Ellipsis in Chinese AMR Corpus

Yihuan Liu[1] **Bin Li** [1] **Peiyi Yan** [1] **Li Song** [1] **Weiguang Qu** [2]
[1]School of Chinese Language and Literature
Nanjing Normal University
[2]School of Computer Science and Technology
Nanjing Normal University
lyh.njnu@gmail.com

Abstract

Ellipsis is very common in language. It's necessary for natural language processing to restore the elided elements in a sentence. However, there's only a few corpora annotating the ellipsis, which draws back the automatic detection and recovery of the ellipsis. This paper introduces the annotation of ellipsis in Chinese sentences, using a novel graph-based representation Abstract Meaning Representation (AMR), which has a good mechanism to restore the elided elements manually. We annotate 5,000 sentences selected from Chinese TreeBank (CTB). We find that 54.98% of sentences have ellipses. 92% of the ellipses are restored by copying the antecedents' concepts. and 12.9% of them are the new added concepts. In addition, we find that the elided element is a word or phrase in most cases, but sometimes only the head of a phrase or parts of a phrase, which is rather hard for the automatic recovery of ellipsis.

1 Introduction

With the rapid development of artificial intelligence (AI), natural language progressing is one of significant applications of AI, and it has made outstanding progress in several basic techniques, such as syntactic analysis and semantic analysis. The former is relatively mature, while the latter needs more efforts (Sun et al., 2014). For example, in the SRL(Semantic Role Labeling)-only task of the CoNLL 2009, the highest score in English is 86.2% and in Chinese it is 78.6% (Hajič et al., 2009). In addition, a common issue for the current semantic parser is that they ignore the elided element which is not overt in the surface form, but necessary in the understanding of the sentence. That elided element is more often referred as ellipsis in linguistic.

Ellipsis is a common linguistic phenomenon across languages. The traditional linguistic researches pay more attention to the formal construction, and don't regard ellipsis as an important factor. Although some theoretical achievements have been made in the classifications and restrictions of ellipsis (Lobeck, 1995; Merchant, 2004, 2007). There are still debates in the definition of ellipsis, the identity constraint between antecedents and the elided element etc. (Phillips and Parker, 2013) .

Most current corpora don't annotate the elided element. A few corpora view ellipsis as an expediency for some irregular sentences, and annotate the elided element roughly. Such as Penn Treebank (PTB for short) (Marcus et al., 1993, 1994), Chinese Treebank (CTB) (Xue et al., 2005), Prague Dependency TreeBank (PDT) (Böhmová et al., 2000; Hajičová et al., 2001) and Universal Treebank (McDonald et al., 2013; Nivre et al., 2016). It is noticeable that Ren et al. (2018) build a treebank with focusing on ellipsis in context for Chinese. But the corpus only contains 572 sentences from a microblog corpus, and the annotations exclude the elided words which can't be said but play an important role in the understanding of the sentence.

This paper uses a novel framework to restore the elided elements in the sentence, which is named Abstract Meaning Representation (AMR)(Banarescu et al., 2013). AMR represents the whole sentence meaning with concepts, which are mainly abstracted from its corresponding words occurring in the sentence. Based on AMR, Chinese AMR (CAMR) makes some adaptations to accommodate Chinese better. What's more, CAMR develops corresponding restoration methods for different types of ellipses, which makes the restoration more reasonable and complete.

The rest of this paper is organized as follows. In Section 2, we discuss the definition of ellipsis and

Proceedings of the First International Workshop on Designing Meaning Representations, pages 92–99
Florence, Italy, August 1st, 2019

gives a broader definition, which refers to all phenomena wherein the elided elements are necessary for the meaning of the sentence but not overt in the sentence. In addition, we introduce the representation for ellipsis in PTB, PDT. In Section 3, we describe three methods to restore ellipsis in CAMR. And in Section 4 , we introduce the Chinese AMR corpus which includes 5,000 sentences from the newspaper portion of CTB. and we present some statistics and analysis based on this corpus. Then we conclude our paper with a summary of our contribution in Section 5.

