

Annotating *the Little Prince* with Chinese AMRs

Bin Li, Yuan Wen, Lijun Bu, Weiguang Qu

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

Nianwen Xue

Computer Science Department
Brandeis University
Waltham, MA, USA
nwxue@brandeis.edu

Abstract

Abstract Meaning Representation (AMR) is an annotation framework in which the meaning of a full sentence is represented as a rooted, acyclic, directed graph. In this paper, we describe a pilot project in which we develop specifications for the annotation of a Chinese AMR corpus: the Chinese translation of *the Little Prince*. The interagreement smatch score between the two annotators is 0.83. We also propose to integrate alignment into Chinese AMR annotation.

1 Introduction

Abstract Meaning Representation (AMR) is an annotation framework designed to capture the “meaning” of a sentence with a single rooted, acyclic¹, directed graph (Banarescu et al., 2013), departing from previous practices of performing partial semantic annotation that focuses on one component of the sentential meaning at a time. For example, Propbank (Palmer et al., 2005; Xue and Palmer, 2009) and NomBank (Meyers et al., 2004) annotations focus on the predicate-argument structure of verbs and predicative or relational nouns. The annotation is done on a predicate basis and the resulting annotation may not necessarily form a fully connected structure for the entire sentence. The practice was necessary as a first attempt to annotate a key aspect of sentential meaning and contributed to a high-quality corpus that has spurred research in automatic Semantic Role Labeling (SRL) (Gildea and Jurafsky, 2002; Pradhan et al., 2004; Xue and Palmer, 2004; Palmer et al., 2010) and downstream applications. This annotation strategy has been adopted for the predicate-

¹Technically, about 0.3% sentences are cyclic in AMR (Banarescu et al., 2015).

argument structure annotation of other languages as well (Xue and Palmer, 2009; Zaghouani et al., 2012).

As we gain more insights on the sentence meaning from annotating individual meaning components, annotating the meaning for the entire sentence becomes a logical next step. The AMR annotation project is such an attempt, along with other similar efforts such as the Universal Dependency annotation project (Nivre, 2015) and the Semantic Dependency Parsing effort (Oepen et al., 2014).

One salient characteristic of AMR annotation is that it abstracts away from elements of surface syntactic structure such as word order and morpho-syntactic markers. Since word order and morpho-syntactic variations account for much of the cross-linguistic variations, this makes the AMR annotation framework more portable across languages, as the preliminary AMR annotation on Chinese and Czech has demonstrated (Xue et al., 2014).

Another consequence of such “decoupling” from the syntactic structure of a sentence is that the AMR annotation framework gives us more freedom in how we handle cases of syntax-semantic mismatch. Words that do not contribute to the meaning of a sentence (e.g., infinitive “to” in English) are left out of the AMR annotation. In light verb constructions such as “take a bath”, since the light verb “take” is semantically impoverished if not vacuous, it is also left out of the AMR annotation. Some discontinuous constructions such as “if... then” can be collapsed into a single relation “:condition”.

With this freedom comes responsibility. Since an annotator is free to drop a word or map discontinuous patterns onto single AMR concepts or relations, for the sake of annotation consistency, it is important to provide detailed annotation speci-

fications for how certain constructions are handled so that they are consistently followed by all annotators.

Since words that “do not carry meaning” are left out of the AMR of a sentence, for purpose of automatic AMR parsing, it is also important to explicitly represent the correspondence between word tokens in the sentence to the concepts and relations in its AMR, that is, the alignment between the input sentence and its AMR presentation. In English AMR annotation (Banarescu et al., 2013), this alignment is performed as a separate step and it is not integrated into AMR annotation itself. For example, to develop a word-to-AMR-concept aligner as the first step of their AMR parser, Flanigan et al. (2014) annotated the alignment between word tokens and AMR concepts and relations for a small corpus which they use as the development set to extract alignment rules. The alignment accuracy of this aligner is about 90%. Pourdamghani et al. (2014) developed an EM-based aligner that yields similar performance without any manual alignment. While these aligners may seem to be very accurate, a 10% error rate in alignment imposes a serious limitation on the overall AMR parsing accuracy.

