
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. 1659

Extracting causal relationships
from Chinese written text

X. LIU¹ AND C. HOEDE

NOVEMBER, 2002

ISSN 0169-2690

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

Extracting Causal Relationships from Chinese Written Text*

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

Expert systems form one of the most important research areas in Artificial Intelligence. The main parts in expert systems are knowledge bases and inference engines. In the knowledge bases the main knowledge is knowledge in the form of "IF-THEN" statements. In knowledge graphs, a new form of knowledge representation, the "IF-THEN" statements are tied up with causal operators (CAU-relations). In this paper, we picked out some Chinese operators with "CAU" meaning, and investigated these operators. We also show by an example how to extract causal relations from a given Chinese writing text.

Key Words: expert system, knowledge graph, knowledge base, CAU-word

AMS Subject Classifications: 05C99, 68F99

1 Introduction

Expert systems are sophisticated computer programs that manipulate knowledge to solve problems effectively and effectively in a narrow problem area. Like real experts, these systems use symbolic logic and heuristics -- rules of thumb -- to find solutions. And like real experts, they make mistakes but they have the capacity to learn from their errors. However, these artificial expertise has some advantages over human expertise: It is permanent, consistent, easy to transfer and

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

document, and cheaper. In sum, by linking the power of computers to the richness of human experience, expert systems enhance the value of expert knowledge by making it readily and widely accessible.

The main parts in expert systems are knowledge bases and inference engines. The knowledge used to solve the problems in a narrow problem area is stored in knowledge bases. The knowledge in knowledge bases is mainly in the form of “IF-THEN” statements. In order to solve different narrow area problems, there are different knowledge representation methods in different knowledge bases. The knowledge in knowledge bases determined the function, efficiency and effectiveness of the expert systems. An inference engine uses the knowledge in knowledge bases to solve the problems.

In a document, one distinguishes *operators* and *arguments* in a sentence. An argument is a self-contained word or group of words, whereas an operator is a word or group of words that is dependent on one or more other arguments. The arguments are viewed as concepts, and the operator one finds is mapped as relation.

In this paper, we picked out some Chinese words (operators) with seemingly causal meanings (*i.e.* CAU-operators), and classified these operators into groups according to the different argument types connected by these causal operators. The main goal is to use such *indicators* of CAU relationships in an automatic extraction system. This is in line with the thesis of de Vries, who studied such indicators for English. However, before we can start automatic extraction, we should investigate which particular problems are encountered in Chinese.

2 CAU-word representation in terms of knowledge graphs

In knowledge graphs theory, the “causal” statements are represented by CAU-operators and corresponding arguments. That is to say, the statement of “A is the causation of B” is represented by the following graph, where CAU is a causal operator, and A and B are corresponding arguments.



Figure 1: A CAU-relationship.

Here, “A CAU B” means: *An occurrence of or a change in A can cause an occurrence of or a change in B.*

Any linguistic phrase that may thus be rephrased is an indicator of a CAU relationship.

The “IF-THEN”-statement is not equivalent to the CAU statement. As was discussed by Hoede and Zhang [Hoede & Zhang, 2001] the knowledge graph representation of “IF-THEN” is in terms of frames (see Figure 2) and expresses the logical equivalence $(p \rightarrow q) \leftrightarrow (\neg(p \wedge \neg q))$.

In Figure 2 we use the NEG-frame to express the negation connective \neg and the AND-frame to express the connective \wedge , that together form a functionally complete set of connectives for

propositional logic.