2 Related Work

As we mentioned above, the definition of ellipsis is an unsolved issue. Many linguists have been trying to define it from different aspects.

2.1 Definition of Ellipsis

To improve the agreement and the accuracy of annotation, it is necessary for annotators to understand what is ellipsis. Arnauld and Lancelot (1975) first mentioned ellipsis in their work General and Rational Grammar. And they defined it as a pragmatic phenomenon which omits some redundant words for concision. Jespersen (1924) gave a semantic ellipsis, He assumed that grammarians should always be wary in admitting ellipses except where they are absolutely necessary and where there can be no doubt as to what is understood. Carnie (2013) assumed that ellipses are phenomena where a string that has already been uttered is omitted in subsequent structures where it would otherwise have to be repeated word for word. While Lobeck (1995) viewed ellipsis as a mismatch of phonological content and semantic content, He thought ellipsis means deleting some words which can be inferred from context.

There are other definitions of ellipsis. Quirk et al. (1972) assumed that ellipsis is purely a surface phenomenon. In the strict sense of ellipsis, words are elided only if they are uniquely recoverable. There is no doubt as to what words are to be supplied, and it is possible to add the recovered words to the sentence. The definition was referred to the restraint of ellipsis. Ren et al. (2018) gave a definition of ellipsis in the practice of natural language processing. It views ellipsis as textual omission of words or phrases expressing a semantic role in a sentence, which are optional but not obligatory.

Comparing all definitions above, the consensus is that there are elided elements that are helpful for the understanding of the sentence, and can be recovered from context. This paper follows that consensus and gives a more broad definition for ellipsis. It encompasses all phenomena wherein the elided elements which are necessary for the understanding of the sentence don't refer to a token in the surface form. There are mainly two differences between this definition and others, which are:

- The restoration do not have to be unique and unambiguous.
- The restoration do not have to be written in the surface form.

The traditional theory requires the restoration of ellipsis must be unique and ambiguous. But sometimes the elided words can't not be uniquely and unambiguously restored. For example, in the sentence 1 is a headless nominal, and the subject of 跳舞(dance) is omitted. Due to lack of contextual information, we only know that the elided elements refer to a dancer or some dancers, but we don't know exactly who it is. Since the elided elements are important in the meaning of the sentence, we add a new concept person in the ellipsis site and consider this special headless nominal as ellipsis.

(1) 跳舞 的 走 了
dance DE go ASP
"The dancer has gone."

(2) 他 想 吃 苹果
he want eat apple
"He wants to eat an apple."

In most cases, The restoration can be said in the surface form, and it makes the sentence regular. But sometimes, the restoration will make the sentence illegal, which means the restoration is only in semantic level. For example, in the sentence 2, the subject of 想(want) and 吃(eat) is 他(he), but 他(he) occurs once in the sentence. According to the theta criterion, each argument is assigned to one and only one theta role, it needs to add another argument to meet the criterion and present the whole sentence meaning. But the recovered sentence "他想他吃苹果。"("He wants him to eat an apple.") is illegal. Considering the semantic importance of the missing argument, we regard this sharing argument as ellipsis, too.

As the goal of the annotation is to present the complete meaning of the sentence, we focus on the semantic aspect than syntactic aspect. And the scope of ellipsis is obviously more extensive than the traditional one. The typical types like VP ellipsis, NP ellipsis and some special phenomena like headless nominal and sharing argument are covered by ellipsis.

2.2 Ellipsis Representation in PTB and PDT

Most current corpora rarely annotate ellipses, only a few corpora have represented part of ellipses with some particular labels, such as PTB, CTB and PDT. Since CTB follows the annotation principles of PTB on the whole, we only describe the representation strategies for ellipsis of PTB and PDT. By comparing the ellipsis representation in these two corpora, we assume that both of them only handle some typical ellipses, and their tree structures are hard to representation ellipsis.

PTB is a large corpus which mainly contains phrase structure annotation. It incorporates the concept of empty category which is introduced in Generative Grammar. Empty category plays a part in syntactic structure and semantic structure, but it has no corresponding phonological content in the sentence, whose performance is similar with ellipsis. In fact, some types of empty categories are covered by ellipsis. So PTB including empty category representation can provide scant help for ellipsis research.

