and Consequences

This means a **single score for the semantic representation of the document**.

AMR system	Coreference	Double	Test
CAMR	CoreNLP	53.6	44.0
gold	none	78.5	80.6
gold	CoreNLP	80.1	82.9
gold	human	87.3	

Table 2: Baseline system performance shows this strongly penalizes within-sentence AMR error

(Caveat: We used the hill-climbing code of [2] for scoring, but this is very memory intensive for longer documents).

Direction of Intended Work

This is viewed as **one step towards document-level representations over AMR**, which adds the important coreference in AMR style. It is hoped that this can be a starting point for adding layers with **temporal, aspectual information, factuality of events, and rhetorical/information structure**.

We believe this approach can also be extended to other languages (*this relates to ongoing work in Chinese AMR at Brandeis*).

References

[1] Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider.
Abstract meaning representation for sembanking.
In *Proceedings of Linguistic Annotation Workshop*, 2013.

[2] Shu Cai and Kevin Knight.
Smatch: an Evaluation Metric for Semantic Feature Structures.
In *ACL (2)*, pages 748--752, 2013.

[3] Wei-Te Chen and Will Styler.
Anafora: a web-based general purpose annotation tool.
In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, volume 2013, page 14. NIH Public Access, 2013.

[4] Matthew Gerber and Joyce Y. Chai.
Semantic role labeling of implicit arguments for nominal predicates.
Computational Linguistics, 38(4):755--798, 2012.

[5] Tatjana Moor, Michael Roth, and Annette Frank.
Predicate-specific annotations for implicit role binding: Corpus annotation, data analysis and evaluation experiments.
In *Proceedings of the 10th International Conference on Computational Semantics*, 2013.