In this paper we present the Chinese AMR (CAMR) annotation specifications that we use to annotate the Chinese translation of *the Little Prince*, which has 1,562 sentences. The CAMR annotation specifications are adapted from the AMR specification for English (Banarescu et al., 2015). We choose *the Little Prince* for our Chinese AMR annotation experiments because its English translation has already been annotated with AMRs and we can easily compare our AMR annotation with that of the English version for cross-linguistic studies. We also propose an integrated annotation approach in which the alignment step is integrated into the CAMR annotation process so that users of this corpus don’t have to create their own alignment. We show that the alignment process shares many of the characteristics of word alignment across languages.

The rest of this paper is organized as follows. In Section 2 we present an overview of the AMR annotation framework and the supporting lexical resources that we use for CAMR annotation. Section 3 describes how we extend AMR to handle Chinese-specific constructions and discourse relations in Chinese. In Section 4, we present re-

sults on our CAMR annotation experiments, and analyze sources of disagreement. In Section 5, we describe how we integrate word-to-concept alignment into the CAMR annotation process, and present a few alignment scenarios between word tokens and CAMR concepts. Section 6 gives a brief introduction of related work on the construction of semantic dependency databases. We conclude our paper in Section 7.

2 AMR Overview

In AMR annotation, each sentence is represented as a rooted, directed, acyclic graph in which the nodes are concepts and the edges are relations between the concepts. The backbone of an AMR graph is the predicate-argument structure of verbal or nominal predicates, though syntactic notions such as verbs and nouns are not part of the AMR vocabulary. AMR draws this aspect from the Proposition Bank (Palmer et al., 2005): the core argument roles (Arg0-Arg5) are defined in the PropBank frame files (Bonial et al., 2014), together with the senses of the predicates as different senses of a predicate have different argument structures. In addition, AMR also annotates named entities (person, location, company, etc.), relations between entities, time expressions, polarity and modality.

(1) The boy wants to go to New York.

```
w/want-01
:arg0 b/boy
:arg1 g/go-01
:arg0 b
:arg1 c/city
:wiki "New York"
:name (n/name :op1 "New" :op2 "York")
```

AMR is a graph, not just a tree, because it allows *reentrancy*. This happens when an argument is shared by more than one predicate as in control structures, or when there is coreference. For example, in (1) *boy* serves as the Arg0 of *want.01* as well as Arg0 of *go.01*. This is illustrated by the fact that the same concept ID *b* is used as an argument for both *want.01* and *go.01*.

In CAMR annotation, we generally adopted the vocabulary used for annotating English AMRs (Banarescu et al., 2013). However, we used argument labels and the predicate senses that are defined in the Chinese Proposition Bank (CPB) (Xue

and Palmer, 2009), which are very similar in convention to that of the English Proposition Bank. When developing the CAMR annotation specifications, most of our effort is expended on how to annotate some Chinese-specific constructions, which we will describe in detail in Section 3. These constructions are described in syntactic terms and are well recognized in Chinese linguistics.

3 Specifications for CAMR Annotation

When annotating CAMRs for *the Little Prince* corpus, we generally adopt the tagset for the (non-lexical) concepts and relations in the English AMR specifications. In English AMR, there are two types of concepts: *lexical concepts* that are grounded to word tokens in a sentence, and *abstract concepts* that are not linked to a specific lexical item. The former are typically lemmatized forms of word tokens with (e.g., *go.01* in (1)) or without word sense (e.g., *boy* in (1)). The latter are inferred from the context and are not tied to a specific lexical item (e.g., *city* in (1)). In the case of *city*, it can be viewed as a named entity tag for “New York”, but not all abstract concepts are named entity tags. There are also labels for numbers, time expressions, dates, as well as concepts that represent discourse relations. We obviously cannot use the lexical concepts in English AMR for CAMR annotation, but we have generally adopted the abstract concepts in AMR. Since Chinese has little inflectional morphology, in most cases the lexical concepts are just the words themselves.

(2) 男孩 想 去 纽约。
 nanhai xiang qu niuyue
 boy want go New York
 “The boy wants to go to New York.”

```
x/想-01
:arg0 x1/男孩
:arg1 x2/去-01
:arg0 x1
:arg1 x3/city
:wiki “纽约”
:name (n/name :op1 “纽约”)
```

The AMR relations include semantic roles as well as nominal relations. We use 6 labels for core arguments (Arg0-Arg5, as they are defined in the CPB frame files), and 42 labels for adjunctive ar-

guments and other semantic relations largely taken from the English AMR vocabulary. As shown in (2), the Chinese translation of “The boy wants to go to New York” is annotated similarly to its English counterpart. Notice that *city* is a non-lexical abstract concept that is not an actual word in the sentence.

