

Transition-based Chinese AMR Parsing

Chuan Wang¹ Bin Li² Nianwen Xue¹

¹Michtom School of Computer Science
Brandeis University

{cwang24;xuen}@brandeis.edu

²School of Chinese Language and Literature
Nanjing Normal University
libin.njnu@gmail.com

Abstract

This paper presents the first AMR parser built on the Chinese AMR bank. By applying a transition-based AMR parsing framework to Chinese, we first investigate how well the transitions first designed for English AMR parsing generalize to Chinese and provide a comparative analysis between the transitions for English and Chinese. We then perform a detailed error analysis to identify the major challenges in Chinese AMR parsing that we hope will inform future research in this area.

1 Introduction

Abstract Meaning Representation (AMR) (Banasescu et al., 2013) is a semantic representation where the meaning of a sentence is encoded as a rooted, directed and acyclic graph. AMR parsing has received a significant amount of attention in the NLP research community. Since the release of the AMR bank a number of AMR parsers have been developed in recent years (Flanigan et al., 2014; Wang et al., 2015b; Artzi et al., 2015; Pust et al., 2015; Peng et al., 2015; Zhou et al., 2016; Goodman et al., 2016). The initial benefit of AMR parsing has also been demonstrated in various downstream applications such as Information Extraction (Pan et al., 2015; Huang et al., 2016), Machine Comprehension (Sachan and Xing, 2016), and Natural Language Generation (Flanigan et al., 2016; Butler, 2016).

In this paper, we present the first AMR parser built using the Chinese AMR Bank (Li et al., 2016). We adopt the transition-based parsing framework first proposed for English (Wang et al., 2015b, 2016), where AMR parsing is modeled as a dependency tree to AMR graph transformation using a set of linguistically motivated actions. We briefly describe the Chinese AMR Bank in Section 2, present the transition-based Chinese AMR

parsing model in Section 3, report and analyze experimental results in Section 4, and conclude our paper in Section 5.

2 The Chinese AMR Bank

In our experiment, we use a pre-release version of the Chinese AMR Bank (Li et al., 2016)¹ consisting of 10,149 sentences extracted from the Chinese Treebank (CTB) 8.0 (Xue et al., 2005)², which mainly consists of Chinese texts of web logs and discussion forums. The average sentence length is 22.43 words.

Similar to English, the Chinese AMRs are also represented as rooted, directed and acyclic graphs that share the *abstract* concepts and relations used in the English AMR Bank. The sense-disambiguated predicates are drawn from the frame files developed for the Chinese Propbank (Xue and Palmer, 2009), just as the sense-disambiguated predicates in the AMR Bank are drawn from the Propbank (Palmer et al., 2005). About 47% of the 10,149 sentences have re-entrancies, meaning that they have a graph structure that cannot be represented with a tree representation.

3 Transition-based AMR Parsing

In a transition-based AMR parsing framework an input sentence is first parsed into a dependency tree and then transformed into an AMR graph via a series of transitions formulated as “actions”. The full set of actions are summarized in Table 1, and we refer the reader to (Wang et al., 2015b,a) for details regarding the training procedure and decoding algorithm. Note that NEXT-EDGE- l_r and NEXT-NODE- l_c are action to label the current node

¹<http://www.cs.brandeis.edu/~clp/camr/camr.html>.

²Available at <https://catalog.ldc.upenn.edu/LDC2013T21>.

or current edge, where the candidate label is defined as a parameter to the action. The $INFER-l_c$ (ifr) is devised to predict *abstract concepts* that are not aligned to any specific word in a sentence. The rest of the actions are responsible for transforming the structure of the partial graph.

4 Experiments

In this section, we present experiments designed to probe the behavior of our Chinese AMR parser, and where appropriate, compare it to its English counterpart. We also devise several ablation tests to further investigate the errors produced by our Chinese AMR parser to gain insight that can be used to guide future research.

4.1 Experiment Settings

We use the 10,149 sentences from the Chinese AMR Bank and split the data according to their original CTB8.0 document IDs, where articles 5061-5558 are used as the training set, articles 5000-5030 are used as the development set and articles 5031-5060 are used as the test set. The train/development/test ratio in this dataset is 7608/1264/1277. As the data are drawn from the Chinese Treebank where words are manually segmented, we will simply use the gold segmentation in our experiments. We then process the whole Chinese dataset using the Stanford CoreNLP (Manning et al., 2014) toolkit to get the POS and Named Entity tags. To get the dependency parse for the Chinese data, we use the transition-based constituent parser in (Wang and Xue, 2014) to first parse the Chinese sentences into constituent trees, which are then transformed into dependency trees using the converter in the Stanford CoreNLP toolkit. Note that this Chinese constituent parser also uses the Chinese Treebank 8.0 to train its model. To avoid training on the parser on AMR test set, we train the constituent parser using a 10-fold cross-validation with each fold parsed using a model trained on the other 9 folds. In order to compare results between Chinese and English, we also train an English AMR parsing model on the LDC2015E86 dataset used in SemEval 2016 Task 8 with the standard split 16833/1368/1371 and the English AMR parser, CAMR, is utilized to train the English model. All the AMR parsing results are evaluated by the

Smatch toolkit (Cai and Knight, 2013)³.

