
Faculty of Mathematical Sciences



University of Twente
The Netherlands

P.O. Box 217
7500 AE Enschede
The Netherlands

Phone: +31-53-4893400

Fax: +31-53-4893114

Email: memo@math.utwente.nl

www.math.utwente.nl/publications

MEMORANDUM No. 1660

Translation: an example from ancient
Chinese to modern Chinese

X. LIU¹ AND C. HOEDE

NOVEMBER, 2002

ISSN 0169-2690

¹Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, P.R. China

Translation: an example from ancient Chinese to modern Chinese*

X. Liu

**Department of Applied Mathematics
Northwestern Polytechnical University
Xi'an, Shaanxi 710072, P.R.China**

C. Hoede

**Department of Applied Mathematics
University of Twente
P.O.Box 217
7500AE Enschede, The Netherlands**

Abstract

In this paper, we gave an idea of translation by means of knowledge graph theory from ancient Chinese to modern Chinese, by using an example story. Actually, we give the details of the method of translation from ancient Chinese to modern Chinese step by step as carried out by hand. From the example, we found that knowledge graphs have a strong ability to represent sentences. And we also found that there are things that should be discussed if we want to translate automatically from one kind of language to another kind of language by using knowledge graph theory.

Key Words: ancient Chinese, modern Chinese, knowledge graph theory, translation

AMS Subject Classifications: 05C99, 68F99

1 Introduction

In ancient times, Chinese was a language with monosyllabic words. Actually, most modern Chinese words have two syllabic words and polysyllabic words, while Chinese morphemes remain monosyllabic.

Each Chinese character is one syllabic word with one or more independent pronunciation(s).

* Research supported in part by SFAC through grant 01J53079.

Historically, there were about fifty thousand Chinese characters altogether. At present, according to some statistics, there are 44908 Chinese characters. Most characters have homonyms. Chinese characters belong to ideographs.

Chinese is an analytic language without inflection, i.e., without form change. The logical relationship in Chinese between character and character, or word and word, or sentence and sentence depends on their order, their meaning and some function words (function words are an important feature of Chinese). Words have no inflection, no matter what position they have in the sentence. A written Chinese sentence is a line of evenly arranged Chinese characters. Chinese has no marks helping to segment words and phrases except punctuation marks. The main way to segment words or phrases is parsing them using a grammar, take into account semantic aspects and using a dictionary, traditionally. All of these peculiarities make Chinese very difficult for computer processing.

Chinese characters came into being at least three thousand and five hundred years ago during the 商 “Shang” and 殷 “Yin” Dynasties. The earliest Chinese characters are 甲骨文 “Jiaguwen” --- inscription on bones or tortoise shells in the Shang Dynasty (16th—11th century B.C.). There are about 4500 characters in Jiaguwen that have been found till now.

Jiaguwen are already a kind of rather mature character. They contain all the basic Chinese characters building methods:

1. 象形 “Xiang4 xing2”: Pictographic
2. 指事 “Zhi3 shi4”: Self-explanatory
3. 会意 “Hui4 yi4”: Associative Compounds
4. 形声 “Xing2 sheng1”: Pictophonetic.

In modern Chinese 90% of characters are built up using the pictophonetic method (in Jiaguwen only 20%). Pictophonetic characters have two parts: form part and sound part, the former decides on the general meaning and the latter part gives the sound and the specific meaning. We can take some examples in the following table.

Chinese character	杆	杠	杖	板	柜	架
spelling	gan3	gang4	zhang4	ban3	gui4	jia4
meaning	Pole	thick stick	weaponry	board	cupboard	rack, shelf

All these words have a common form part: 木 (mu4, WOOD); this means that all are things made of wood (in ancient times). Yet all these words have their own sound part, so they have their own particular meaning and pronunciation.

Not only the pictophonetic method, but also the other methods are still interesting. For example, 旧 “jiu4”, OLD, is formed by two parts: 一 “yi1”, ONE, and 日 “ri4”, DAY. Its meaning is OLD. That is to say “one day” = “old”. 闻 “wen2”, HEAR, is still formed by two parts: 门 “men2”, DOOR, and 耳 “er3”, EAR. Its meaning is HEAR, i.e. “the ear in the door” = “hear”.

Generally speaking, the radicals in Chinese characters are used as form part of characters (words).

There are 540 radicals in ancient times and 188 radicals in modern Chinese. The sound part of Chinese characters is important as well. There are about one hundred different sound parts to correspond to every radical; many radicals do correspond to the same sound part.

Chinese characters are different from other ideograph characters in history, such as ancient Egyptian characters and cuneiforms (a kind of syllabic language). All these forms died out long ago. Chinese characters still keep their basic feature of ideographs, and make Chinese a root-isolated language.

Most commonly used Chinese characters have powerful word-building abilities, and every Chinese character keeps its form invariable. Mostly, a Chinese character is therefore a basic meaning unit. By combining two or more of these meaning units, hundreds of thousands of Chinese words can be built (some characters are words themselves). Chinese characters are powerful in intelligent information transmission and expression as well. Children, who have learned about 4000 characters in common use, would be able to understand Chinese texts without any difficulty even though they have no further knowledge, while an American student should know about 50000 English words to understand the New York Times [Tang, 1982]. Chinese characters make Chinese a powerful language, since they ensure a high efficiency in communication.

To make learning and processing Chinese characters convenient, many schemas about Chinese characters coding have been proposed and used in computers.

From the above discussion, we see that the Chinese characters and the English letters are not to be considered to be on the same level. The Chinese characters are more basic for semantics than the English letters are, because English letters do not have any meaning. The analysis of Chinese characters can start from the parts of the characters.

2 Ancient Chinese and modern Chinese

Ancient Chinese is the writing language in ancient times in China. The articles in ancient Chinese we saw mostly are written in standard language used during the pre-Qin dynasty.

Ancient Chinese and modern Chinese are just two writing language systems in China. They were used in different ages. The relations between these two languages are inheritance and development. We can study the relations from vocabulary and grammar, the two main elements in writing languages.

In vocabulary, the inheritance is shown by the identity in word-buildings, the identity of basic vocabulary and some ancient Chinese letters and words that have remained in modern Chinese or as parts of modern Chinese words. The development is shown by the replacement of words, the shift of word meanings and that of monosyllabic words to disyllabic words or polysyllabic words from ancient Chinese to modern Chinese. The outstanding evolution from ancient Chinese to modern Chinese is that from monosyllabic words in ancient Chinese to disyllabic or polysyllabic words in modern Chinese [Chen, 1979].

In grammar, both ancient Chinese and modern Chinese are undeveloped in the change of the form ---- declension. The main methods in Chinese grammar are word order and the use of function

words. So, Chinese grammar is very flexible. The rules in Chinese grammar are simple. The parts of speech and the structures of grammar rules are almost the same in ancient Chinese as in modern Chinese. The main differences include the following three things. (1) It is a general phenomenon that in ancient Chinese a word in one part of speech can be used in another part of speech. (2) There are some differences in word order and sentence pattern between ancient Chinese and modern Chinese. (3) The replacement of function words. The function words used in ancient Chinese are nearly not used in modern Chinese.

The articles in ancient Chinese are very short. It is an outstanding impression when one wants to compare ancient Chinese and modern Chinese. If one wants to translate one article in ancient Chinese to modern Chinese, the length will increase at least one time. The main reasons are: (1) Words in ancient Chinese are almost all monosyllabic, but in modern Chinese, words are almost all disyllabic or polysyllabic; (2) Authors in ancient times stress omission in writing, because the writing instruments in ancient times were very burdensome, expensive and hard to use.

Because of this, it is difficult to read articles in ancient Chinese. So, it is very meaningful and useful to study the automatically translation from ancient Chinese to modern Chinese.

In this paper, we will make use of an article in ancient Chinese named “Fox Has the Aid of Tiger’s Stateliness” to discuss the translation from ancient Chinese to modern Chinese by means of knowledge graphs.

The main steps in translation from ancient Chinese to modern Chinese by means of knowledge graphs are the following six. (1) Input the sentence in ancient Chinese. (2) Make the word graphs in ancient Chinese. (3) Construct the sentence graph by gluing the word graphs of the words in the sentence in ancient Chinese. (4) Translate the word graphs in ancient Chinese to word graphs in modern Chinese. (5) Translate the sentence graph in ancient Chinese to a sentence graph in modern Chinese. (6) The final step is expressing the sentence graph in modern Chinese in a syntactically correct way.

Note that step (2) and (3) form the structural parsing of the sentence. Step (4) and (5) form essential steps of the translation. Step (6) finally determines the uttering of the sentence graph in modern Chinese.

3 An example to illustrate the method by an article

In this paper, we will give in detail the method to translate an article from ancient Chinese to modern Chinese.

Sentence 1

The first sentence is “狐假虎威” (hu2 jia2 hu3 wei1).

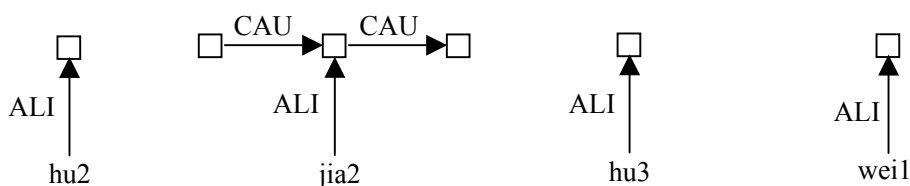
Step 1. The sentence in AC (Ancient Chinese) is the following.

狐 假 虎 威。

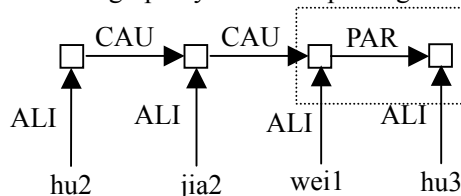
hu2 jia2 hu3 wei1.

fox have the aid of tiger stateliness.

Step 2. We have the following word graphs.



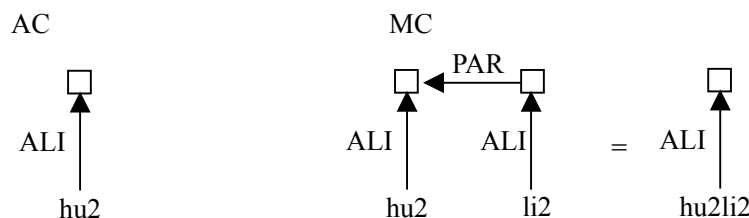
Step 3. We construct the sentence graph by structural parsing. The sentence graph is



That is structural parsing and we will not discuss that for the moment. Note that the PAR-link that is indicated by the broken block shown is not mentioned in the sentence. It takes considerable expansion of the word graphs and their syntactical aspects to find the structure given.

Actually, according to Chinese grammar, two Chinese words can be combined into one Chinese phrase. The meaning of the phrase is from the meanings of its components. We have discussed this in [Hoede *et al.*, 2000]. From that paper we know that the order of the components is very important. So, here, “虎 hu3, TIGER” is the attribute, but “威 wei1, STATELINESS” is the center word. Therefore we get the sentence graph.

Step 4. We have to express this sentence graph in AC to MC (Modern Chinese). That means that we have to express AC words in terms of MC words. (We need a vocabulary in graph form.)