Even though we use the same annotation conventions and mostly the same vocabulary as used in the English AMR, we still need to specify how to annotate Chinese-specific constructions that are not in English so that these constructions are consistently annotated. Due to the limitation of space, we only describe six such constructions: the number and classifier construction, the serial-verb construction, the headless relative construction, the verb-complement (VC) construction, the split verb construction, and reduplications. We will also discuss how to represent discourse relations in Chinese AMR, an area where there are significant adaptations.

3.1 Number and Classifier Construction

When a number modifies a Chinese noun or verb, it is always followed by a classifier. A classifier can be a measure like 吨, which has an equivalent word in English, “ton”. However, there is also another type of classifier which does not have an English equivalent. It serves as a cognitive measure of things and its meaning is hard to represent. The word 只 in (3) is such an example. It is also very idiosyncratic in the type of nouns it can modify. For example 只 can be used to modify sheep/goats or chickens, but not other types of animals such as cows or pigs. They are generally referred to as “individual classifiers” in Chinese linguistics. As AMR is concerned with the abstract meaning, we keep the measure words in the AMR representation but leave out the individual classifiers. Notice that the numbers are also normalized to Arabic numerals.

(3) 三 千 只 羊
 san qian zhi yang
 three thousand CL sheep
 “3000 sheep”

```
x/羊
:quant 3000
```

3.2 Serial-Verb Construction

Serial-verb constructions are very common in Chinese. They are characterized by having several verbs in a sequence, but it is sometimes very hard to determine the grammatical relations between them. For example, one verb can modify another or the two can be semantically equally important as in a coordinate structure. We choose to avoid making this hard decision for now for the sake of consistent annotation and consider these verbs to be in a coordination structure and create a non-lexical “and” concept to connect them.

- (4)他 走 过去 说...
ta zou guoqu shuo
he walk over say
“He walked over and said...”

a/and
:op1 x/走-01
:arg0 x1/他
:direction x2/过去
:op2 x3/说-01
:arg0 x1

3.3 Headless Relative Construction

Headless relative constructions are relative constructions without an explicit noun head. Syntactically it is realized as a relative clause followed by 的(DE), a function word that serves multiple purposes, one of which is as the marker of a relative clause. The dropped noun head of the relative clause could play any number of roles with regard to the verb in the relative clause: agent, patient, instrument, location, etc. When annotating the AMR, we use an abstract concept to represent the dropped noun head. In (5), for example, the abstract noun head is a “person”, and it is Arg0 of the verb 跳舞(dance).

- (5)跳舞 的 走 了。
tiaowu de zou le
dance DE leave COMPLETE
“The person who danced has left.”

x/走-01
:arg0 p/person
:arg0-of x1/跳舞-01

3.4 Verb-Complement Construction

A Verb-Complement (VC) construction is composed of a verb followed by another verb that indicates possibility, result, etc. The function word 得(DE) can optionally come between those two words. In AMR annotation, we make the meaning of the construction explicit using abstract concepts or relations. In (6), for example, the VC construction has a modal meaning, represented by “possible”, although there isn’t one word that specifically means possible. This meaning comes from the VC construction. In (7), there is a causal relationship between the two verbs 跑(pao) and 丢(diu), represented as a “cause” relation between the two verbs.

- (6)买 得 起 房子。
mai de qi fangzi
buy DE rise house
“Can buy a house.”

p/possible
:arg0 x/买-01
:arg1 x1/房子

- (7)它会 跑 丢 的。
ta hui pao diu de
it will run lost DE
“It will run and then get lost.”

x/会-02
:arg0 x1/丢-02
:arg1 x2/它
:cause x3/跑-01
:arg0 x2

3.5 Split Verb Construction

A “split verb” is a verb whose two parts can be separated by other linguistic material. 帮忙(help) is a typical example. When it is separated, it takes the form of a verb (帮) followed by an object (忙), separated by some modifiers. Its syntactic representation is quite a paradox: on the one hand, the semantics of the two parts are not separable, and it simply means “help” in its totality. On the other hand, it takes the form of a verb-object construction, and needs to be represented that way. AMR solves this paradox by just representing the entire construction as one concept, 帮忙, regardless of whether it is split or

not.