The specific representation method for ellipsis includes two steps. Firstly, PTB annotates the corresponding empty category label in the ellipsis site. Secondly, PTB attaches the id to the labels to contact the empty category and the related elements in the sentence (Xue et al., 2005).

In Figure 1, 公司(company) is a sharing argument, which is shared by the verb 计划(plan) and 增加(increase). PTB regards the elided argument as PRO, and assigns the label NONE - * PRO * to the ellipsis site. The id -1 behind the empty category label corresponds to the superior node NP-PN-SBJ, which indicates that the elided element is 公司(company).

(3) 公司 计划 增加 产量
Company plan increase output
"The company plans to increase output."

PDT includes three layers which are morphological layer, syntactic layer and semantic layer.

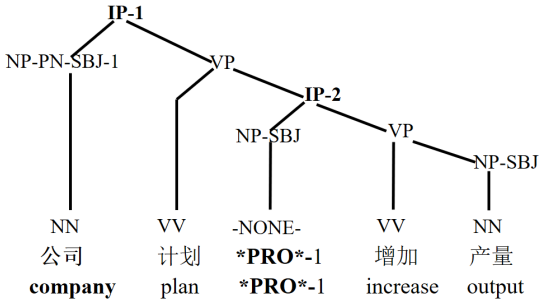

Figure 1: Empty categories in PTB

Each level annotates the morphological, syntactic and semantic information respectively. At the syntactic layer, it annotates the overt words in the sentence, and it restores the elided elements at the semantic layer. The methods of representing ellipsis in PDT are more complex than PTB, which mainly include three steps. Firstly, it adds a new node. Then it judges the category of ellipsis and represent it with corresponding label. At last, if there is an antecedent, it will use the coreference link to associate the new node with its antecedent node (Mikulová, 2014; Hajič et al., 2015).

(4) 公司 计划 增加 产量
Company plan increase output
"The company plans to increase output."

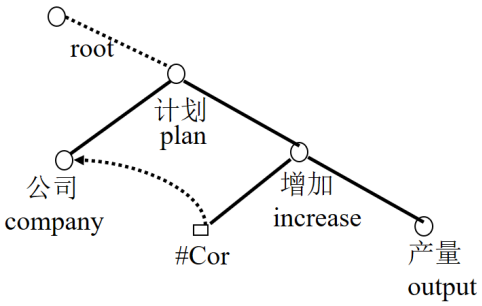

Figure 2: Ellipsis representation in PDT

Figure 2 shows the annotation of Example 3 in PDT. Similar with PTB, PDT also adds a new node for the elided element, and marks it as #Cor, which means the elided element is the subject in the object clause of the control verb 计划(plan). Because of the antecedent 公司(company), coreference link is also added to contact the restored element with its antecedent, as shown by the dotted arrow.

Although PTB and PDT have designed special labels for ellipsis, but they lack complete resolution to deal with some special ellipses. For exam-

ple, the two corpora have no ability to represent the subtle semantic difference between the elided elements and its antecedent. And both of them restore the elided elements by adding a new node, which make the tree structure more complex, especially when the elided elements occur repeatedly in the same sentence. What's more, to represent the identity of the elided element and its antecedent, a coreference link or other similar marks is added to contact them. In that case, the tree structure is changed into a graph structure.

2.3 Concept-to-word Alignment in CAMR

To represent the whole meaning of the sentence in Chinese, CAMR has made some adaptations to accommodate the linguistic facts of Chinese, and one of the special adaptations is alignment. It uses the sequence number of words in the sentence as the concept id of the notional word, which realizes the concept-to-word alignment in the annotation (Li et al., 2017).And this adaptation helps to represent the elided element more intuitional and convenient.