Here, the added word “狸 li2” is another animal named “raccoon dog”. It looks like the animal “狐 hu2” in form, as we mentioned in [Hoede *et al.*, 2000]. The meaning of the word “狐 hu2” is FOX. The meaning of the word “狐狸 hu2li2” is still FOX. Why did we have to add the word “狸 li2” after the word “狐 hu2”, whereas they have the same meaning? We cannot give the exact answer. We think the following two aspects may be part of the reason. One is that “狸 li2” looks like the animal “狐 hu2” in form, we do not need to separate these two kinds of animals. So, we use “狐狸 hu2li2” to indicate these animals. Another one is the need of the disyllabic in MC.

The next word is “假 jie2”. Here we use “借助 jie4zhu4, HAVE THE AID OF” in MC to express the word “假 jie2” (see Figure 1). “Jie4” has the basic meaning BORROW, “zhu4” means HELP. So, “jie4zhu4” has its meaning HAVE THE AID OF. The expression of the word graph of “借助 jie4zhu4” is very profound. It indeed expresses the meaning “to borrow something to help oneself”.

The next one is “虎 hu3”. Here, “lao3” is added in modern Chinese. “Lao3” has basic meaning OLD. Although “lao3” has basic meaning OLD, the “lao3” in this sentence does not have the meaning OLD. The word “lao3” is only the prefix of the word “hu3” according to the “Advanced

Chinese Dictionary”. “Hu3” in AC and “lao3hu3” in MC have the same meaning TIGER (see Figure 2). Another example is the word rat that is “鼠, shu3” in AC and “老鼠, lao2shu3” in MC. There “lao3” is the prefix of the word “shu3”. The situation of replacing the thing expressed by one word (or monosyllabic) in AC by two words (disyllabic) or more word (polysyllabic) in MC is an example of the development of Chinese as we mentioned before.

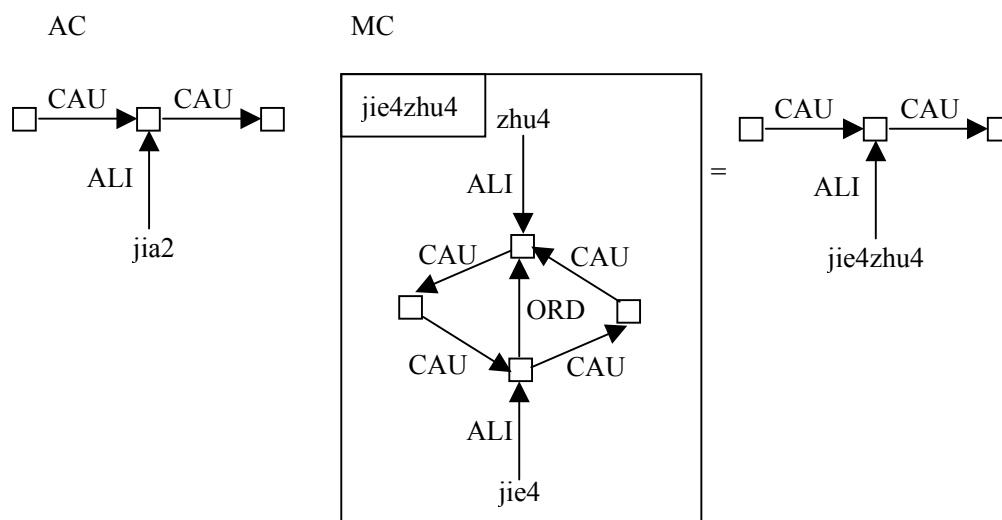


Figure 1

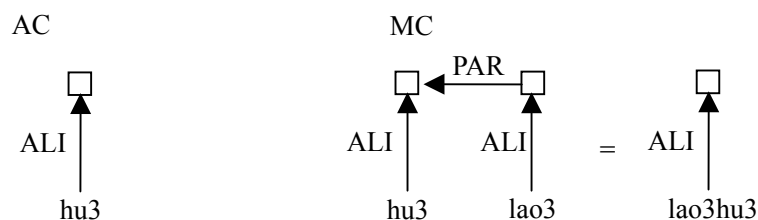
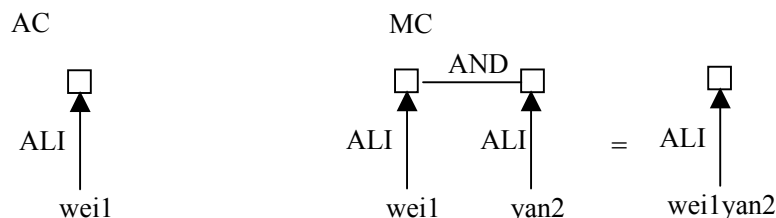
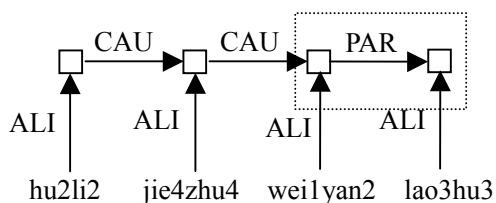


Figure 2

The last word in the sentence is “威 wei1”. Here “威 wei1” means POWER, and “^{JE} yan2” means MAJESTIC. (The basic meaning is URGENT, majestic is its extended meaning). “威^{JE} wei1yan2” means STATELINESS.



Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

狐狸 借助 老虎 的 威严。
 hu2li2 jie4zhu4 lao3hu3 de wei1yan2.
 fox have the aid of tiger of stateliness.

Note that uttering is a problem once the graph is constructed. Also note that the PAR-link is expressed by “的 de, OF”.

From this analysis, we see that there are two important problems in translation. The first problem is to glue word graphs into a sentence graph. That as we said is called *structural parsing*. The second one is the problem of *uttering* the transformed sentence graph.

The following sentences will be discussed less extensively.

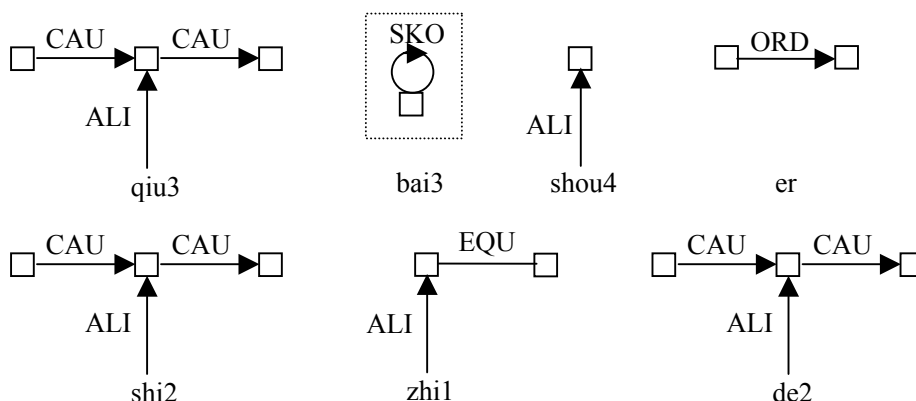
Sentence 2

The second sentence is “虎求百兽而食之，得狐” (hu3 qiu2 bai3 shou4 er3 shi2 zhi1, de2 hu2).

Step 1. The sentence in AC (Ancient Chinese) is the following.

虎 求 百 兽 而 食 之， 得 狐。
 hu3 qiu2 bai3 shou4 er3 shi2 zhi1, de2 hu2.
 tiger seek many animal to eat them, get fox.

Step 2. We have the following word graphs (The words “虎 hu3” and “狐 hu2” have been discussed, so we consider the remaining words).



Step 3. We construct the sentence graph by structural parsing. The sentence graph is shown in Figure 3.

Note that there are three relations not represented by words in the graph but indicated by three broken blocks. The first one is the SKO relation that represents the word “百, bai3, MANY”; the second one is the relation ORD that represent the conjunction of verbs “而, er3, TO AND”; the third one is another ORD relation that represent the comma. How to glue the word graphs obtained in the step 1 into the sentence graph? We only should glue the word graphs according to their orders except for the pronoun “之, zhi1, THEM”. The pronoun “之, zhi1, THEM” indicates

the word mentioned before. From the meaning of the sentence, we can obtain that the pronoun “之, zhi1, THEM” indicates the word “百兽, bai3shou4, MANY ANIMAL”.

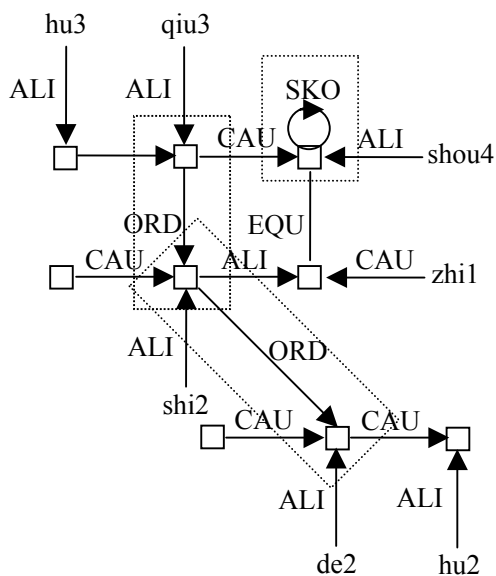
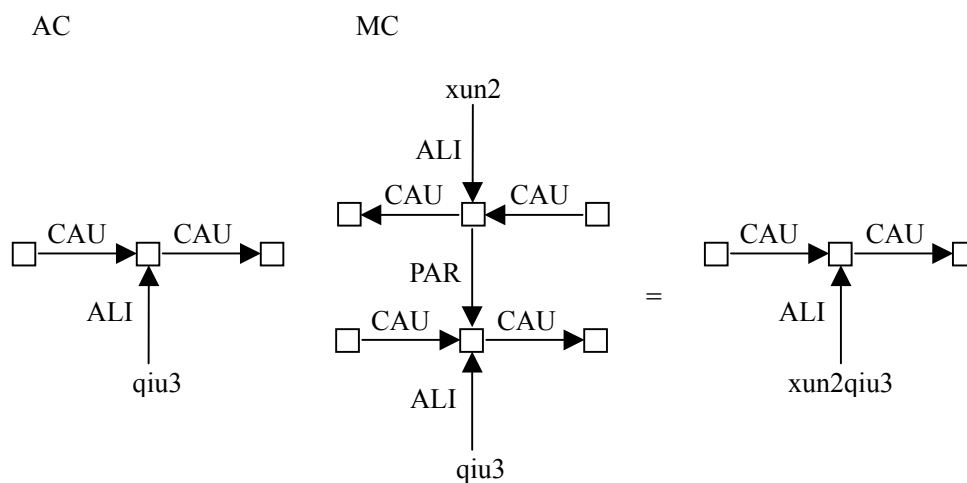


Figure 3

Step 4. We have to express this sentence graph in AC to MC (Modern Chinese). That means that we have to express AC words in terms of MC words.



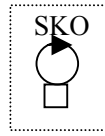
Here, the basic meaning of the word “寻 xun2” is equal to the basic meaning of the word “求 qiu2” in modern Chinese, and the meaning is SEEK. Why did one need to add the word “寻 xun2” before the word “求 qiu2” in modern Chinese? The reason is, as we mentioned above, that almost all words are disyllabic or polysyllabic in modern Chinese.

