International Journal of Knowledge and Language Processing
Volume 11, Number 1, **2020**

www.ijklp.org
KLP International ©2020 ISSN 2191-2734
pp.1--13

Ellipsis in Chinese AMR Corpus

Yihuan Liu[1], Bin Li[1]✉, Peiyi Yan[1], Li Song[2,1] and Weiguang Qu[3]

[1] School of Chinese Language and Literature
Nanjing Normal University
No.122, Ninghai Road, Gulou District, Nanjing, Jiangsu, China
lyh.njnu@gmail.com; libin.njnu@gmail.com

[2] Department of Chinese Language and Literature
Tsinghua University
No. 30 Shuangqing Road, Haidian District, Beijing, China
songli1105@sina.com

[3] School of Computer Science and Technology
Nanjing Normal University
No.1, Wenyuan Road, Qixia District, Nanjing, Jiangsu, China
wgqu_nj@163.com

Selected paper from CLSW2019

ABSTRACT. *Ellipsis is very common in language. It's necessary for natural language processing to restore the elided elements in the sentence. However, there are only a few corpora annotating ellipsis, and the mechanic of how to restore ellipsis is not well studied yet. This paper investigates ellipsis in Chinese sentences, using the graph-based representation Abstract Meaning Representation (AMR), which has a novel mechanism to restore elided elements. We annotate 5,000 sentences selected from Chinese TreeBank (CTB), finding that 54.98% of sentences have ellipses. 92% of the ellipses are restored by copying the antecedents' concepts. And 12.9% of them are newly 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.*

Keywords: Ellipsis, Abstract Meaning Representation, Semantic Representation

1. **Introduction.** With the rapid development of artificial intelligence (AI), natural language progressing as one of significant application areas of AI, is bringing great concern

in recent years. 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 [1]. 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% [2]. In addition, a common issue of current semantic parsers is that they ignore the elided elements which are not overt in surface form, but necessary in the understanding of the sentence. That elided elements are more often referred as *ellipsis* in linguistics.

Ellipsis is a common linguistic phenomenon across languages. Almost all kinds of constituents can be deleted as long as the corresponding syntactic or semantic restrictions are met, such as NP-ellipsis, VP-ellipsis. Traditional linguistic researches pay more attention to 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 [3-5], there are still debates in the definition of ellipsis, the identity constraint between antecedents and elided elements etc. [6].

Most current corpora view ellipsis as an expediency for some irregular sentences, and don't restore elided elements. But there are still some corpora noticing the importance of ellipsis. They annotate elided elements by adding some simple marks roughly, such as Penn Treebank (PTB for short) [7-8], Chinese Treebank (CTB for short) [9], Prague Dependency TreeBank (PDT for short) [10-11] and Universal Treebank (UTB for short) [12-13]. It is noticeable that Ren et al. builds a treebank focusing on ellipsis in context for Chinese [14]. But this treebank only contains 572 sentences from a microblog corpus, and it excludes 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 elided elements in sentences, which is named Abstract Meaning Representation (AMR for short) [15]. AMR represents the whole sentence meaning with concepts which are mainly abstracted from its corresponding words in the sentence. Based on AMR, Chinese AMR (CAMR for short) 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.

The rest of this paper is organized as follows. In Section 2, we are going to give a broader definition of ellipsis. In addition, we will 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 analyses 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. This paper mainly considers the semantic aspect of ellipsis and gives it a broader definition. Although the current corpora like CTB and PDT represent ellipses in the sentence, the annotation is rough and incomplete, which draws back the identification and recovery of ellipsis.

2.1. **The Definition of Ellipsis**. To improve the agreement and the accuracy of annotation, it is necessary for annotators to understand what ellipsis is. Arnald and Lancelot [16] firstly 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 [17] gave us a semantic ellipsis. He assumed that grammarians should always be wary in admitting ellipses except where they are absolutely necessary and where they are clearly understood. Carnie [18] assumed that ellipsis is a phenomenon where a string that has already been uttered is omitted in subsequent structures. Lobeck [3] 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 some other definitions of ellipsis. Quirk et al. [19] assumed that ellipsis is purely a surface phenomenon. In the strict sense of ellipsis, words are deleted only if they are uniquely recoverable. This definition was referred to the restriction of ellipsis. Ren et al. [14] 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 is optional but not obligatory.