(5) 他[1] 想[2] 吃[3] 苹果[4]
he want eat apple
"He want to eat an apple."

w/want-01
:arg0() **h/he**
:arg1() e/eat-01
:arg0 **h**
:arg1 h2/apple

x2/想-02
:arg0() **x1/他**
:arg1() x3/吃-01
:arg0 **x1/他**
:arg1 x4/苹果

As shown on the textual representation on the left, English AMR does not align the concepts with the words, it assigns the first letter of the word to its concept. When the elided element is restored, its antecedent is not very straightforward, especially when the sentence is complex and there are some other words that have same first letter as the antecedent. Specifically, the elided element 他(he) is represented by the initial letter "h" of its antecedent. To annotate and understand the sentence, we need spend time in finding what the initial letter exactly denotes. It is more likely to cause lower efficiency and higher error rate. While CAMR aligns the concepts to their words, and makes the ellipsis representation more clearly.

3 Ellipsis Presentation in CAMR

As we described above, PTB and PDT mainly restore the elided element by referring to its antecedent. CAMR also represents ellipsis with the help of antecedent, but sometimes the sentence has no antecedent, or the reference of the elided element is not identical but similar with its antecedent. Referring to its antecedent is not reasonable any more. Considering these different linguistic performances of ellipsis, CAMR develops corresponding methods to represent them reasonably, which are:

- Copy the antecedent, if there is an antecedent, and the reference of antecedent and the elided element is identical.
- Add a new concept, if there is no antecedent.
- Add a new concept and copy the antecedent, if there is an antecedent, but the reference of antecedent and the elided element is not identical.

3.1 Copy the Antecedent

When the antecedent can be found in context, CAMR directly copies the antecedent's concept and fills the copied concept in ellipsis site to restore the elided element. It is noticeable that CAMR does not increase new concept like PTB and PDT. The concept of the elided element and antecedent will be merged into one concept. In CAMR graph, the concept of the elided element and antecedent share the same concept node. the elided element and its antecedent are dominated by different elements, thus the semantic structure of the sentence becomes a graph.

(6) 公司[1] 计划[2] 增加[3] 产量[4]
Company plan increase output
"The company plans to increase output."

x2/计划-01
:arg0() **x1/公司**
:arg1() x3/计划-01
:arg0 **x1/公司**
:arg1 x4/产量

Comparing Figure 1, Figure 2 and Figure 3 , CAMR does not add a new concept NONE - * PRO * or #Cor for the elided element like PTB and PDT. It copies the node of antecedent 公司(company) directly, and combines the two arguments into one node. The node 公司(company) represents the elided element and its antecedent at the same time. Since the node 计划(plan) and

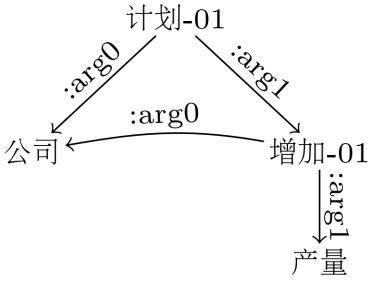

Figure 3: Copy the antecedent in CAMR

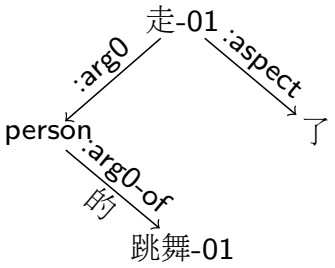

Figure 4: Add a new concept in CAMR

the node 增加(increase) both are fathers of 公司(company), which makes the structure of this sentence a typical graph.

This representation method in CAMR can reduce the total amount of node and make the structure of the whole sentence as clear as possible. The advantage of graph structure benefits when the same elided element occurring repeatedly many times in the sentence. Since no matter how many times the elided element occurs, the number of nodes in the graph will not increase.

3.2 Add New Concepts

When the elided element has no corresponding antecedent in the sentence, the method of copying the concept of antecedent directly is no longer applicable. In this case, CAMR adds a new concept for ellipsis. Specifically, CAMR firstly judges the semantic categories of the elided element and adds an appropriate abstract concepts, such as person and thing. Then it analyses the semantic relationship between the new concept and other concepts. And the whole sentence's meaning is to represent completely.