- (8) 地理学 帮 了 我 很 大 的 忙。
 dilixue bang le wo hen da de mang
 geography help PAST me very big DE business
 “Geography helped me a lot”

x/帮忙-01
:arg0 x1/地理学
:arg1 x2/我
:degree x3/大
:degree x4/很

3.6 Reduplications

There are two types of reduplications in Chinese. In the first type of reduplications (9a-9b), the reduplicated form has roughly the same meaning as the root form. The reduplication has either an aspectual meaning that the root form does not have (9a), or has its meaning intensified (9b). For the moment, we do not represent such subtle aspectual meanings or intensification. In the second type, however, the reduplicated form clearly adds meaning to its root form (9c-9d). We annotate their actual meaning by adding an abstract concept. The root form is in brackets in the following examples:

- (9a) 看看 (看)
 kankan (kan)
 “take a look”
- (9b) 干干净净 (干净)
 ganganjingjing (ganjing)
 “very clean”
- (9c) 人人 (人)
 renren (ren)
 “every person”
- (9d) 他天天 (天) 跑步。
 ta tiantian (tian) paobu
 he everyday (day) run
 “He runs everyday.”

x/跑步-01
:arg0 x1/他
:frequency x2/rate-entity-91
:arg3 x3/temporal-quantity :quant 1
:unit x4/天

3.7 Discourse Relations

One of the more significant adaptations to the AMR annotation specifications in CAMR is how discourse relations are annotated. Since for the moment AMR is a sentence-level meaning representation, here we only discuss intra-sentential discourse relations to the exclusion of inter-sentential relations.

- (10) 他 专 心 地 看 着, 随 后 又 说:
 ta zhuanxin de kan zhe suihou you shuo
 he carefully DE watch ASP then also say
 “我 不 要 。”
 wo bu yao
 I not want
 “He looked at it carefully, then he said: ‘I don’t want it.’”

x/temporal
:arg1 x2/看-02
:arg0 x3/他
:manner x4/专心-01
:arg2 x5/说-01
:mod x6/又
:arg0 x3
:arg1 x8/要-04 :polarity -
:arg0 x3

In English AMR, discourse relations are represented with a combination of abstract concepts (e.g., *and*, *or*, *contrast.01*) and relations (:*cause*, :*condition*, :*concession*, :*purpose*). In CAMR, we represent discourse relations with 10 concepts defined in the Chinese Discourse TreeBank (CDTB) (Zhou and Xue, 2015). These 10 discourse relations include *and*, *or*, which are also used in English AMR, but they also include *causation*, *condition*, *contrast*, *expansion*, *purpose*, *temporal*, *progression*, *concession*. Some of these discourse relations, e.g., *causation*, *condition*, *purpose*, and *concession* are treated as relations in AMR, and others are not part of the AMR vocabulary (*expansion*, *progression*, and *temporal*). In particular *temporal* represents the temporal precedence of a sequence of discourse segments while *progression* means one argument represents a progression from the other, in extent, intensity, scale, etc. As CDTB discourse relations are formally predicates that take two or more discourse segments as their arguments, the argument labels are meaningful as well. (10) is an example of temporal relation.

The arguments are arranged in chronological order, with Arg1 temporally preceding Arg2.

4 Annotation Experiments

We annotated all of the 1562 sentences in the Chinese version of *the Little Prince* following the CAMR specifications. Two linguistic undergraduate students were trained to perform the annotation. Each completed the annotation for all of the 1562 sentences, and the inter-agreement is calculated by Smatch toolkit (Cai and Knight, 2013). The overall Smatch score between the two annotators is 0.83.²

We analyzed the annotated data from the two annotators to see to what extent the graph representation of the meaning of a sentence is necessary. Out of the 1,562 sentences of the two annotated files, 576 and 548 of them have non-tree CAMR graphs in Chinese version of *the Little Prince*. This is in comparison with the 663 sentences that have non-tree AMR graphs in the English version, which is sentence-aligned with the Chinese version. The Pearson correlation of the sentences having non-tree graphs is around 0.56, indicating the bilingual semantic representation of the same sentence pair is similar. When one Chinese sentence has a non-tree graph structure, its English translation does too.