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

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

Traditional ellipsis theory requires the restoration is unique and unambiguous. But sometimes the deleted words can't be uniquely and unambiguously restored. For example, the subject of the sentence (1) is a headless nominal. The elided head is the subject of the verb 跳舞(dance). Due to lack of contextual information, we only know that the elided element refers to a dancer or some dancers. Since the elided element is important in the meaning of this sentence, we fill something ambiguous in the corresponding site and consider this special headless nominal as ellipsis.

(1) 跳舞 的 走 了 。
Dance DE go LE
The dancer has gone.

In most cases, the restoration can be restored 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 necessary in semantic level. For example, in the sentence (2), the subject of 想(want) and 吃(eat) is 他(he), but 他(he) occurs only once in the sentence. According to theta criterion [18, 20], each argument is assigned to one and only one theta role. For this sentence, 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.

(2) 他 想 吃 苹果 。

He want eat apple

He wants to eat an apple.

As the goal of the annotation is to present the complete meaning of the sentence, we focus on the semantic aspect rather 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 all 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 guidelines of PTB on the whole, we only describe the representation strategies of ellipsis in PTB and PDT. By comparing the ellipsis representation in these two corpora, we find that both of them only handle partial typical ellipses.

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, but it has no corresponding phonological content in the sentence, whose performance is similar to 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 of ellipsis includes two steps. Firstly, PTB annotates the corresponding empty category label in the ellipsis site. Secondly, PTB contacts the empty category and the related elements in sentence by the same id number.

(3) 公司 计划 增加 产量。

Company plan increase output

The company plans to increase output.

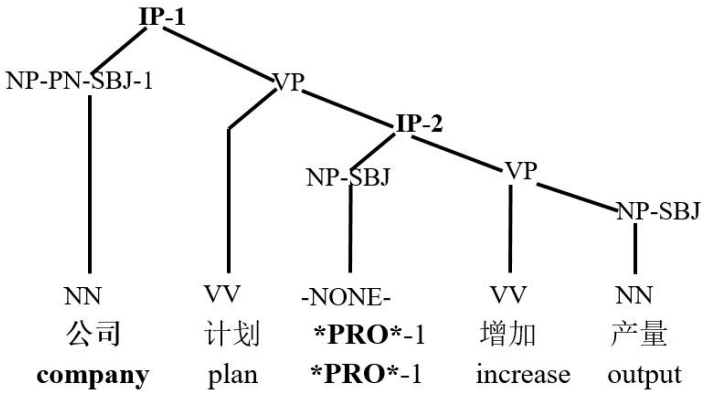

FIGURE 1. EMPTY CATEGORY IN PTB

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 number *-1* behind the empty category label corresponds to the superior node NP-PN-SBJ, which indicates that the elided element is 公司(company).

PDT includes three layers which are morphological layer, syntactic layer and semantic layer, annotating morphological, syntactic and semantic information respectively. At the syntactic layer, it annotates overt words in sentence, and it restores elided elements at the semantic layer. The method of representing ellipsis in PDT is more complex than PTB, which mainly includes three steps. Firstly, it adds a new node. Then it judges the category of ellipsis and represents it with corresponding label. Finally, if there is an antecedent, it will use coreference link to contact the new node with its antecedent node [20-21].

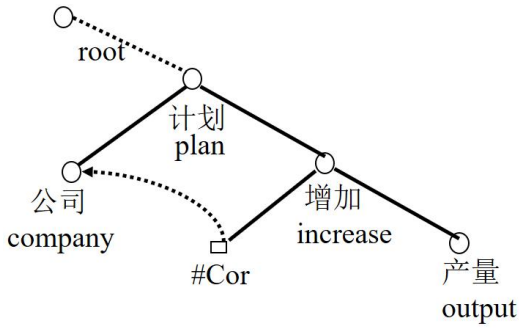

FIGURE 2. ELLIPSIS REPRESENTATION IN PDT

Figure 2 shows the annotation of sentence (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, they lack complete resolution to deal with some special ellipses. For example, the two corpora have no ability to represent the subtle semantic difference between the elided element and its antecedent. And both of them restore the elided elements by adding a new node, which makes the tree structure more complex, especially when the elided elements occur many times 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 are added to contact them, which makes the tree structure changed into a graph structure.