4.2 Action Distribution

Before we train the parser, we first perform a quantitative comparison of the actions that are invoked in English and Chinese AMR parsing. We run the *oracle* function separately on the training data of both languages and record the distribution of the actions invoked, as shown in Figure 1. Note that without any modification of the action set designed for English, the “pseudo-gold” graphs generated by the *oracle* function have reached F1-score of 0.99 when evaluated against gold Chinese AMR graphs, and this indicates that the action set is readily generalizable to Chinese. The histograms in Figure 1 shows the distribution of action types for both English and Chinese. We leave out the $NEXT-EDGE-l_r$ and $NEXT-NODE-l_c$ actions in the histogram as they do not trigger structural transformations like other actions, and thus are not our point of interest.

In Figure 1 we can see that there is a large difference in action distribution between Chinese and English. First of all, there are a lot fewer DELETE-NODE actions applied in the dependency-to-AMR transformation process for Chinese, which indicates that in Chinese data there is a smaller percentage of “stop words” that do not encode semantic information. Also, in the Chinese data, more $INFER-l_c$ actions are invoked than in English, implying that Chinese AMRs use more abstract concepts that don’t align to any word token.

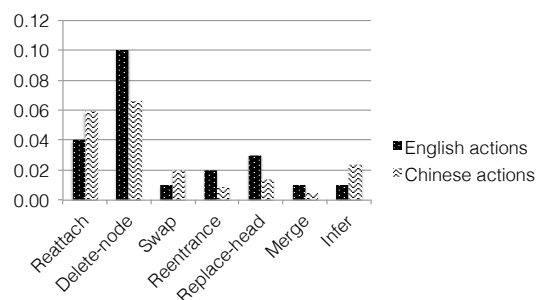


Figure 1: Action distribution on English and Chinese

To further investigate the different linguistic patterns associated with each action in the two languages, for each action type t , we randomly sample 100 sentences in which action t is invoked for both English and Chinese. We then conduct

³<http://alt.qcri.org/semeval2016/task8/data/uploads/smatch-v2.0.2.tar.gz>

Action	Description
NEXT-EDGE- l_r (ned)	Assign the current edge with edge label l_r and go to next edge.
SWAP- l_r (sw)	Swap the current edge, make the current dependent as the new head, and assign edge label l_r to the swapped edge.
REATTACH $_k$ - l_r (reat)	Reattach current dependent to node k and assign label l_r to new edge.
REPLACE-HEAD (rph)	Replace current head node with current dependent node.
REENTRANCE $_k$ - l_r (reen)	Add another head node k to current dependent and assign label l_r to edge between k and current dependent.
MERGE (mrg)	Merge two nodes connected by the edge into one node.
NEXT-NODE- l_c (nnd)	Assign the current node with concept label l_c and go to next node.
DELETE-NODE (dnd)	Delete the current node and all edges associated with current node.
INFER- l_c (ifr)	Insert concept with label l_c between current node and its parent.

Table 1: Action set in Chinese AMR Parsing, where k, l_r, l_c are parameters of the action.

a detailed analysis of the sampled data. We find that MERGE is mostly responsible for combining spans of words to form a named entity in English parsing. However, in Chinese AMR parsing, in addition to forming named entity concepts, MERGE also handles a large portion of split verb constructions. A “split verb” is a linguistic phenomenon in Chinese in which the characters in a multi-character verb are split into two discontinuous parts by other lexical items. For example, in (1), the sentence has a split verb “帮 /help ... 忙 /business” that are merged by the MERGE action to form the AMR concept “帮忙-01”, as shown in Figure 2.

In the cases of SWAP and REPLACE-HEAD, we notice that the linguistic patterns associated with the two actions are mostly consistent across the two languages. For example, as we already mentioned, the SWAP action is used to handle the structural divergence between the dependency tree and AMR graph of coordination constructions. This holds for both English and Chinese. Similarly, the REPLACE-HEAD action is designed to resolve the structural divergence between the dependency tree and AMR graph of propositional phrases. Based on our analysis of sampled data, the REPLACE-HEAD action resolves the same dependency-AMR divergence in Chinese AMR parsing.

- (1) 他¹帮²了³我⁴很⁵大⁶的⁷忙⁸。
 He helped PAST me very big DE business
 “He helped me a lot.”