We also analyzed the sources of disagreement between the annotators. The causes of disagreement between the annotators are mostly from two sources. The first source of disagreement is that the two annotators have different interpretations of the same sentence. Disagreement also occurs when either or both annotators missed some concepts when annotating long sentences. The annotation tool that they use does not keep track of which words have been covered and which have not, and this contributes to this problem. Another issue with the annotation tool, which may or may not lead to disagreement in annotation, is that the annotator has to constantly shift between Chinese and English input modes to type Chinese characters for lexical concepts and English alphabets for abstract concepts. Motivated by the need to address these shortcomings of the annotation tool, as well as the need to incorporate word-to-concept alignment in the CAMR annotation process, we have redesigned the annotation framework, which

²The annotation data is available at <http://www.cs.brandeis.edu/~clp/camr/camr.html>.

we describe in detail in the next section.

5 Integrating Alignment to Annotation

When annotating *the Little Prince*, we followed the English AMR approach in which the concepts in CAMR are not aligned to word tokens in the sentence. Since AMR abstracts away from surface forms of a sentence, there is a non-trivial alignment problem between the AMR concepts and word tokens in a sentence. Some word tokens are considered to be devoid of meaning and are not represented in the AMR. Words that are not represented in AMR include determiners such as “a”, “an” and “the”, infinitive marker “to”. As we have discussed in Section 3, individual measure words are not represented in CAMR. On the other hand, abstract concepts in AMR are not grounded to any specific lexical item and are inferred from the context. In some cases, one word token is analyzed into multiple AMR concepts. For example, the English word “teacher” is represented in a similar way to “person who teaches” in the AMR. In other cases, multiple word tokens in a sentence may represent a single AMR concept. These word tokens do not even have to be contiguous. For example, Chinese parallel discourse connective 因为(because)...所以(therefore)... is mapped to one single discourse relation concept *causation*. So other than straight-forward one-to-one mappings between word tokens and AMR concepts, there are also complex alignment patterns such as one-to-zero, zero-to-one, one-to-many and many-to-one. In many ways, this is not too different from the word alignment between two languages. As we mentioned briefly above, having this alignment is important to AMR parsing, which is a process of mapping the input sentence to its AMR. Word-to-concept alignment is essential to this process, not unlike the role of word alignment to statistical machine translation.

The word-to-concept alignment is not integrated into the English AMR, mainly out of concern that it will slow down AMR annotation too much and it’s too complex to provide annotation to support for this. There was also hope that the alignment can be learned in an unsupervised manner with EM-based algorithms, just like word alignment between different languages can be learned automatically without the need for manual annotation. Although this expectation has been partially met in the work of Pourdamghani

et al. (2014), but we argue that an error rate of around 10% is too much of a deficit in the AMR parsing process, especially considering the level of difficulty in syntactic parsing where there is no alignment issue (or where there is perfect alignment between word tokens in the input sentence and terminal nodes in its parser tree).

We propose an annotation approach in which we integrate alignment with Chinese AMR annotation. Its basic idea is to use the index of a word token as the ID of the concept it aligns to in the AMR representation, thus establishing the alignment between the AMR concepts. We develop an annotation tool that allows an annotator to simply input the index of a word token in place of a concept during the annotation process. The tool will automatically retrieve the word token based on its index and generate the concept as well as the concept ID for it. This assumes that the tool does automatic lemmatization, which fortunately is very straightforward for Chinese where there is little inflectional morphology and the concepts are generally the same as their word forms. The tool also allows the annotator to revise the concept, and this is useful when a word does have inflections in a limited number of cases or when the word is misspelled. Words that do not correspond to a concept will of course receive no concept IDs and are not aligned. For abstract concepts that do not correspond to any word token, they are assigned IDs that have a value that is higher than the number of word tokens in the sentence. An example is given below. The IDs for the concepts are prefaced with “x”. Note that the *city* concept has an ID of “x5”, which does not correspond to any word token in the sentence.