The next word is “百 bai3”. Its basic meaning is HUNDRED. Its extended meaning is “多 duo1, MANY” even in AC. Another example of its use is given in the sentence

百	战	不	殆。
bai3	zhan4	bu2	dai4.
hundred	war	not	danger.

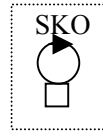
There, the meaning of the word “百 bai3” is still its extended meaning MANY. So, we have

AC



bai3

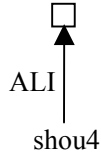
MC



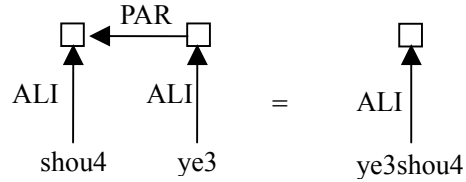
duo1zhong3

and we picked out the word “多种 duo1zhong3, MANY” in MC to express it.

AC

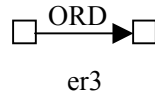


MC

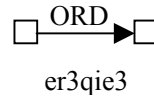


There the word “野 ye3” has the basic meaning WILD. The reason adding the word “野 ye3, WILD” before the word “兽 shou4, ANIMAL” is the same as mentioned above, i.e., the development to using disyllabic words.

AC



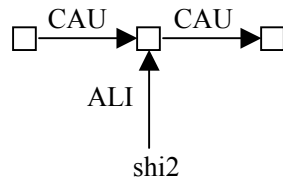
MC



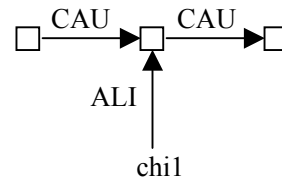
The word “而 er3” is a conjunction. The function of the word is to add two sentences. Moreover, it gives the “order” of two actions expressed by the two sentences. The “purpose” of the first sentence is the second sentence. Because the “order” or “purpose” is represented by TO in English, we use the ORD-relation to express “而 er3” here. Here, the meaning of the word “且, qie3” is EVEN.

The next word “食 shi2” is a very simple one. Its meaning is “EAT”. At present, we still use “食 shi2” in MC to express EAT, especially in certain phrases. For instance, “美食家 mei3 shi2 jia1” is a phrase in Chinese. Its meaning is “belly-god”. Here, we use “食 shi2” to express EAT. But, in most cases, we use “吃 chi1” instead of “食 shi2” in MC to express EAT. Therefore we have

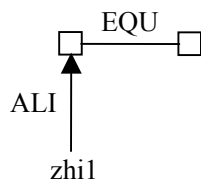
AC



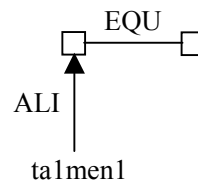
MC



AC

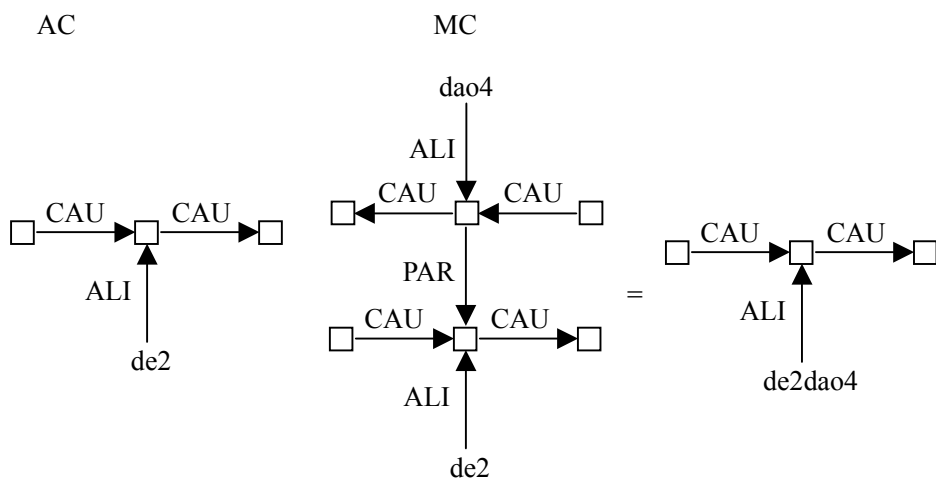


MC



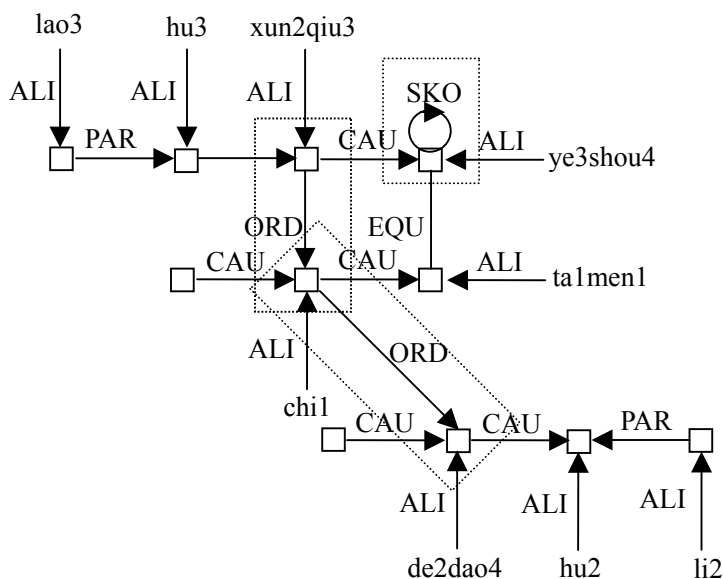
The word “之 zhi1” is a pronoun in AC to indicate a person or thing. In this sentence, it indicates animals. So we picked out the word “它们 ta1men1, them” to express it.

The next word “得 de2” is very simple. We have



Here, the word “到 dao4” has the basic meaning TO.

Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

老虎 寻求 多种 野兽 而且 吃 它们, 得到 狐狸。
 lao3hu3 xun2qiu2 duo1zhong3 ye3shou4 er2qie3 chi1 ta1men1, de2dao4 hu2li2.
 tiger seek many animal to eat them, get fox.

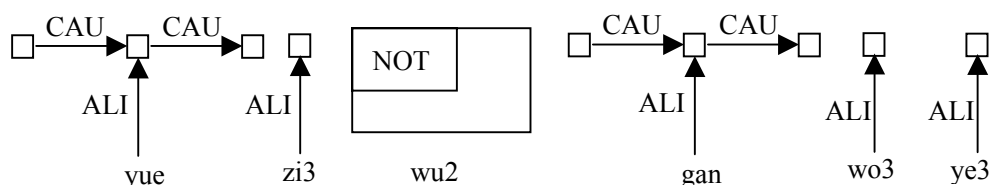
Sentence 3

The third sentence is “狐曰：子无敢食我也！” (hu2 yue4: zi3 wu2 gan3 shi2 wo3 ye).

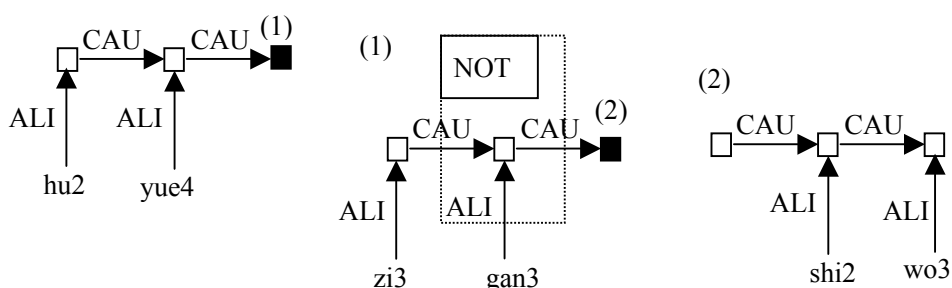
Step 1. The sentence in AC is the following.

狐 曰： 子 无 敢 食 我 也！
 hu2 yue4: zi3 wu2 gan3 shi2 wo3 ye!
 fox say: you not dare eat me!

Step 2. We have the following word graphs. (The words “狐 hu2” and “食 shi2” have been discussed, so we consider the remaining words).



Step 3. We construct the sentence graph by structural parsing. The sentence graph is



From the graph, we can see that the structural parsing of this sentence is the most difficult so far. Sentences in Chinese are not like the ones in English. In English, there is only ONE central verb in ONE sentence. But as we mentioned in [Hoede *et al.*, 2000], there is a situation that many verbs can exist in ONE Chinese sentence. In this sentence, there are three verbs: “曰 yue4, SAY”, “敢 gan3, DARE” and “食 shi2, EAT”.

The object of “say” is always a frame, this is EQU to “you not dare eat me”, which is “not you dare eat me”. The object of “dare” is also a frame, namely “(you) eat me”.

From the word graphs that we got and the sentence itself, we still cannot know *WHAT* is the subject of the verb EAT. This is not syntax problem. It is really a semantic problem. We see the following two examples.

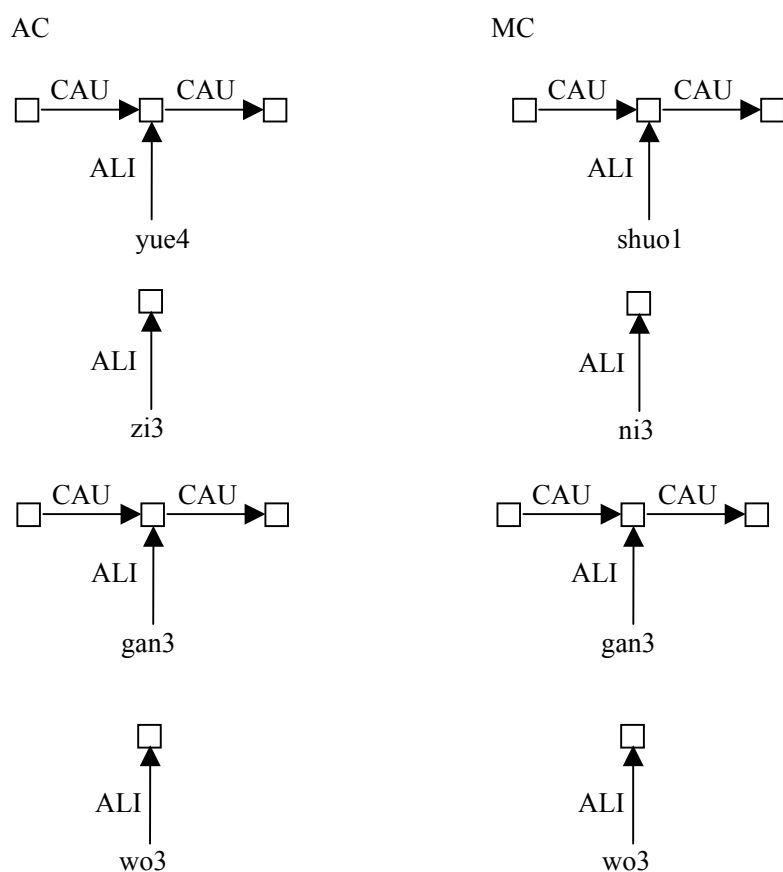
Example 1: “我帮你拿 wo3 bang1 ni3 na2” is a Chinese sentence. Where “我 wo3” means I, “帮 bang1” means HELP, “你 ni3” means YOU, and “拿 na2” means TAKE. Like in sentence 3, there are two verbs here. The SUBJECT of the verb TAKE is I, that is I TAKE.