(7) 跳舞[1] 的[2] 走[3] 了[4]
dance DE go ASP
"The dancer has gone."

x3/走-01
:arg0() **x6/person**
:arg1(x2/的) x1/跳舞-01
:aspect x4/了

Traditionally, it is assumed that the headless relative construction such as 跳舞的(the dancer), is a contextual variant of the formal nominal structure. When the head is the subject or object of the adjunct in this nominal structure, it can be elided (Huang, 1982). In general, there is no antecedent, and the elided elements are abstract. In Example 7, the elided head of 跳舞的(the dancer) is vague. It might be a dancer or some dancers. So CAMR adds an abstract concept person to contact 走(walk) and 跳舞(dance), and completes the whole sentence meaning. In these relations, the semantic relation label arg0-of between person and 走(walk) is an inverse relation of arg0, which is used to maintain a single-rooted structure of CAMR graph.

3.3 Add a New Concept and Copy the Antecedent

There is a special ellipsis where the antecedent can be found in the sentence, but the reference of the elided element and its antecedent is not identity. Previous ellipsis researches tend to neglect that semantic nonidentity. Even though PDT has realized that there are differences between the two items in the comparison structure, the annotation schemes can't represent this semantic difference properly. To represent the whole sentence meaning reasonably, CAMR combines the two method described above. That is adding new concepts and then copying the antecedent. Specifically, according to the semantic category of the elided element, CAMR adds a new concept. Then it analyzes the relation between the elided element and its antecedent, and represents this relationship with special semantic relation labels.

(8) 你[1] 的[2] 收入[3] 比[4] 我[5] 高[6]
you DE income than I high
"Your income is higher than mine."

x6/高-01
:arg0() **x3/收入**
:arg1(x2/的) x1/你
:compare-to(x4/比) **x8/thing**
:poss() x5/我
:dcopy() **x3_s/收入**

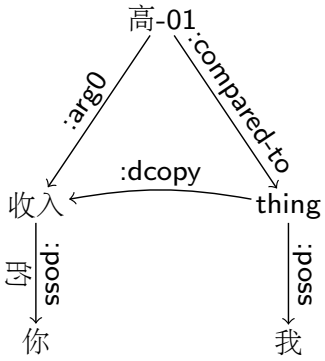

Figure 5: Add a new concept and copy the antecedent

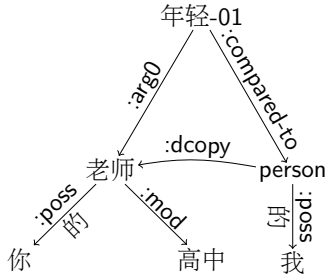

Figure 6: The elided elements are parts of a phrase

The Example 8 is a comparative structure. 你的收入(your income) and 我(I) are asymmetrical in syntactic structure. 我(I) is an incomplete and abbreviated form in semantic expression (Li, 1982). Since the purpose of this sentence is actually to emphasize the difference between the two items 你的收入(Your income) and 我的收入(my income), it is obviously unreasonable to copy the concept directly. So we first add a concept thing and then use a special semantic relation label dcopy, which is added in CAMR to indicate that the elided element and the antecedent belong to the same category, but they refer to different objects in real world.

We further find that there are residual modifiers of the elided elements in Chinese sentence, and these modifiers are the cues which remind us to pay attention to the reference of the elided elements and its antecedent. In Example 6, Example 7, the elided element is a word or a complete phrase exactly. While in Example 8, the elided element is the head of the phrase 我的收入(I income, my income). Sometime it might be more complex. the elided elements are parts of a phrase.

(9) 你[1] 的[2] 高中[3] 老师[4] 比[5] 我[6] 的[7] 年轻[8]
you DE high school teacher than I DE young
"Your high school teacher is younger than mine."

x8/年轻-01:
:arg0() x4/老师
:arg1(x2/的) x1/你
:mod() x3/高中
:compare-to(x5/比) **x10/person**
:poss(x7/的) x4/我
:dcopy() **x3_x4/高中老师**

In Example 9, the elided elements are 高中老师(high school teacher), which are parts of the phrase 我的高中老师(my high school teacher). We are trying to refine the guidelines to represent these different elided elements reasonably, and we will discuss this type of ellipsis in the future.