Being able to identify the linguistic environment for each action helps us understand what the parser actually does when actions are applied. More importantly, making the relation between the

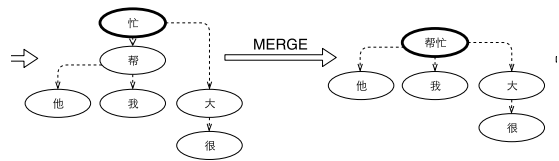


Figure 2: AMR for Example (1)

linguistic structure and the parser actions transparent is crucial to our ability to devise effective features for the parsing model which directly impacts the performance of the parser. For example, knowing that the MERGE action is responsible for producing concepts from split verb constructions helps us understand the need to design character-level features in addition to features targeting named entities.

4.3 Main results for Chinese AMR Parsing

Using the configuration in Section 4.1, we train our Chinese AMR parser with 5 iterations and report results on both the development and test set.

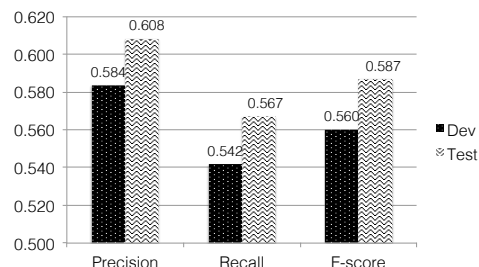


Figure 3: Parsing performance on the development and test set

Figure 3 presents the parsing performance on the development and test set in terms of the

Smatch score. Compared with the state of the art in English AMR parsing, which is in the high 60 percentage points (May, 2016), this initial parsing performance here is very strong, considering the model is trained on a smaller training set. The Chinese AMR parsing model also does not benefit from the more extensive feature engineering that has been done for English AMR parsing. For example, the English AMR parser, CAMR, uses semantic roles and coreference features that are not available to the Chinese AMR parser. The other important factor is that most of the Chinese linguistic analyzers (dependency parsers, named entity taggers, etc.) have a lower accuracy than their English counterparts, and when used as preprocessors for the AMR parser, could further disadvantage the Chinese AMR parsing model.

4.4 Fine-grained Error Analysis

So far, all of our experiments are evaluated using the Smatch score, where only precision, recall and F-score are reported based on the overall performance of the parser. To gain more insights, we further break down the Smatch score and report the performance for each component using the evaluation tool from Damonte et al. (2017). The evaluation tool examines different aspects of the AMR parsing result through different ablation tests that we summarize as follows. The detailed description of the ablation test can be found in Damonte et al. (2017).

- **Unlabeled.** Smatch score obtained by ignoring the edge labels (relation).
- **No WSD.** Smatch score without the word sense disambiguation.
- **NP (Noun Phrase)-only.** Only evaluating the noun phrases.
- **Reentrancy.** Only evaluating reentrancy edges.
- **Concepts.** Evaluating the node labels (concept).
- **Named Ent.** Named entity evaluation.
- **Negation.** Evaluation on negation detection.
- **SRL.** Semantic Role Labeling, which only evaluates triples in AMR that have relations starting with *:ARG*.

Note that we simply ignore the wikification evaluation as Chinese AMRs do not have wikification annotation at the current stage.

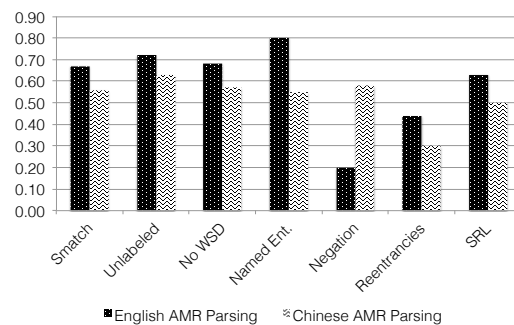


Figure 4: Fine-grained AMR parsing evaluation on dev

Figure 4 shows the performance breakdown on the Chinese and English development sets, where we can see that the overall performance gap between English and Chinese is around 0.11 Smatch score and there is a similar gap for **Unlabeled**, **No WSD** and **SRL** evaluations. However, the largest performance comes from **Named Ent.**, where the F-score for Chinese is 0.55 which is 0.25 lower than English. This indicates that named entity is one of the bottlenecks in Chinese AMR parsing. This indicates that improving named entity recognition, either as a preprocessing step or as an integral part of the parsing model, is crucial to Chinese AMR parsing.

5 Conclusion

We present the first Chinese AMR parser trained on the Chinese AMR Bank. We show that a transition-based AMR parsing framework first proposed for English is general enough to handle the linguistic phenomena in Chinese and has produced a strong baseline that future research can build on. In addition, we perform a detailed comparative analysis of the transition distributions for English and Chinese as well as errors in Chinese AMR parsing that we hope will inform future Chinese AMR parsing research.