Example 2: “我给你看 wo3 gei3 ni3 kan4” is another Chinese sentence. Where “我 wo3” means I, “给 gei3” means GIVE, “你 ni3” means YOU, and “看 kan4” means SEE. There are two verbs. But the SUBJECT of the verb SEE is not I, and it is YOU, that is YOU SEE.

So, in order to parse sentences, we have to use semantics. On the other hand, the syntactic analysis is not sufficient, but the semantic analysis is most important, especially in Chinese.

In [Liu & Hoede, to appear], we have discussed this problem.

Step 4. We have to express AC words in terms of MC words.



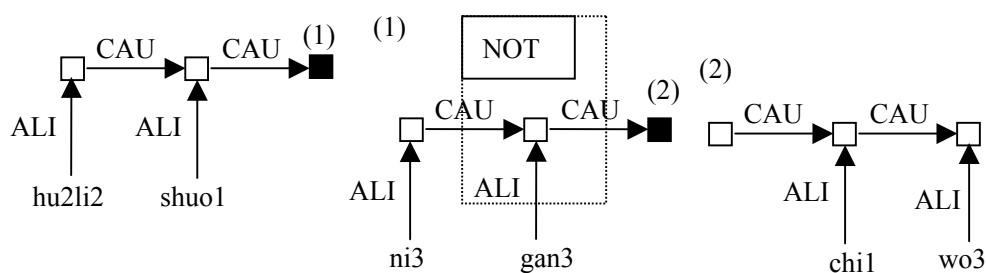
The word “曰 yue4” in AC is replaced by “说 shuo1” in MC. “子 zi3” is replaced by “你 ni3”. “敢 gan3” in AC is still “敢 gan3” in MC. “食 shi2” is replaced by “吃 chi1”. “我 wo3” in AC is still “我 wo3” in MC. These are very simple transformations.



“无 wu2” is the NEG-frame. In MC we use “不 bu4” expressing it.

“也 ye” is a modal particle.

Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

狐狸 说： 你 不 敢 吃 我 啊！
 hu2li2 shuo1: ni3 bu4 gan3 chi1 wo3 a!
 fox say: you not dare eat me a!

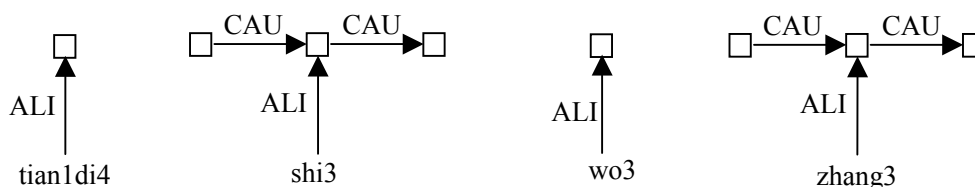
Sentence 4

The fourth sentence is “天帝使我长百兽” (tian1di4 shi3 wo3 zhang3 bai3 shou4).

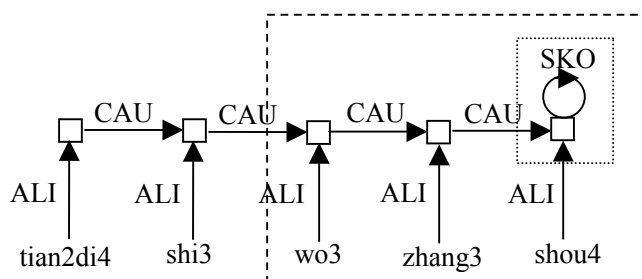
Step 1. The sentence in AC is the following.

天帝 使 我 长 百 兽。
 tian1di4 shi3 wo3 zhang3 bai3 shou4.
 God order me lead many animal.

Step 2. We have the following word graphs. (The words “百 bai3” and “兽 shou4” have been discussed).



Step 3. We construct the sentence graph by structural parsing. The sentence graph is



We see that this sentence is very simple. The graph is constructed by gluing the word graphs created in step 2 according to the order of the words.

Although this sentence is very simple to understand for Chinese people, it is not easy to understand for foreigners. There are two verbs in this sentence. As we discussed in [Hoede *et al.*, 2000], it is allowed to use many verbs in one Chinese sentence.

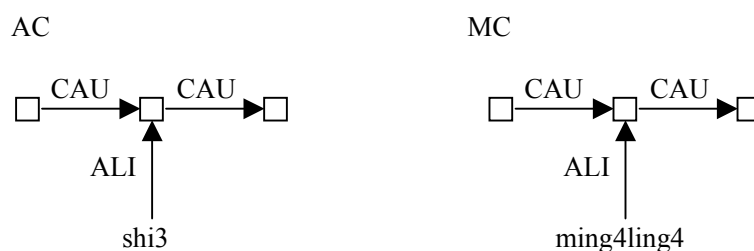
Step 4. We have to express AC words in terms of MC words.

The first word in the sentence is “天帝 tian1di4”.



Although we cannot say that “天帝 tian1di4” is not used in modern Chinese, we still use “上帝 shang4di4” in most cases.

The second word is “使 shi3”.



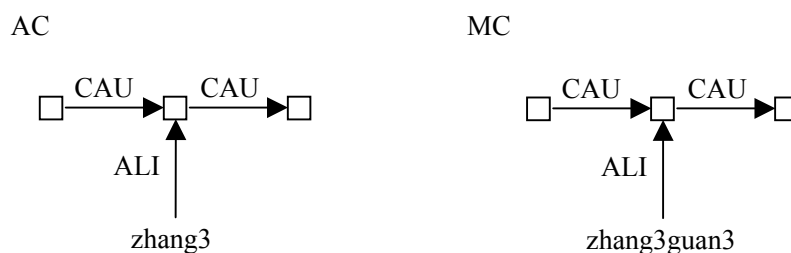
The word “使 shi3” in AC is expressed by the word “命令 ming4ling4” in MC. The basic meaning of “命 ming4” is ORDER, the basic meaning of “令 ling4” is still ORDER. There are a few differences. “命 ming4” is used as a verb to express an action “order”, “令 ling4” is used as a noun to express a “directive”. That is to say, the basic meaning of the word “命令 ming4ling4” is “order a directive”. But in modern Chinese, we use “命令 ming4ling4” to express the meaning ORDER, a verb.

The third word is “我 wo3”.



The representation of the word “我 wo3” is the same in AC and MC. As we discussed in [Hoede *et al.*, 2000], some words in AC are still used in MC.

The fourth word is “长 zhang3”.



This word is very difficult to express. In AC, the basic meaning of “长 zhang3” is MANAGER or LEADER as a noun. But its meaning in this sentence is “as the leader of”, or MANAGE. The same situation occurs in English in some cases. For example, the basic meaning of the word WATER is a noun. But in some case, it can be used as a verb “WATER the garden”. The word “掌管 zhang3guan3” has the meaning MANAGE or LEAD in MC. Here, the basic meaning of “掌 zhang3” is “palm”, the extension meaning is “be in charge of”. And “管, guan3” means a WIND INSTRUMENT originally, but now ADMINISTER is its extended meaning.

Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get the graph shown in Figure 4.

Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

上帝 命令 我 掌管 多种 野兽。
shang4di4 ming4ling4 wo3 zhang3guan3 duo1zhong3 ye3shou4.
God order me lead many animal.

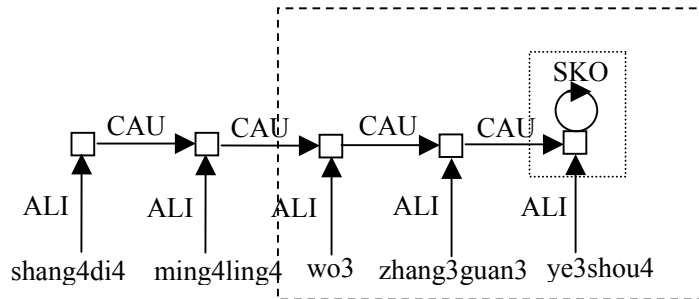


Figure 4

Sentence 5

The fifth sentence is “今子食我” (jin1 zi3 shi2 wo3).

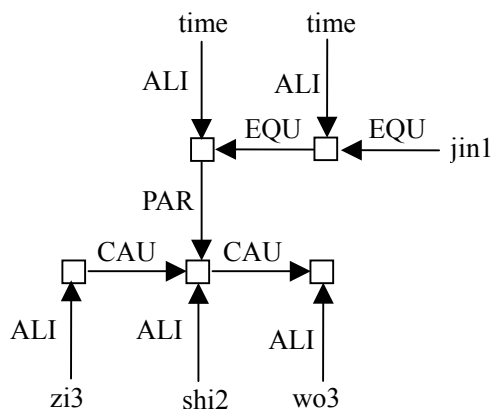
Step 1. The sentence in AC is the following.

今 子 食 我，
jin1 zi3 shi2 wo3,
now you eat me,

Step 2. We have the following word graphs. (The words “子 zi3”, “食 shi2” and “我 wo3” have been discussed, so we consider the remaining words).



Step 3. We construct the sentence graph by structural parsing. The sentence graph is



Here, we meet the expression of time. We use a PAR-link to express it.

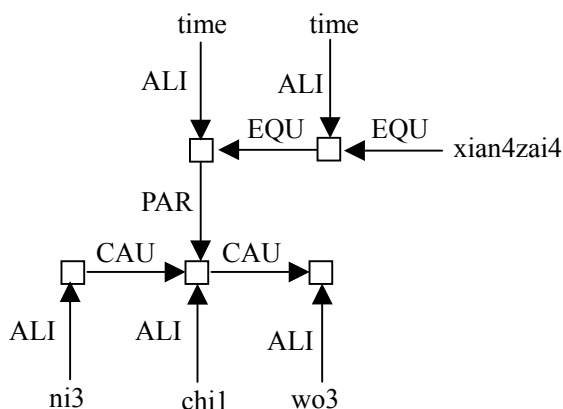
The interesting word is JIN or NOW. This is expressing that the time of speaking (YUE in AC) is equal to the time of eating (SHI in AC). There should be two PAR-arcs from tokens of type time,

whereas the tokens should be connected by an EQU-link.

Step 4. We have to express AC words in terms of MC words.

There are four words here. But three of them have been discussed in the above sentence. That is “子 zi3, 食 shi2, and 我 wo3”. The remaining word is “今 jin1”. We use “现在 xian4zai4” in MC to express it. The graph is the same. Where, the basic meaning of “现 xian4” is APPEAR, its extended meaning is PRESENT. The basic meaning of “在 zai4” is EXIST, its extended meaning is JUST. Now, “现在 xian4zai4” means NOW.

Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

现在 你 吃 我,
xian4zai4 ni3 chi1 wo3,
now you eat me,

Sentence 6

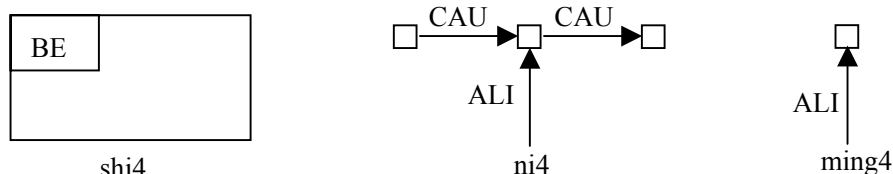
The second sentence is “是逆天帝命也!” (shi4 ni4 tian1di4 ming4 ye).