2.3. **Concept-to-word Alignment in CAMR**. To represent the whole meaning of the sentence more intuitively, CAMR has made some adaptations to accommodate the linguistic facts of Chinese, and one of the outstanding adaptations is to align word and concept in the AMR graph by assigning the sequence number of words to the concept [22-23]. And this adaptation helps to represent elided elements more intuitional and convenient.

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

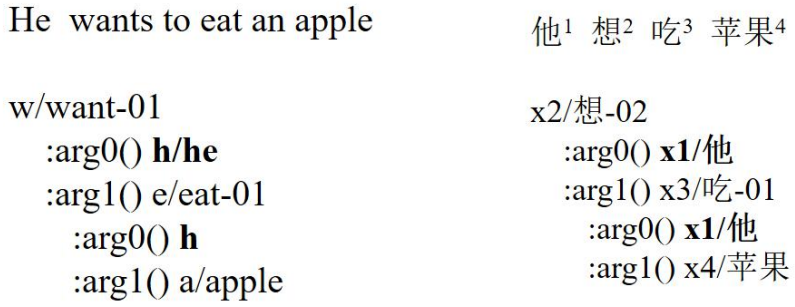

FIGURE 3. TEXTUAL REPRESENTATION OF ENGLISH AND CHINESE AMR

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

3. **Ellipsis Representation in CAMR.** As we mentioned 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 a sentence has no antecedent, or the reference of elided element is not identical but similar to 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 not an 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 a new concept like PTB and PDT. The concept of the elided element and its antecedent will be merged into one concept. That is to say in CAMR graph, the concept of elided element and its antecedent share the same node. The elided element and its antecedent are dominated by different elements. Thus the semantic structure of the sentence becomes a graph.

(5) 公司[1] 计划[2] 增加[3] 产量[4] 。[5]
Company plan increase output
The company plans to increase output.

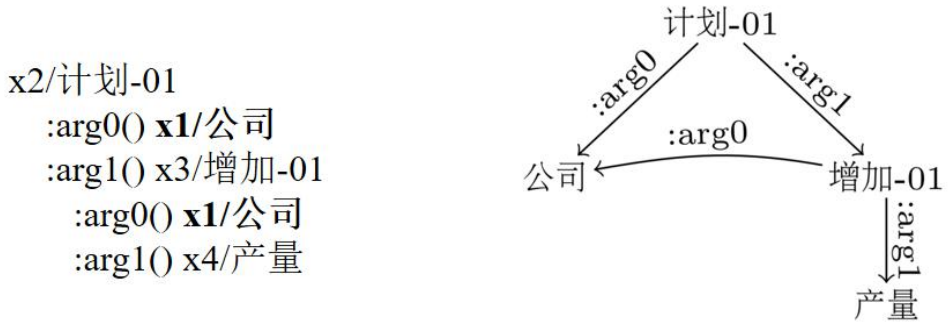

FIGURE 4. COPY THE ANTECEDENT IN CAMR

Comparing Figure 1, Figure 2 and Figure 4, CAMR does not add a new concept NONE -* PRO * or #Cor for 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 the node 增加(increase) 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 nodes and make the structure of whole sentence as clearer as possible. The advantage of graph structure benefits when the same elided element occurs repeatedly 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 a New Concept**. When the elided element has no corresponding antecedent in sentence, the method of copying the antecedent directly is no longer applicable. In this case, CAMR adds a new concept for ellipsis. Specifically, CAMR firstly judges the semantic category of elided element and adds an appropriate abstract (ambiguous) concept, such as *person* and *thing*. Then it analyzes the semantic relationship between the new concept and other concepts. And the whole sentence's meaning is represented completely.