In conclusion, CAMR can represent the elided element more concisely and show the relationship between the elided element and its antecedent in detail. These three methods can handle most ellipses and represent the semantics of the whole sentence, which determines it is a more reasonable annotation scheme to represent ellipsis.

4 Statistics and Analysis

We annotate 5,000 sentences from Penn Chinese Treebank CTB8.0. Based on this data, we show the proportion of ellipsis and how common it is in Chinese. And we find that the length of the sentence affect the distribution of ellipsis indeed. We also analyze how the added concept work in ellipsis.

4.1 Proportion of Ellipsis in Chinese

As shown in Table 1, the first column **Type** contains three items. Among them, **Overall** means all 5,000 sentences in the corpus. The rest columns represent three statistical indicators,which show the number of tokens, concepts and sentences of ellipsis and overall. In Chinese AMR corpus, we restore 5,787 tokens and 4,178 concepts. And we find that 2,749 sentences are with ellipsis. That is, 54.98% of sentences contain ellipsis, which proves that ellipsis is very common in Chinese.

We further show the proportion of three methods for ellipsis mentioned in Section 3. As shown in Table 2, copying the antecedent is the most popular methods in the corpus, which means that among all elliptical sentences (2,749 sentences), 2,537 sentences appear the identical antecedent. Almost 92% of ellipses can be restored by copy-

Type	Token	Concept	Sentence
Ellipsis	5,787	4,178	2,749
Overall	13,2981	12,0991	5,000
Ratio	4.35%	3.45%	54.98%

Table 1: Proportion of ellipsis in Chinese AMR Corpus

Type	Token	Concept	Sentence
Copy the antecedent	5,143	3,567	2,537
Add a new concept	284	258	230
*Add & Copy	360	353	267

* is the abbreviation of *Add a New Concept and Copy the Antecedent*

Table 2: Frequency of three methods for ellipsis

Type	Token	Concept
Ellipsis	32.58	31.11
Overall	26.6	24.2

Table 3: Average token count and concept count in per sentence

ing its antecedent directly. This high proportion shows that the antecedents are of great importance to restore the elided element, which explains why most current ellipsis models rely on antecedents for ellipsis recognition and restoration.

4.2 The Length of the Elliptical Sentence

The statistics also prove that length of the sentence will affect the distribution of ellipsis. There are two ways to measure the length of a sentence. One is based on words, the length of a sentence refers to the number of words that make up the sentence. The other is based on concepts, the length of a sentence refers to the number of concepts that make up the semantic meaning of a sentence.

The average length of elliptical sentences is about 6 units longer than the regular sentences in the corpus, whether in terms of words or concepts. The reason is that the longer the sentence is, the more complex the semantic structure is and the richer the semantic information is. Therefore, it is more likely to delete some words from the sentence.

4.3 The Added Concept for Ellipsis

CAMR adds new concepts to represent ellipsis when there is no antecedent or the reference of the elided element and its antecedent is different.

Type	Concept	Frequency	Ratio
Add a new concept	thing	110	38.73%
	person	103	36.27%
	country	8	2.82%
Add & Copy	thing	294	81.67%
	person	35	9.72%
	animal	4	1.11%

Table 4: The added concept for ellipsis

CAMR also adds abstract concepts when we annotate proper nouns, special quantity types and special semantic relationships. For example, when annotating quantitative phrases for weight, we first add a concept mass-quantity . These added concepts should be excluded in statistics.

As shown in Table 4, the frequency of thing and person is much higher than other concepts. The reason is mainly that they are more abstract. We usually add *thing* and *person* when the elided element is not clear.

5 Conclusion and Future Work

In this paper, we uses a novel graph-based framework AMR, which mainly represents the elided element by copying its antecedent, adding a new concepts, or we combining the two methods when the reference of the elided elements and its antecedent is not identical. On the basis of Chinese AMR corpus, which contains 5,000 sentences selected from CTB, we show how common ellipsis is in Chinese, and we prove that the length of the sentence affect the distribution of ellipsis indeed. The average length of elliptical sentences is about 6 units longer than the regular. We further show the added concept for ellipsis.