Step 1. The sentence in AC is the following.

是 逆 天帝 命 也!
shi4 ni4 tian1di4 ming4 ye.
be disobey God order (affirmation mood).

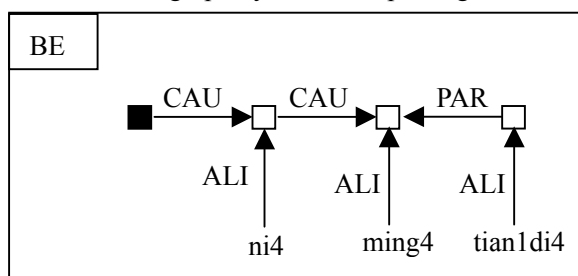
NOTE: This sentence is not a complete sentence. It is the latter part of the full sentence, the former part of the full sentence is the sentence 5.

Step 2. We have the following word graphs. (The word “天帝 tian1di4” has been discussed).



Here, the word “也 ye” is not represented. It is a modal particle. Its meaning is the affirmation for the sentence.

Step 3. We construct the sentence graph by structural parsing. The sentence graph is



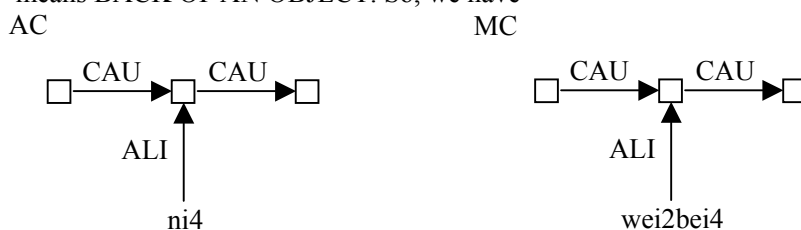
The solid token represents the former sentence. As in sentence 1, the PAR-link is not mentioned in the sentence. Adding it in the sentence graph is according to the method to build Chinese word. That is, “天帝 tian1di4” is the attribute of the word “命 ming4”.

Step 4. We have to express AC words in terms of MC words.

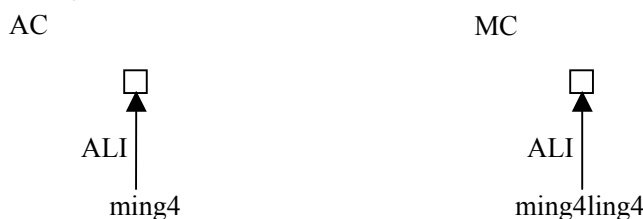
The word “是 shi4” is the same in AC and MC, so is the same representation.

The second word is “逆 ni4”.

As we discussed in [Hoede *et al.*, 2000], some words in AC have changed in MC. This word is one of them. In MC, it is represented by the word “违背 wei2bei4”. “违 wei2” means VIOLATE, and “背 bei4” means BACK OF AN OBJECT. So, we have



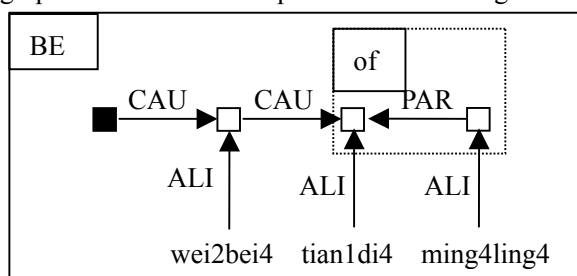
The next word is “命 ming4”.



The word “命 ming4” and “令 ling4” have the same basic meaning “order”.

The last word is “也 ye”, it is a mood particle to indicate the affirmation mood of the sentence. In MC, it is represented by the word “啊 a”.

Step 5. The sentence graph in AC has to be expressed in MC. we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

是 违背 天帝 的 命令 啊！

shi4 wei2bei4 shang4di4 de ming4ling4 a.

be disobey God of order (affirmation mood).

NOTE: Here, (的 de, of) is the word corresponding to the PAR-link.

Sentence 7

The seventh sentence is “子以我为不信 zi3 yi3 wo3 wei2 bu2 xin4”.

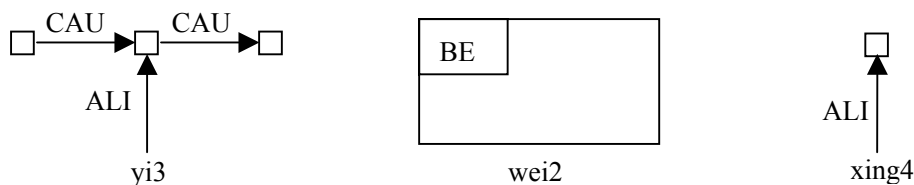
Step 1. The sentence in AC is the following.

子 以 我 为 不 信，

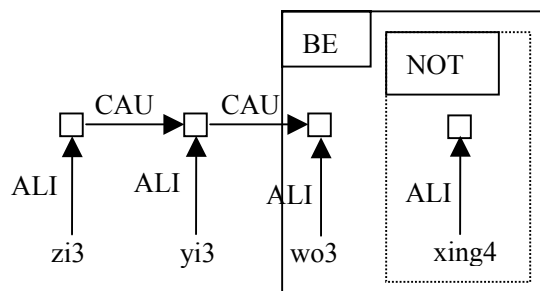
zi3 yi3 wo3 wei2 bu2 xin4,

you think I be not honest,

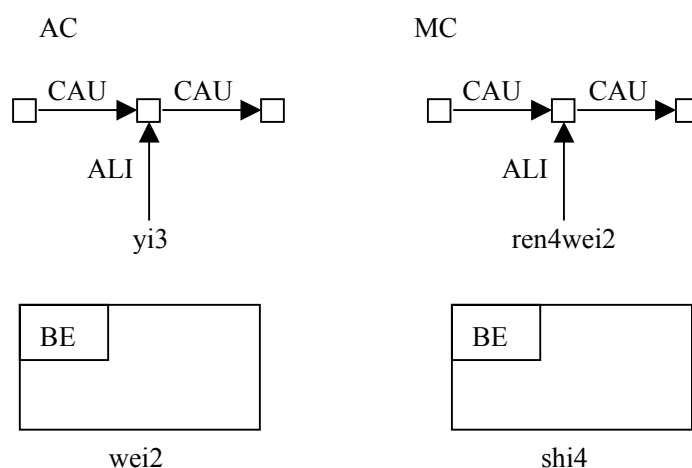
Step 2. We have the following word graphs. (The words “Zi3”, “Wo3”, and “Bu4” have been discussed, so we consider the remaining words).

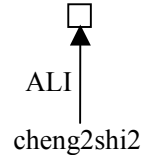


Step 3. We construct the sentence graph by structural parsing. The sentence graph is



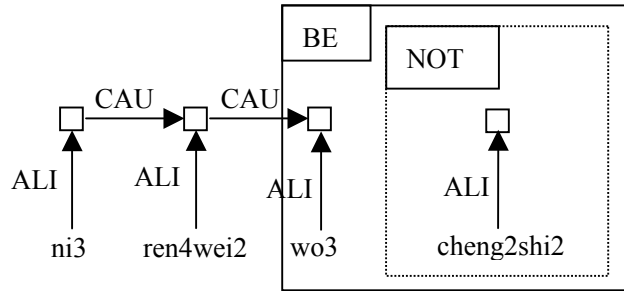
Step 4. We have to express AC words in terms of MC words.





The expressions are very clear. “以 yǐ3” in AC is replaced by “认为 ren4wei2” with the meaning THINK, where “认 ren4” means RECOGNIZE, and “为 wei2” means BELIEVE. (The basic meaning is FEMALE MONKEY). “为 wei2” is replaced by “是 shì4” with the meaning BE. “信 xìn4” is replaced by “诚实 cheng2shi2” with the meaning HONEST. Where, “诚 cheng2” has the meaning HONEST, and “实 shí2” is HONEST too.

Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

你 认为 我 是 不 诚实，
 ni3 ren4wei2 wo3 shi4 bu2 cheng2shi2,
 you think I be not honest,

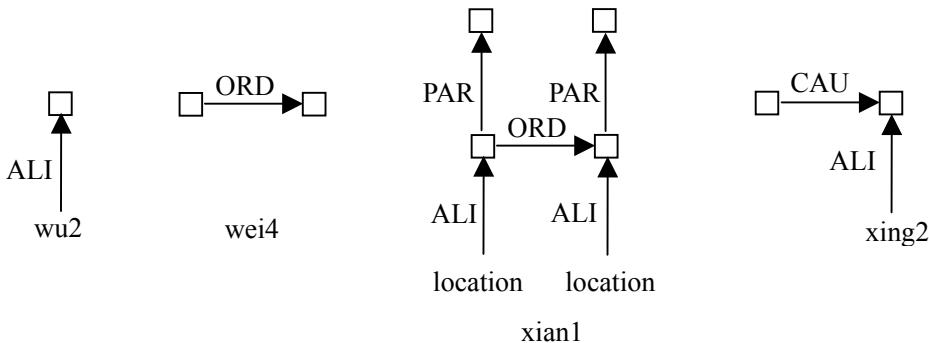
Sentence 8

The eighth sentence is “吾为子先行，” (wú wéi zǐ xián xíng).

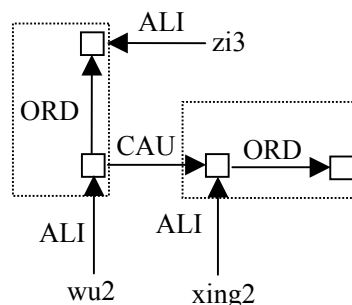
Step 1. The sentence in AC is the following.

吾 为 子 先 行，
 wu2 wei4 zi3 xian1 xing2,
 I for you ahead walk,

Step 2. Let us assume that we have the following word graphs.



Step 3. We construct the sentence graph by structural parsing. The sentence is very difficult to analyze. We know that the order of the words in the sentences is very important in Chinese sentences. We understand the simple sentence according to it. But in some complex sentences, we have to use the syntax (semantic) to give an analysis. The sentence graph that we found is



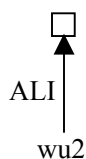
The two broken boxes indicate the function words “为 wei4, for” and “先 xian1, ahead”, respectively.

Step 4. We have to express AC words in terms of MC words.

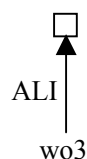
The first word is “吾 wu2”.

“吾 wu2” is a AC word. Its meaning is I. In MC, we use “我 wo3” to express it. So, we have

AC



MC

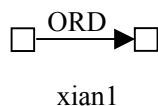


The second word is “为 wei4”, it is still a MC word. The representation is the same.

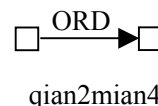
The next word is “先 xian1”.

“先 xian1” is an AC word, it is still a MC word. Its meaning is AHEAD. But, as we mentioned in [Hoede *et al.*, 2000], in MC we always use the disyllabic or polysyllabic instead of the monosyllabic. So we use “前面 qian2mian4” to express AHEAD.