(6) 跳舞[1] 的[2] 走[3] 了[4] 。[5]
Dance DE go LE
The dancer has gone.

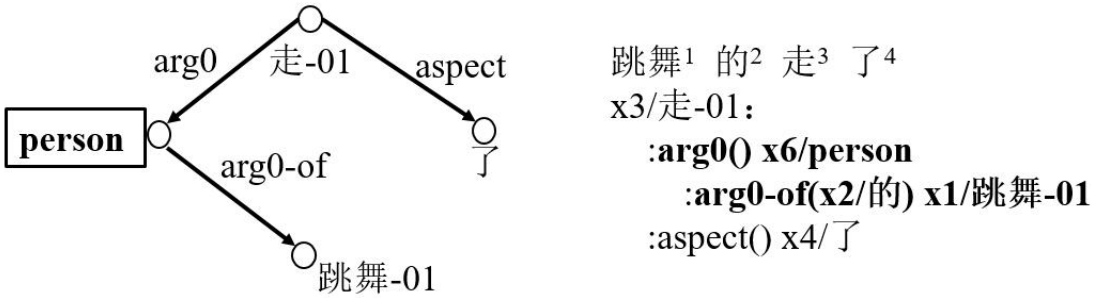

FIGURE 5. ADD A NEW CONCEPT IN CAMR

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 [24-25]. In general, there is no antecedent, and the elided elements are abstract. In sentence (6), 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. 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 sentence, but the reference of the elided element and its antecedent is not identical. 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 methods 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.

(7) 你[1] 的[2] 收入[3] 比[4] 我[5] 高[6] 。[7]
You DE income than I high
Your income is higher than mine.

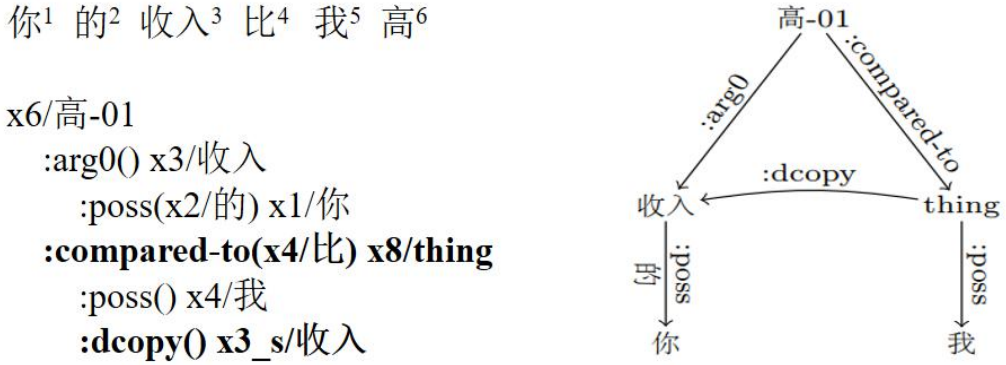

FIGURE 6. ADD A NEW CONCEPT AND COPY CONCEPT IN CAMR

The sentence (7) 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. Since the function 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 elided elements in Chinese sentences, and these modifiers are the cues which remind us to pay attention to the reference of the elided element and its antecedent. In sentence (5) (6), the elided element is a word or a complete phrase exactly. While in sentence (7), the elided element is the head of the phrase 我的收入(my income). Sometime it might be more complex. The elided elements are parts of a phrase.

(8) 你[1] 的[2] 高中[3] 老师[4] 比[5] 我[6] 的[7] 年轻[8] 。[9]

You DE high school teacher than I DE young

Your income is higher than mine.

你[1] 的[2] 高中[3] 老师[4] 比[5] 我[6] 的[7] 年轻[8]

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

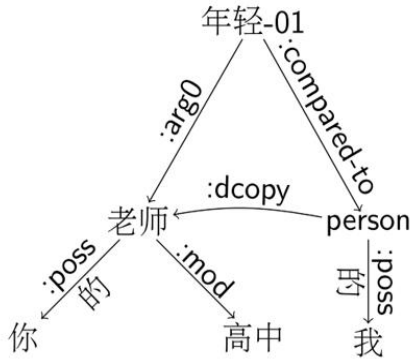

FIGURE 7. THE ELIDED ELEMENTS ARE PARTS OF A PHRASE

In sentence 8, 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 elided elements 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 meaning 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 CTB8.0. Based on this corpus, we show the proportion of ellipsis and how common it is in Chinese. And we find that the length of the sentence affects the distribution of ellipsis. We also analyze how added concepts work in ellipsis.