(2)男孩¹想²去³纽约⁴。
 boy want go New York
 “The boy wants to go to New York.”

```
x2/想-01
:arg0 x1/男孩
:arg1 x3/去-01
:arg0 x1
:arg1 x5/city
:wiki “纽约”
:name (n/name :op1 x4/“纽约”)
```

The annotation tool also keeps track of which words in the sentence have been “covered” by the

AMR by highlighting words that the annotator has created concepts for. This is an especially useful feature when annotating long sentences, as it is very easy for the annotator to miss some words.

In addition to one-to-one, one-to-zero, and zero-to-one alignments, there are also one-to-many and many-to-one alignments between word tokens in a sentence and concepts in its AMR. The following is the AMR for Example (8) where one AMR concept is aligned to two word tokens that are also discontinuous. This is a case of split verbs that we discussed in Section 3. The word tokens are “帮...忙” and the AMR concept is simply 帮忙. Its ID is a concatenation of the indices of the two word tokens “x2_x8”.

(8)地理学¹帮²了³我⁴很⁵大⁶的⁷忙⁸。
 geography help PAST me very big DE business
 “Geography helped me a lot.”

```
x2_x8/帮忙-01
:arg0 x1/地理学
:arg1 x4/我
:degree x6/大
:degree x5/很
```

(11) is an example where one word is aligned to multiple concepts. This usually happens when the word has a complicated internal structure and each morpheme corresponds to an AMR concept. Chinese has very little derivational or inflectional morphology, but compounding is a highly productive morphological process.

(11)市场-分析-家¹说²
 shichang-fenxi-jia shuo
 market-analyze-expert say
 “The market analyst says...”

```
x2/说-01
:arg0 x1_5/家
:arg0-of x1_3_4/分析-01
:arg1 x1_1_2/市场
```

In (11), the compound word 市场分析师 (market analyst) has 5 characters and corresponds to three AMR concepts: 市场 (market), 分析 (analyze) and 家 (expert). In this case, we represent the alignment with the character offsets within the compound word. Notice that the character offsets, unlike the word indices, are not prefixed with “x”. This is how we differentiate word

indices from character offsets. For example, the concept ID for 市场 is “x1_1_2”, meaning that it is aligned with the first two characters of the first word. Similarly, 分析 is aligned with the third and fourth character of the first word, and its ID is “x1_3_4”. Finally, the ID for the concept 家 is “x1_5”, meaning that it is aligned with the fifth character of the first word.

In sum, this new annotation framework integrates word-to-concept alignment to the entire AMR annotation process. It also has the advantage of being able to keep track of word tokens that have been accounted for in the AMR (and those that have not), and helping to address the “missing word” problem in AMR annotation. In CAMR annotation, the annotator needs to switch back and forth between different input methods to input the lexical concepts that are composed of Chinese characters and the abstract concepts that are in English alphabets. The new annotation tool allows the annotator to use word indices and character offsets to input the lexical concepts and thus avoids the need to shift input modes, thus improving annotation efficiency.

6 Related work

Other than the AMR annotation project, other efforts aimed at annotating and parsing the semantic representation of a sentence with a graph structure include the semantic dependency parsing effort of Oepen et al. (2014). The difference is that the work of Oepen et al. (2014) does not abstract away from the surface word order and “semantically empty” words, and as far as we know, does not make use of abstract concepts as AMR does.

There are several efforts for constructing the Chinese semantic dependency resources. Li et al. (2004) reported parsing experiments on a one million word Chinese corpus annotated with semantic dependencies, but their dependency structure is tree-based rather than graph-based. Chen and Ji (2011) described a three thousand sentence corpus annotated with semantic graphs. Corpora annotated with semantic graphs also include those reported in Ding et al. (2014) and Zheng et al. (2014). These semantic resources vary in the types of semantic relations they use, but they all differ from the work we report here in that they define semantic relations between word tokens instead of abstract concepts.

7 Conclusion and Future Work

In this paper, we present our effort in developing specifications as well as an annotation tool for Chinese AMR (CAMR) annotation. We first annotate all 1,562 sentences of the Chinese translation of *the Little Prince* following the English AMR annotation framework while developing annotation guidelines to handle certain Chinese-specific syntactic constructions. We show that while we have achieved consistent annotation, there are shortcomings with this annotation approach. We then develop a new annotation tool and redesigned our annotation framework to address these shortcomings. In the future we plan to annotate additional data with this new framework.