Acknowledgements

We want to thank the anonymous reviewers for their suggestions. We also want to thank Yuan Wen, Lijun Bu and Li Song on their contributions to the creation of the Chinese AMR corpus. The second author would also like to acknowledge support by the National Science Foundation of China (61772278, 61472191).

References

- Yoav Artzi, Kenton Lee, and Luke Zettlemoyer. 2015. [Broad-coverage CCG semantic parsing with AMR](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1699–1710, Lisbon, Portugal. Association for Computational Linguistics.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. [Abstract meaning representation for sembanking](#). In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria.
- Alastair Butler. 2016. [Deterministic natural language generation from meaning representations for machine translation](#). In *Proceedings of the 2nd Workshop on Semantics-Driven Machine Translation (SedMT 2016)*, pages 1–9. 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. 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.
- Jeffrey Flanigan, Chris Dyer, A. Noah Smith, and Jaime Carbonell. 2016. [Generation from Abstract Meaning Representation using tree transducers](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 731–739. Association for Computational Linguistics.
- Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. 2014. [A Discriminative Graph-Based Parser for the Abstract Meaning Representation](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1426–1436, Baltimore, Maryland. Association for Computational Linguistics.
- James Goodman, Andreas Vlachos, and Jason Naradowsky. 2016. [Noise reduction and targeted exploration in imitation learning for abstract meaning representation parsing](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1–11. Association for Computational Linguistics.
- Lifu Huang, Taylor Cassidy, Xiaocheng Feng, Heng Ji, R. Clare Voss, Jiawei Han, and Avirup Sil. 2016. [Liberal event extraction and event schema induction](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 258–268. Association for Computational Linguistics.
- Bin Li, Yuan Wen, Q. U. Weiguang, Lijun Bu, and Nianwen Xue. 2016. [Annotating the little prince with chinese amrs](#). In *Linguistic Annotation Workshop Held in Conjunction with ACL*, pages 7–15.
- Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. [The stanford corenlp natural language processing toolkit](#). In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland. Association for Computational Linguistics.
- Jonathan May. 2016. Semeval-2016 task 8: Meaning representation parsing. *Proceedings of SemEval*, pages 1063–1073.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–105.
- Xiaoman Pan, Taylor Cassidy, Ulf Hermjakob, Heng Ji, and Kevin Knight. 2015. [Unsupervised entity linking with abstract meaning representation](#). In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1130–1139.
- Xiaochang Peng, Linfeng Song, and Daniel Gildea. 2015. [A synchronous hyperedge replacement grammar based approach for AMR parsing](#). In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*.
- Michael Pust, Ulf Hermjakob, Kevin Knight, Daniel Marcu, and Jonathan May. 2015. [Parsing English into abstract meaning representation using syntax-based machine translation](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1143–1154, Lisbon, Portugal. Association for Computational Linguistics.
- Mrinmaya Sachan and Eric Xing. 2016. [Machine comprehension using rich semantic representations](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 486–492. Association for Computational Linguistics.
- Chuan Wang, Sameer Pradhan, Xiaoman Pan, Heng Ji, and Nianwen Xue. 2016. [CAMR at semeval-2016 task 8: An extended transition-based AMR parser](#). In *Proceedings of the 10th International*

Workshop on Semantic Evaluation (SemEval-2016), pages 1173–1178, San Diego, California. Association for Computational Linguistics.

Chuan Wang, Nianwen Xue, and Sameer Pradhan. 2015a. Boosting transition-based AMR parsing with refined actions and auxiliary analyzers. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers)*, pages 857–862.

Chuan Wang, Nianwen Xue, and Sameer Pradhan. 2015b. [A transition-based algorithm for AMR parsing](#). In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 366–375, Denver, Colorado. Association for Computational Linguistics.

Zhiguo Wang and Nianwen Xue. 2014. [Joint pos tagging and transition-based constituent parsing in chinese with non-local features](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 733–742, Baltimore, Maryland. Association for Computational Linguistics.

Nianwen Xue and Martha Palmer. 2009. Adding semantic roles to the chinese treebank. *Natural Language Engineering*, 15(1):143–172.

Nianwen Xue, Fei Xia, Fudong Chiou, and Martha Palmer. 2005. The penn chinese treebank: Phrase structure annotation of a large corpus. *Natural Language Engineering*, 11(2):207–238.

Junsheng Zhou, Feiyu Xu, Hans Uszkoreit, Weiguang QU, Ran Li, and Yanhui Gu. 2016. [AMR Parsing with an Incremental Joint Model](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 680–689, Austin, Texas. Association for Computational Linguistics.