AC



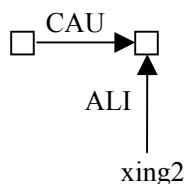
MC



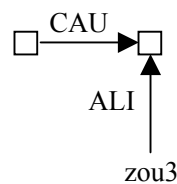
Here “前 qian2” means AHEAD, and “面 mian4” means FACE. The meaning of “前面 qian2mian4” is still AHEAD.

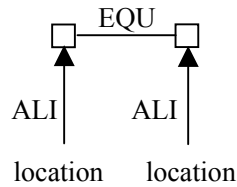
The fifth word is “行 xing2”.

AC



MC





can be interpreted as that of “在 zai4” or “at”. So we have found an example where “先 xian1” is naming a subgraph that is larger than that of “前 qian2” and also contains the “face” concept “面 mian4”.

Sentence 9

The ninth sentence is “子随我后 zi3 sui2 wo3 hou4”. It is a simple sentence.

Step 1. The sentence in AC is the following.

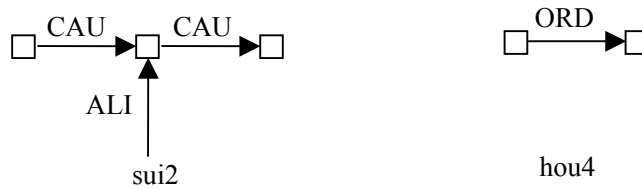
子 随 我 后。

zi3 sui2 wo3 hou4.

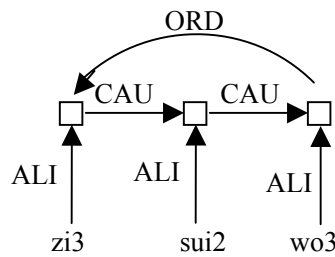
you follow I behind.

Step 2. We have the following word graphs.

(The words “子 zi3” and “我 wo3” have been discussed, so we consider the remaining words).



Step 3. We construct the sentence graph by structural parsing. The sentence graph is

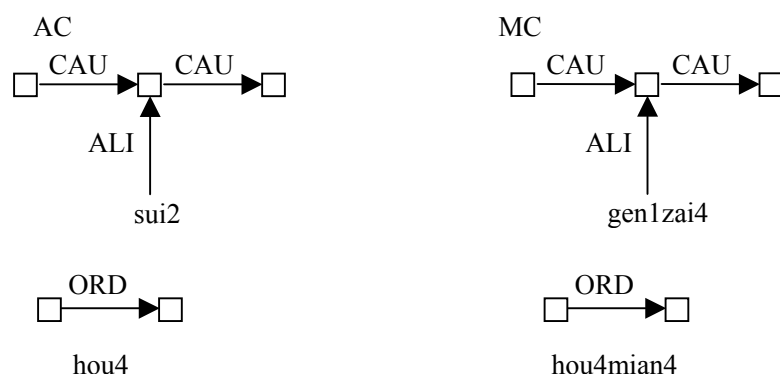


“后 Hou4” is a preposition expressing the ordering of two locations, namely of “zi3” and of “wo3”. It is similar to “zai4” or “at” but now an ORD-arc is present instead of an EQU-link. We refer to [Hoede & Li, 1996], in which the prepositions are discussed. Without further explanation it is not clear how the ordering is to be understood. One might even say that the ordering is somehow included in “sui2”, as FOLLOW implies an ordering in space.

Step 4. We have to express AC words in terms of MC words.

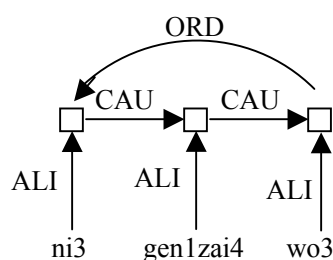
The remaining words are “随 Sui2” and “后 Hou4”. “随 Sui2” is expressed by “跟在 gen1zai4” in MC with the meaning FOLLOW. Where “跟 gen1” has the meaning FOLLOW (the basic meaning is HEEL), and “在 zai4” means AT (the basic meaning is EXIST). “后 Hou4” is expressed by “后面 hou4mian4” in MC with the meaning BEHIND. “后 Hou4” has the meaning BEHIND (the basic meaning is SOVEREIGN), and “面 mian4” has the meaning FACE. The

structure is the follows.



“在 zai4” means AT. This is indicating a location of “跟 gen1”, the verb. “随 sui2” has developed into “跟在 gen1zai4”, that is now seen as one word. “后 hou4” has the same graph as “后 hou4” in AC, but the locations have shifted to “面 mian4”. In the MC version it says something like *you follow my backface*, i.e. *behind me*. Again it may be that “后面 hou4mian4” is now standard in MC. But it is interesting how the concept BEHIND has evolved from “后 hou4” to “后面 hou4mian4” in connection with the possessive “的 de” in “我的 wo3de”.

Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

你 跟 在 我 的 后 面。
 ni3 gen1zai4 wo3 de hou4mian4.
 you follow I of behind.

Like in sentence 8 we could have given a larger word graph for “qian2mian4”, here “hou4mian4” could be described by a larger word graph so that the development from the AC “hou4” to MC becomes clear.

Sentence 10

The tenth sentence is “观百兽之见我而敢不走乎？”(guan1 bai3 shou4 zhi1 jian4 wo3 er3 gan3 bu4 zou3 hu?)

Step 1. The sentence in AC is the following.

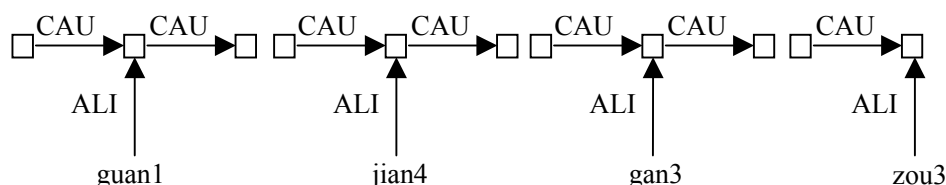
观 百 兽 之 见 我 而 敢 不 走 乎？
 guan1 bai3 shou4 zhi1 jian4 wo3 er3 gan3 bu4 zou3 hu?
 observe many animal see I and dare not run (ask in retort)?

Here, there are two function words in the sentence. One is “之 zhi1”, it has only grammatical meaning. It is used to abrogate the independence of the sentence when it is located between the subject “百兽 bai3shou4” and the predicate “见 jian4”. The other is “乎 hu”, it is a modal particle to express asking in retort.

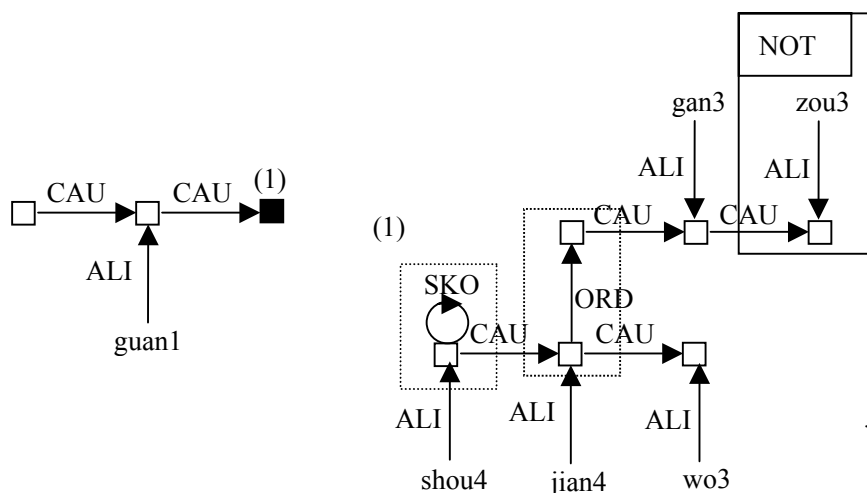
Note: (1) Asking the sentence in retort means that the sentence is not an real question. There are no questions but it uses the question form.

(2) There are two explanations for the word “之 zhi1”. The first is that it has not a word meaning, it is just used to combine two parts of the sentence. The second is that it is a pronoun to indicate the former noun. In this sentence, it indicates “百兽,bai3shou4”. The purpose is to emphasize.

Step 2. We have the following word graphs.



Step 3. We construct the sentence graph by structural parsing. The sentence graph is

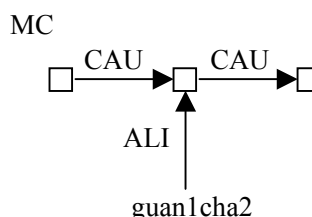
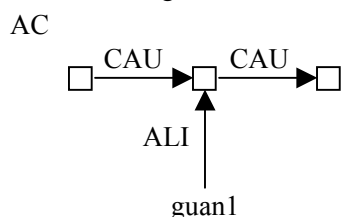


This is a complex sentence. So, we represent the two parts. There exists an FPAR-link between these two parts. In the sentence, there are three verbs (观 guan1 OBSERVE, 见 jian4 SEE, and 走 zou3 RUN), and one auxiliary (敢 gan3 DARE).

Step 4. We have to express AC words in terms of MC words.

The words “百 bai3, 兽 shou4, 我 wo3, 不 bu4 and 而 er3” have been represented, we need not discuss them here.

The first word is “观 guan1”.

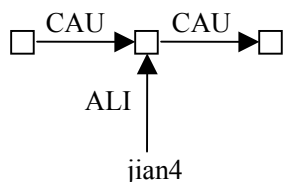


Here, in MC “观 guan1” and “察 cha2” have the same meaning. The meaning is SEE CLEARLY or OBSERVE.

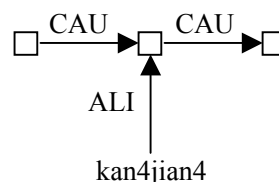
The next word is “之 zhi1”. There is no corresponding usage in MC.

The fifth word is “见 jian4”.

AC



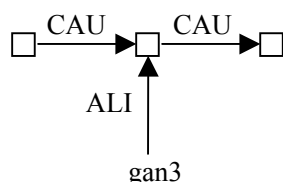
MC



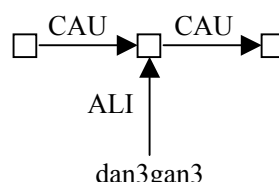
In MC, we use “看见 kan4jian4” to express “见 jian4” in AC. Here, “看 kan4” means WATCH, and “见 jian4” means SEE. But “看见 kan4jian4” still has the meaning SEE.

The eighth word is “敢 gan3”. It is an auxiliary word.

AC



MC

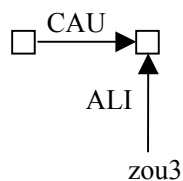


In MC, we use “胆敢 dan3gan3” to express “敢 gan3” in AC. Here, “胆 dan3” means “courage”, and “敢 gan3” means “dare”. “胆敢 dan3gan3” means still “dare”.

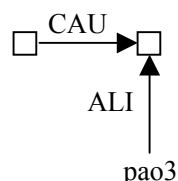
The tenth word is “走 zou3”.

As we mentioned in [Hoede *et al.*, 2000] the meaning of the word “走 zou3” in AC shifted to the word “跑 pao3” in MC, that is RUN. So, we have

AC



MC



The last word is “乎 hu”. It is a modal particle to express asking in retort. In MC, we use “吗 ma” to express it.

Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get the graph shown in Figure 5.

Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

观察 许多 野兽 看见 我 而 胆敢 不 跑 吗?

guan1cha2 xu3duo1 ye3shou4 kan4jian4 wo3 er3 dan3gan3 bu4 pao3 ma1?

observe many animal see I dare not run (ask in retort)?

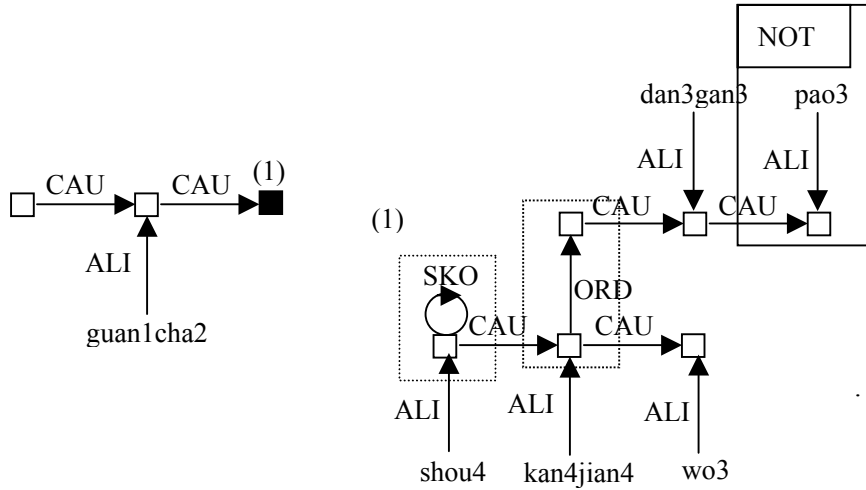


Figure 5

Sentence 11

The eleventh sentence is “虎以为然 hu3 yi3 wei2 ran2”. It is a simple sentence.

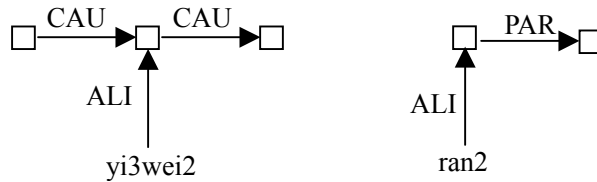
Step 1. The sentence in AC is the following.

虎 以为 然。

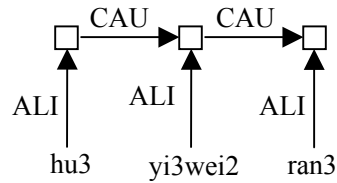
hu3 yi3wei2 ran2.

tiger think right.

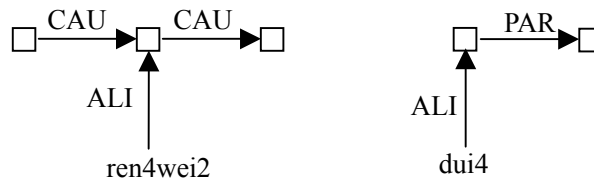
Step 2. We have the following word graphs.



Step 3. We construct the sentence graph by structural parsing. The sentence graph is



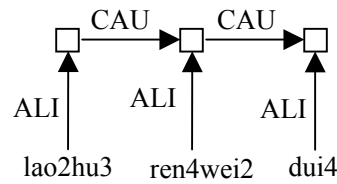
Step 4. We have to express AC words in terms of MC words.



“以为 yi3wei2” is replaced by “认为 ren4wei2” in MC with the meaning THINK, and “然 ran2” is replaced by “对 dui4” in MC with the meaning RIGHT. The analysis is simple.

Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4,

we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

老虎 认为 对。

lao2hu3 ren4wei2 dui4.

tiger think right.

Note that both for “ran2” and for “dui4” we have indicated that the “rightness” is an attribute of something, to which the unlabeled token refers. In this case the reference is, of course, to the former sentences, the proposals of the fox.

Sentence 12

The twelfth sentence is “故遂与之行。” (gu4 sui2 yu3 zhi1 xing2).

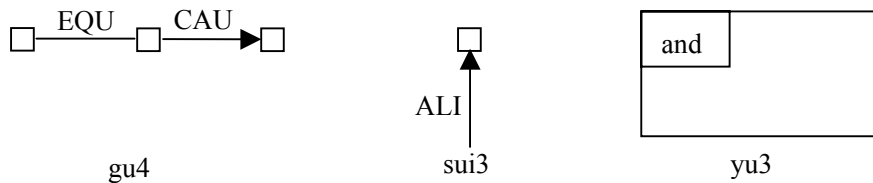
Step 1. The sentence in AC is the following.

故 遂 与 之 行。

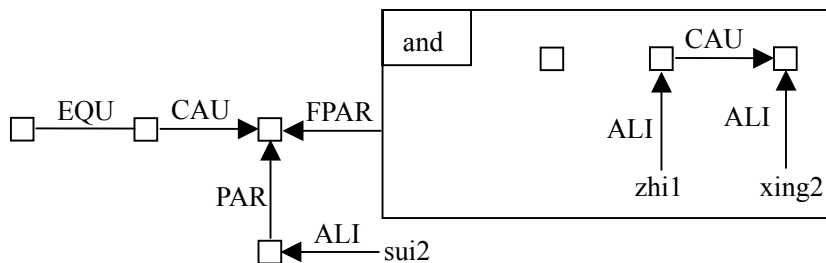
gu4 sui2 yu3 zhi1 xing2.

therefore right now and it go.

Step 2. We have the following word graphs. (The words we did not discuss before).



Step 3. We construct the sentence graph by structural parsing. The sentence graph is



Step 4. We have to express AC words in terms of MC words.

The first word “故 gu4” is still an MC word, but we use “因此 yin1ci3” to express it here. Its meaning is THEREFORE.



Here, “因 yin1” means BECAUSE, “此 ci3” means THIS, and is describing the EQU-link. The word “因此 yin1ci3” means THEREFORE or BECAUSE THIS.

The second word is “遂 sui2”.

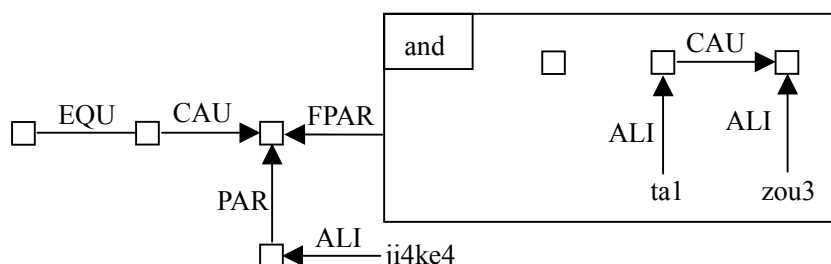
The meaning of the word “遂 sui2” is RIGHT NOW. There is no usage of this word in MC. In MC, we use “即刻 ji4ke4” to express it. Here, “即 ji4” means APPROACH, and “刻 ke4” means RESTRICT. In principle this asks for a detailed word graph explaining why this replacement has evolved in MC. We have another example of the fact that in AC words are used with large word graphs, which are brought under words in MC by two syllables. Hence we have a case where the intersection of meanings occurs.



Because the word “即可 ji4ke4” is an adverb, we add a PAR-link in the sentence graph.

The third word is “与 yu3”. It is still an MC word, and has the same meaning and representation.

Step 5. The sentence graph in AC has to be expressed in MC. We get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

因此 即刻 与 它 走。
 yin1ci3 ji4ke4 yu3 ta1 zou3.
 therefore right now and it go.

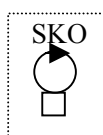
Sentence 13

The thirteenth sentence is “兽见之皆走 shou4 jian4 zhi1 jie2 zou3”.

Step 1. The sentence in AC is the following.

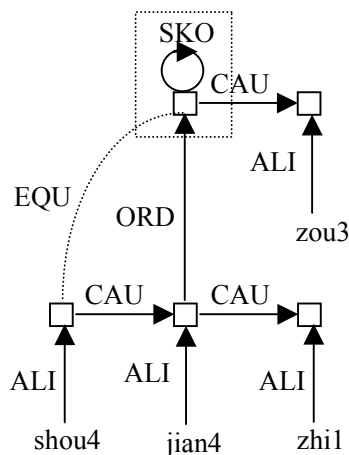
兽 见 之 皆 走。
 shou4 jian4 zhi1 jie2 zou3.
 animal see them all run.

Step 2. We have the following word graphs. (The words we did not discuss before).



jie2

Step 3. We construct the sentence graph by structural parsing. The sentence graph is

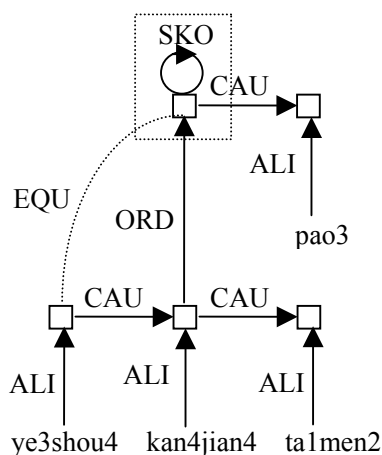


It is not a complex sentence. But there are two verbs here. One is “见 jian4 SEE”, and another is “走 zou3 RUN”. The subject of “见 jian4 SEE” is “兽 Shou4 ANIMAL”, and the object is “之 zhi1 THEM”. That is OK. But what is the subject of “走 zou3 RUN”? “兽 Shou4 ANIMAL” or “之 zhi1 THEM”? As we discussed for sentence 3, we need to use semantics to decide.

Step 4. We have to express AC words in terms of MC words.

The remaining word is “皆 jie2, ALL”. It is a quantity word. In MC we use “都 dou1, ALL” to express it.

Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

野兽 看见 它们 都 跑。
 ye3shou4 kan4jian4 ta1men2 dou1 pao3.
 animal see them all run.

Sentence 14

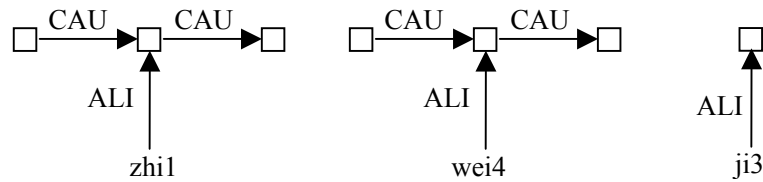
The fourteenth sentence is “虎不知兽畏己而走之，”(hu3 bu4 zhi1 shou4 wei4 ji3 er3 zou3 zhi1).

Step 1. The sentence in AC is the following.