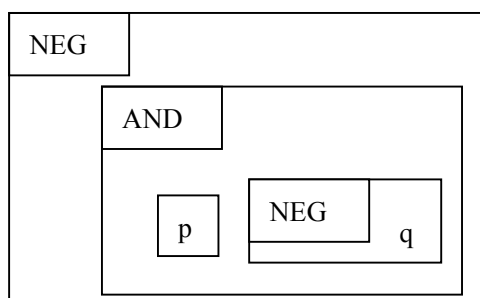


Figure 2: Knowledge graph representation of the “IF-THEN” statement

Let a certain change ΔA in A cause a certain change ΔB in B . We may now state “IF ΔA occurs THEN ΔB occurs”. So, “ ΔA occurs” plays the role of p and “ ΔB occurs” plays the role of q in the representation given above. In this way the “IF-THEN”-relationship ties up with the CAU relationship. In Chinese, there are many words that can be seen as operators with “causal” meaning, but especially argument types connected by these causal operators are different.

Example 1 “按照计划,我们写文章。” is a Chinese sentence. In “pinyin”, spelling, the sentence reads, “an4zhao4 ji4hua4 wo3men2 xie3 wen2zhang1”. The numbers refer to the four kinds of intonation. The literal meaning is “According to plan we write paper”. The partial representation by a sentence graph is shown in Figure 3.

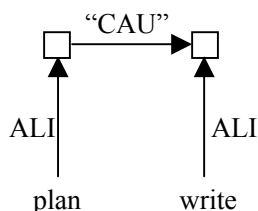


Figure 3: A partial sentence graph

We see that the central words in the sentence are “plan” and “write”. That is to say, the “plan” is related to “writing”. Here, “按照, an4zhao4”, according to, is the operator to express this relationship and connects two arguments: “plan” and “write”.

We have indicated this in Figure 3 by an arc with label CAU, as the relationship is certainly not that of a subject “计划, ji1hua4”, plan, to a verb “写, xie3”, write. In fact “according to” may be replaced by “conform to” or just by “alike”. Some aspect of the writing shows similarity with the plan, let this be the “chosen way” of writing, something not mentioned in the sentence, then the likeness of “计划, ji1hua4”, plan, and “chosen way”, to which “按照, an4zhao4”, according to, refers, becomes clear, but also the causal nature of the relationship between “计划, ji1hua4”, plan, to a verb “写, xie3”, write. The plan *influences* the choice of the way of writing of the paper. Another plan would *lead to* another way of writing. So “按照, an4zhao4”, according to, can be seen as a causal operator, be it that this only becomes clear on *expansion* of the concepts. Representation by a simple causal arc can be done but then its type should be “CAU”, indicating that the causality only becomes clear when a much more elaborate sentence graph would be given. This phenomenon occurs quite often in natural language.

Example 2 “黑暗暗含着危险。 heilan4 an4han2zhe1 weilxian3” is another Chinese sentence.

The meaning is “darkness imply danger”. The representation by a sentence graph is as in the Figure 4.

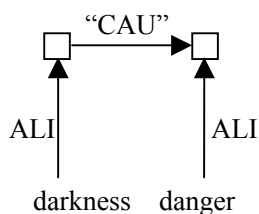


Figure 4: A sentence graph

We see that the “黑暗, he1an4”, darkness, is said to be the cause of “危险, wei1xian3”, danger.

“暗含着, an4huan2zhe3”, imply, is the operator expressing “causation”, and connecting two arguments, *i.e.* “黑暗, he1an4”, darkness, and “危险, wei1xian3”, danger. As in Example 1, the “causation” only becomes clear on expansion. Dark places may be dangerous, but the implication is by no means strict. The main point, however, is that “暗含着, an4huan2zhe3”, imply, is a causal operator, which might have been taken up into these graph as a verb, like the verb “写, xie3”, write, in Figure 3.

From these two examples we can see that different operators with CAU meanings (Chinese “causal” words) connect different types of arguments. In Chinese sentences often other pairs of types than noun and verb are connected by CAU-words.

3 The collected operators

Like in English certain Chinese words are clearly CAU-words. Examples of easily recognized indicators of a CAU-link in English are “causes” or “leads to”, see de Vries [de Vries, 1989]. Also the word “by” in “hereby” is such a clear indicator, although the arguments are posing the problem to determine where “here” refers. However, we saw in the examples that indicators may less obviously show their causal content. Only a more detailed analysis of the meaning of the word, its word graph, may bring this content forward.