4.1. **The Proportion of Ellipsis in Chinese**. Ellipsis is a common phenomenon in natural languages. That statement can be seen frequently in traditional ellipsis research. But there are rare data to support it, and we don't know exactly how common it is. Depending on these copied and added concepts for elided elements in the corpus, we can obtain the exact proportion of ellipsis in Chinese sentence.

TABLE 1. STATISTICS OF ELLIPSIS IN CHINESE AMR CORPUS.

	#sentence	#Token	#Concept
Ellipsis	2,749	5,787	4,178
Overall	5,000	132,981	120,991
Percentage (%)	54.98	4.35	3.45

As shown in table 1, we restore 5,787 tokens and 4,178 concepts in Chinese AMR corpus. And we find that 2,749 sentences are with ellipsis. That is, 54.98% of sentences contain ellipsis, which proves that ellipsis is really common in Chinese. Besides, we further show the proportion of three types of ellipsis mentioned in Section 3.

TABLE 2. STATISTICS OF THREE METHODS FOR ELLIPSIS IN CHINESE AMR CORPUS.

	#Sentence	#Token	#Concept
There is an identical antecedent	2537	5143	3567
There is no antecedent	230	284	258
There is an nonidentical antecedent	267	360	353

Among all elliptical sentences (2,749 sentences), 2,537 sentences appear identical antecedent, which means that almost 92% of ellipses can be restored by copying its antecedent directly. This high proportion shows that antecedents are of great importance to restore elided elements, which could explain why most current models rely on antecedents for ellipsis recognition and restoration.

4.2. **The Length of the Elliptical Sentence.** The statistics also prove that the 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 word, and the length of a sentence refers to the number of words that make up the sentence. The other is based on concept, and the length of a sentence refers to the number of concepts that make up the semantic meaning of a sentence.

TABLE 3. Statistics of the length of elliptical sentence.

	#Token	#Concept
Overall	26.6	24.2
Ellipsis	32.58	31.11

The average length of elliptical sentences is about 6 units longer than that of 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 sentences.

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. 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 from statistics.

TABLE 4. STATISTICS OF NEW CONCEPTS FOR ELLIPSIS.

	#concept	#frequency	#Percentage (%)
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

As shown in Table 4, the frequency of *thing* and *person* is much higher than that of other concepts. The reason is mainly that they are more abstract. We usually add *thing* and *person* when the elided element is vague.

5. **Conclusion and Future Work**. In this paper, we use a novel graph-based framework AMR, which mainly represents the elided element by coping its antecedent, adding a new concept, or combining these two methods when the reference of the elided element and its antecedent are 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 affects the distribution of ellipsis indeed. The average length of elliptical sentences is about 6 units longer than that of the regular. We further show the added concepts 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 research results to Chinese AMR parser, to improve its ability to identify and restore ellipsis in sentences.

Acknowledgment. We are grateful for the comments of all reviewers. This work is the staged achievement of the projects supported by National Social Science Foundation of China (18BYY127), National Science Foundation of China (61772278) and the project for Jiangsu Higher Institutions' Excellent Innovative Team for Philosophy and Social Sciences..

REFERENCES

[1] M.S.Sun, T.Liu, et al., Frontiers of language computing. *Journal of Chinese Information Processing*, vol.28, no.1, pp: 1-8, 2014.

[2] J.Hajič, M.Ciaramita, et al., The CoNLL-2009 shared task: syntactic and semantic dependencies in multiple language. *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL): Shared Task*, 2009.

[3] A.Lobeck, *Ellipsis Functional Heads, Licensing, and Identification*. Oxford University Press, New York, 1995.