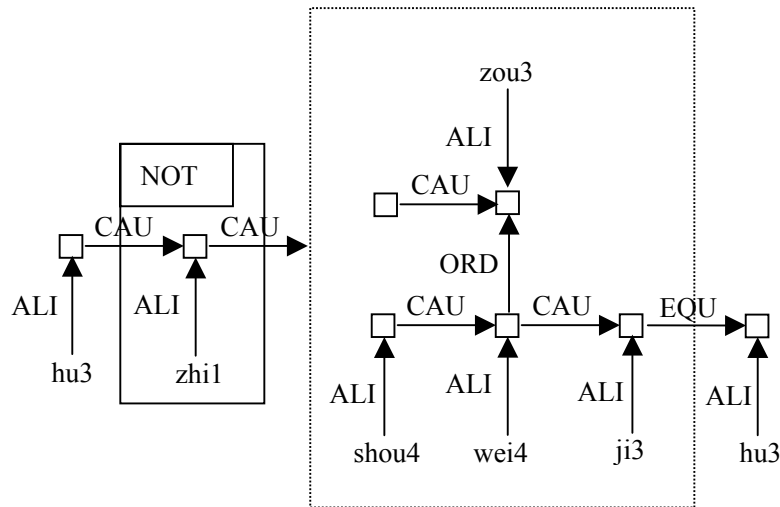
虎 不 知 兽 畏 己 而 走 之，
 hu3 bu4 zhi1 shou4 wei4 ji3 er3 zou3 zhi1,
 tiger not know animal fear itself to run,

Note, we have another function word “之 zhi1”. It has no meaning, or rather we do not know its meaning now. It is a void. We can check this by deleting it without change the meaning of the sentence.

Step 2. We have the following word graphs. (The words we did not discuss before).

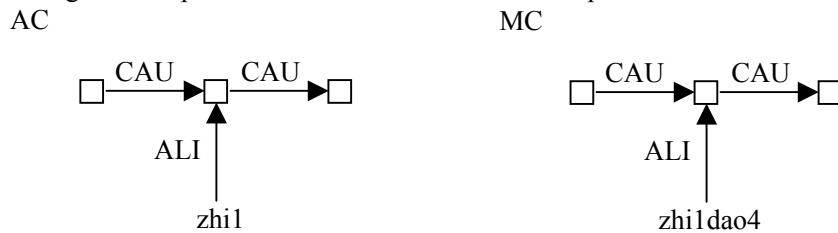


Step 3. We construct the sentence graph by structural parsing. This sentence is not so difficult, but there still exist three verbs here. The sentence graph is



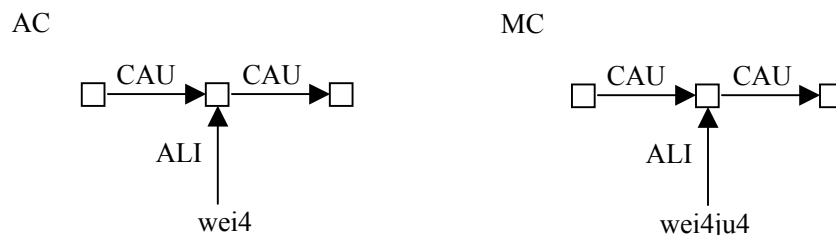
Step 4. We have to express AC words in terms of MC words.

The first remained word is “知 zhi1”. It is still an MC word. Now we still use it to express the same meaning. But we prefer to use “知道 zhi1dao4” to express it.



Here, “知 zhi1” means KNOW, and “道 dao4” means SAY. But “知道 zhi1dao4” is still KNOW.

The second remaining word is “畏 wei4”. Like the word “知 zhi1”, it is still an MC word. Now we still use it to express the same meaning. But we prefer to use “畏惧 wei4ju4” to express it.

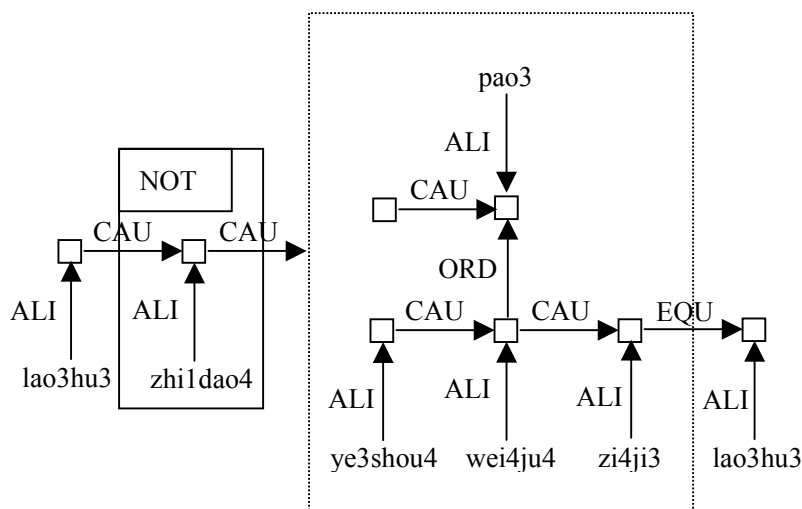


Here, “畏 wei4” means FEAR, and “惧 ju4” also means FEAR. So “畏惧 wei4ju4” is FEAR.

The third remained word is “己 ji3”. It is a pronoun. In MC, we use “自己 zi4ji3” to express it. “自 zi4” basically means self. The pronoun aspect is covered by “己 ji3”, and describes the EQU-link.



Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

老虎 不 知道 野兽 畏惧 自己 而 跑,
lao3hu3 bu4 zhi1dao4 ye3shou4 wei4ju4 zi4ji3 er3 pao3,
tiger not know animal fear itself to run,

Sentence 15

The last sentence is “以为畏狐也 yi3wei2 wei4 hu2 ye3”.

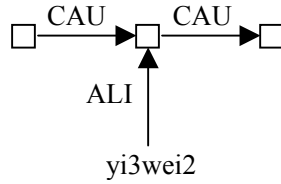
Step 1. The sentence in AC (Ancient Chinese) is the following.

以为 畏 狐 也。

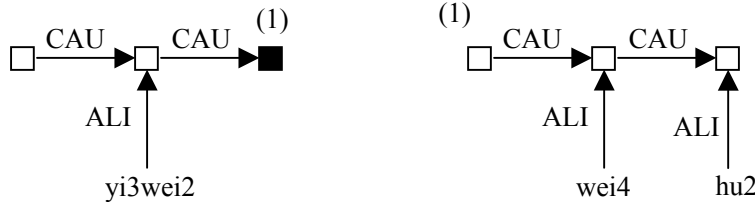
yi3wei2 wei4 hu2 ye3.

think fear fox.

Step 2. We have the following word graphs. (The words we did not discuss before).

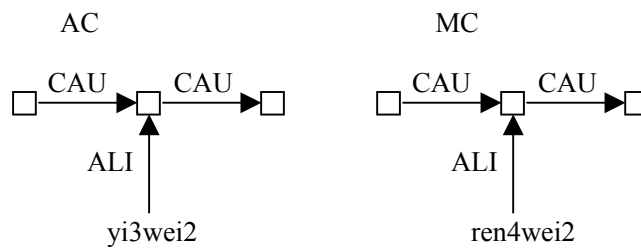


Step 3. We construct the sentence graph by structural parsing. The sentence graph is



One token is “free” here, i.e. it refers to something because it is not a complete sentence. Actually, it refers to “虎 Hu3, TIGER”.

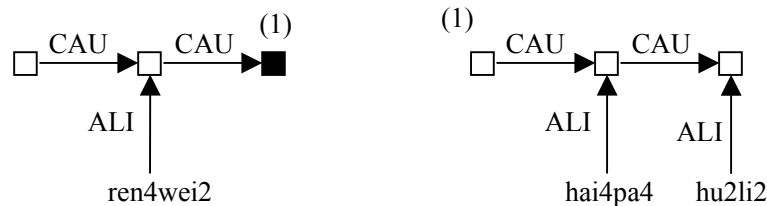
Step 4. We have to express AC words in terms of MC words.



The word “以为 yi3wei2, THINK” is a disyllabic word, like “天帝 tian1di4, GOD”. In MC we use “认为 ren4wei2, THINK” to express it.

The word “也 ye” is a modal particle.

Step 5. The sentence graph in AC has to be expressed in MC. According to the analysis in step 4, we get



Step 6. The final step is expressing this sentence graph in syntactically correct MC. That is

认为 害怕 狐狸 呢。

ren4wei2 hai4pa4 hu2li2 ne.

think fear fox.

4 Concluding remarks

In this paper, we gave an idea of translation from ancient Chinese to modern Chinese by using an example story by means of knowledge graphs. Actually, we give the details of the method of translation from ancient Chinese to modern Chinese step by step as carried out by hand. From the example, we found that knowledge graphs have a strong ability to represent sentences. But we also found that there are things that should be discussed if we want to translate from one kind of language to another kind of language by using knowledge graph theory automatically.

(1) Although in ancient Chinese, almost all words are monosyllabic words, there are still some disyllabic words, such as “天帝 tian1di4, GOD” and “以为 yi3wei2, THINK” in the article. That is to say, still there is a problem to segment words in sentences.

(2) Another problem we encountered is structuring parsing, that is how to glue word graphs for words in a sentence into sentence graphs. As we mentioned in this paper, the word order and function words in sentences are very important in Chinese. But, only according to these two features we cannot glue the word graphs into sentence graphs correctly. We found that semantics is the third feature to understand Chinese. So, how to make structuring parsing is still a problem for achieving translation from one kind of language to another kind of language automatically. Even for “close” languages as ancient and modern Chinese we saw in several sentences that the chosen word graphs for the AC words should be considerably large, in order to understand how the development to the MC way of expressing has taken place.

In most sentences, however, we see that the translation step, replacing word graphs in AC by word graphs in MC is straight forward, as was to be expected. The most important features are the ones already recovered in Section 2.

The fact that both syntax and semantics play an important role in mapping a sentence on a sentence graph was also found in the paper of Zhang [Zhang, 2002], that focuses on structural parsing, in an attempt to develop an automatic way of knowledge extraction.

(3) The third problem is the uttering, that is to say, how to generate the syntactically correct target sentence from the given sentence graphs. In order to do this, the grammar and sentence patterns in target language is very important. Zhang [Zhang, in preparation] also made a start with the study of utterance paths.

References

[Chen, 1979] Chen PuQing, *Basic knowledge of ancient Chinese*, Yuelu Publishing House, 1979. (in Chinese)

[Hoede *et al.*, 2000] C. Hoede, X. Li, X. Liu and L. Zhang, *Knowledge Graph Analysis of some Particular Problems in the Semantics of Chinese*, Memorandum No.1516, Department of Applied Mathematics, University of Twente, Enschede, The Netherlands, 2000.

[Hoede & Li, 1996] C. Hoede and X. Li, *Word Graphs: The First Set*, in *Conceptual Structures: Knowledge Representation as Interlingua*, Auxiliary Proceedings of the Fourth International

Conference on Conceptual Structures, Bondi Beach, Sydney, Australia (P.W. Eklund, G. Ellis and G. Mann, eds.), ICCS '96, (1996) 81-93.

[Hoede & Zhang, 2001] C. Hoede and L. Zhang, *Structural Parsing*, in: Conceptual Structures: Extracting and Representing Semantics, Aux. Proc. of the 9th International Conference on Conceptual Structures (Ed. G. W. Mineau), CA, USA, 75-88, 2001.

[Liu & Hoede, to appear] X. Liu and C. Hoede, *Passive Sentences and Structural Parsing*, to appear.

[Tang, 1982] Tang Yan, The American scholar Tang Degang's option to latinization of Chinese character, *Zhongbao monthly*, Hong Kong, 4, 1982.

[Zhang, 2002] L. Zhang, *Knowledge graph theory and structural parsing*, Ph.D. thesis, University of Twente, Enschede, The Netherlands, ISBN 9036518350, 2002.