Acknowledgements

We are grateful for the comments of the anonymous reviewers. This work is supported in part by National Natural Science Funds of China(No.61272221, 61170181) and Jiangsu University Philosophy and Social Science Fund(No.2016SJB740004).

References

- 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, August.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2015. *Abstract Meaning Representation (AMR) 1.2.2 Specification*. <https://github.com/amrisi/amr-guidelines/blob/master/amr.md>.
- Claire Bonial, Julia Bonn, Kathryn Conger, D. Hwang, Jena, and Martha Palmer. 2014. Propbank: Semantics of new predicate types. In *The 9th edition of the LanguageResources and Evaluation Conference*, pages 3013–3019, Reykjavik, Iceland, May.
- 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, August. Association for Computational Linguistics.
- Bo Chen and Donghong Ji. 2011. Chinese semantic parsing based on dependency graph and feature structure. In *International Conference on Electronic*

- and *Mechanical Engineering and Information Technology*, volume 4, pages 1731–1734, Aug.
- Yu Ding, Yanqiu Shao, Wanxiang Che, and Ting Liu. 2014. Dependency graph based chinese semantic parsing. *Lecture Notes in Computer Science*, 8801:58–69.
- 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*, pages 1426–1436.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–2883.
- Mingqin Li, Juanzi Li, Zuoying Wang, and Lu Dajin. 2004. A statistical model for parsing semantic dependency relations in a chinese sentence. *Chinese Journal of Computers*, 27(12):1679–1687.
- Adam Meyers, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika Zielinska, Brian Young, and Ralph Grishman. 2004. The nombank project: An interim report. In A. Meyers, editor, *Proceedings of the Workshop Frontiers in Corpus Annotation at HLT-NAACL 2004*, pages 24–31, Boston, Massachusetts, USA, May 2 - May 7.
- Joakim Nivre. 2015. Towards a universal grammar for natural language processing. *Computational Linguistics and Intelligent Text Processing*, 9041:3–16.
- Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Dan Flickinger, Jan Hajic, Angelina Ivanova, and Yi Zhang. 2014. Semeval 2014 task 8: Broad-coverage semantic dependency parsing. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 63–72, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–105.
- Martha Palmer, Daniel Gildea, and Nianwen Xue. 2010. Semantic role labeling. In Graeme Hirst, editor, *Synthesis Lectures on Human Language Technology Series*. Morgan & Claypoole Publishers.
- Nima Pourdamghani, Yang Gao, Ulf Hermjakob, and Kevin Knight. 2014. Aligning english strings with abstract meaning representation graphs. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 425–429, Doha, Qatar, October. Association for Computational Linguistics.
- Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James H. Martin, and Daniel Jurafsky. 2004. Shallow semantic parsing using support vector machines. In *Proceedings of the Human Language Technology Conference/North American chapter of the Association of Computational Linguistics (HLT/NAACL)*, Boston, USA.
- Nianwen Xue and Martha Palmer. 2004. Shallow semantic parsing using support vector machines. In *Proceedings of 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain.
- Nianwen Xue and Martha Palmer. 2009. Adding semantic roles to the chinese treebank. *Natural Language Engineering*, 15(1):143–172.
- Nianwen Xue, Ondrej Bojar, Jan Hajic, Martha Palmer, Zdenka Uresova, and Xiuhong Zhang. 2014. Not an interlingua, but close: Comparison of english amrs to chinese and czech. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, pages 1765–1772, Reykjavik, Iceland, May. European Language Resources Association (ELRA).
- Wajdi Zaghouni, Abdelati Hawwari, and Mona Diab. 2012. A pilot propbank annotation for quranic arabic. In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, pages 78–83, Montréal, Canada, June.
- Lijuan Zheng, Yanqiu Shao, and Erhong Yang. 2014. Analysis of the non-projective phenomenon in chinese semantic dependency graph. *Journal of Chinese Information Processing*, 28(6):41–47.
- Yuping Zhou and Nianwen Xue. 2015. The chinese discourse treebank: a chinese corpus annotated with discourse relations. *Language Resources & Evaluation*, 49(2):1–35.