[4] M.Jason, Fragments and ellipsis. *Linguistics and Philosophy*, no.27, pp: 661–738, 2004.

[5] M.Jason, There kinds of ellipsis: syntactic, semantic, pragmatic?. *Semantic Workshop*. Rutgers University, 2007.

[6] C.Phillips and D.Parker, The psycholinguistics of ellipsis. *Lingua*, vol.151, no.1, pp: 78-95, 2013.

[7] M.P.Marcus, B. Santorini, et al., Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*. Vol.19, no.2, pp: 313-330, 1993.

[8] M.P.Marcus, M.A.Marcinkiewicz, et al., The Penn Treebank: annotating predicate argument structure. *Proceedings of the ARPA Human Language Technology Workshop*, 1994.

[9] N.W.Xue, E.Xia, F.D.Chiou, et al., The Penn Chinese TreeBank: phrase structure annotation of a large corpus. *Natural Language Engineering*, vol.11, no.2, pp: 207–238, 2005.

[10] A.Böhmová, J.Hajič, E.Hajičová, et al., The Prague Dependency TreeBank: a three-level annotation scenario. *TreeBank: Building and Using Parsed Corpora*, 2000.

[11] E.Hajičová, J. Hajic, et al., The current status of the Prague Dependency TreeBank. *Proceedings of the 5th International Conference on Text, Speech and Dialogue*.2001.

[12] R.McDonald, J. Nivre, et al., Universal dependency annotation for multilingual parsing. *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics*. 2013.

[13] J.Nivre, M.Marneffe, et al., Universal Dependencies v1: a multilingual treebank collection. *Proceedings of the Tenth International Conference on Language Resources and Evaluation*. 2016.

[14] X.C.Ren, X.Sun, et al., Building an ellipsis-aware Chinese Dependency Treebank for web text. *Proceedings of the 12th International Conference on Language Resources and Evaluation*. 2018.

[15] L.Banarescu, C.Bonial, S.Cai, et al., Abstract meaning representation for sembanking. *Linguistic Annotation Workshop and Interoperability with Discourse*, Sofia, Bulgaria, pp.178–186, 2013.

[16] A.Arnauld and C.Lancelot, *General and R.Grammar: the Port-royal Grammar.* Hague: Mouton, 1975.

[17] O.Jespersen, *The Philosophy of Grammar*. Groge Allen & Unwin LTD, London, 1924.

[18] A.Carnie, *Syntax-a Generative Introduction*. Wiley-Blackwell, London, 2013.

[19] R.Quirk, S.Greenbaum, G.Leech and J.Svartvik. *A Grammar of Contemporary English*. Longman Singapore, Singapore, 1972.

[20] N.Chomsky, *Some Concepts and Consequences of the Theory of Government and Binding*. MIT Press, Cambridge, 1982.

[21] M.Mikulová, Semantic representation of ellipsis in the Prague Dependency Treebanks. *Proceedings of the Conference on Computational Linguistics and Speech Processing*. pp: 125-138, 2014.

[22] J.Hajič, E.Hajičová, et al., Deletions and node reconstruction a dependency-based multilevel annotation scheme. *Proceedings of Computational Linguistics and Intelligent Text Processing*, Cairo, Egypt, pp: 17-31, 2015.

[23] B.Li, Y.Wen, L.J.Bu, et al., Annotating the Little Prince with Chinese AMRs. *Linguistic Annotation Workshop*, Berlin, Germany, 2016.

[24] B.Li, Y.Wen, L.J.Bu, et al., A comparative analysis of the AMR graphs from English and Chinese

corpus of the Little Prince. *Journal of Chinese Information Processing*, Nanjing, China, vol.31, no.1, pp: 50-57, 2017.

[25] G.Huang, The syntactic function and semantic function of 'DE' structure. *Studies in Language and Linguistics*, no.1, pp: 101-129, 1982.

[26] B.Q.Shi, R.B.Dai., W.G.Qu, et al., Hybrid neural network for recognition of the 'DE' structure. *Acta Scientiarum Naturalium Universitatis Pekinesis*. vol.55, no.1, pp: 75-83, 2018.