In the future, we will discuss ellipses which are the head of a phrase or just parts of a phrase in detail. And we intend to apply the research result to Chinese AMR parser, to improve its ability to identify and restore ellipsis in Chinese sentences.

Acknowledgments

We are grateful for the comments of the reviewers. This work is the staged achievement of the projects supported by National Social Science Foundation of China (18BYY127) and National Science Foundation of China (61772278).

References

Antoine Arnauld and Claude Lancelot. 1975. *General and Rational Grammar: the Port-royal Grammar*. Mouton Hague, Paris,France.

L Banarescu, C Bonial, S Cai, and et al. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop Interoperability with Discourse*, pages 178–186.

Alena Böhmová, Jan Hajič, Eva Hajičová, and Barbora Hladká. 2000. The prague dependency treebank: a three-level annotation scenario. In *TreeBank: Building and Using Parsed Corpora*.

Andrew Carnie. 2013. *Syntax-a generative introduction*. Wiley Blackwell, London, UK.

Jan Hajič, Massimiliano Ciaramita, Richard Johansson, and et al. 2009. The conll-2019 shared task:syntactic and semantic dependencies in multiple language. In *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL): Shared Task*.

Jan Hajič, Eva Hajičová, and et al. 2015. Deletions and node reconstruction a dependency-based multilevel annotation scheme. In *Proceedings of Computational Linguistics and Intelligent Text Processing*, pages 17–31.

Eva Hajičová, Jan Hajič, Barbora Hladká, Martin Holub, and et al. 2001. The current status of the prague dependency treebank. In *Proceedings of the 5th International Conference on Text, Speech and Dialogue*.

G Huang. 1982. The syntactic function and semantic function of 'de' structure. *Studies in Language and Linguistics*, 1.

Otto Jespersen. 1924. *The Philosophy of Grammar*. Groge Allen & Unwin LTD, London, UK.

Bin Li, Yuan Wen, Lijun Bu, and et al. Annotating the little prince with chinese amrs.

Bin Li, Yuan Wen, Lijun Bu, and et al. 2017. A comparative analysis of the amr graphs from english and chinese corpus of the little prince. *Journal of Chinese Information Processing*, 31.

L Li. 1982. *The Sentence Pattern in Modern Chinese*. The Commercial Press, Beijing, China.

Anne Lobeck. 1995. *Ellipsis Functional Heads, Licensing, and Identification*. Oxford University Press, New York, US.

Mitchell Marcus, Grace Kim, Mary Ann Marcinkiewicz, and et al. 1994. The penn treebank: annotating predicate argument structure. In *Proceedings of the ARPA Human Language Technology Workshop*.

Mitchell Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of english: the penn treebank. *Computational Linguistics*, 19.

Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, and et al. 2013. Universal dependency annotation for multilingual parsing. In *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics*.

Jason Merchant. 2004. Fragments and ellipsis. *Linguistics and Philosophy*, 27(6):661–738.

Jason Merchant. 2007. There kinds of ellipsis: Syntactic, semantic, pragmatic? In *Semantic Workshop*.

Marie Mikulová. 2014. Semantic representation of ellipsis in the prague dependency treebanks. In *Proceedings of the Conference on Computational Linguistics and Speech Processing*, pages 125–138.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, and et al. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*.

Colin Phillips and Dan Parker. 2013. The psycholinguistics of ellipsis. *Lingua*, 151.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1972. *A Grammar of Contemporary English*. Longman Singapore, Singapore.

Xuancheng Ren, Xu Sun, Bingzhen Wei, Weidong Zhan, and et al. 2018. Building an ellipsis-aware chinese dependency treebank for web text. In *Proceedings of the 12th International Conference on Language Resources and Evaluation*.

Maosong Sun, Ting Liu, Donghong Ji, and et al. 2014. Frontiers of language computing. *Journal of Chinese Information Processing*, 28.

N Xue, F Xia, F Chiou, and et al. 2005. The penn chinese treebank: Phrase structure annotation of a large corpus. *Natural Language Engineering*, 11(2):207–238.