The problem that we are facing here again is that of the undetermined boundary of a relationship. If A and B are related in some way, this is expressed in knowledge graph theory by a knowledge graph containing A and B. The structure of the graph describes the way A and B are related. A CAU-link may be part of the graph like in the analysis of “按照, an4zhao4”, according to, in Example 1. For that reason we may call the relationship between “计划, ji1hua4”, plan, and “写, xie3”, write, a causal relationship. However, the graph that describes the precise way of influencing the writing by the plan is much more complicated than just a CAU-link between two tokens with label “plan” and “write”.

As a matter of fact, in the knowledge graph formalism transitive verbs are represented by a labeled token related to two other tokens by CAU-links. One of these is the *agent* of the verb and the other is the *patient* of the verb. In the syntactic graph of a verb SKO-links are used, one token is the *subject* and the other is the *object*. An agent is seen as “causing” the verb act to take place and the verb act is seen as “causing” the patient the experience that is due to the act. In both cases the causal relationship is of complex nature too. In “we write paper” both the precise content of the relationship between “we” and “write” and that of the relationship of “write” and “paper” would

ask for a much larger graph.

In principle each word should have a word graph that characteristically differs from the word graph of another word. This also holds for words used as CAU-operators. We therefore have in principle the duty to study each of these operators and to give a characteristic word graph for them. In this analysis it may turn out that the same graph is used for two different words, but that the difference consists of the focus laid on one of its tokens. This way $\boxed{X} \text{---CAU} \rightarrow \square$ and $\square \text{---CAU} \rightarrow \boxed{X}$ should be described by different words, when the crosses indicate the focus. We shall see how this is encountered in Chinese.

The extraction of CAU relationship from a text leads to a set of arcs of type CAU that may be used to construct a directed graph. This directed graph can then be used either as an expert system or as a decision support system. These points of the graph will in many cases be labeled by nouns. As we mentioned in the introduction, usually expert systems are formulated in terms of IF-THEN rules. We choose to focus on the graph with causal links and therefore represent operators between arguments that are statements also by causal links, which asks for some remarks.

The sentence “IF p THEN q” could be represented by $p \text{---CAU} \rightarrow q$. In accordance with our interpretation of the CAU-link we have chosen “IF p THEN q” to mean “The occurrence (truth) of p cause the occurrence (truth) of q”. We have restricted the truth table to the case in which both statements p and q are true. In pure logic the statements p and q need not have a relationship at all. In case we want to build an expert system it is, of course, important to determine where the causal relationship between p and q lies in detail. In “IF p THEN q”, p can e.g. play the role of a *condition* for the occurrence of q. When q is the description of a process, a chemical reaction, and p is the description of a numerical condition, say on the temperature, we can represent the essence of the statement by a causal arc between temperature, a noun, and the verb describing the reaction, e.g. “combine” if q is the statement “molecules A and B combine to molecule C”. This example shows that the operator may connect a noun and a verb. That verbs can be represented by nouns, “combine” replaced by “combining” or “combination”, suggest that, in principle, all CAU-links can be seen as connecting nouns. We will come back to this point after analyzing the chosen Chinese CAU-words, with respect to the chosen example sentences. We will now discuss the chosen CAU-operators and we will give a literal translation next to a more usual translation.

Group 1

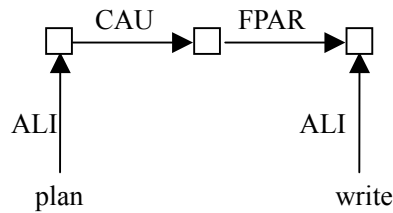
Noun relates to verb

1. 按照, an4zhao4 = puch copy = according to

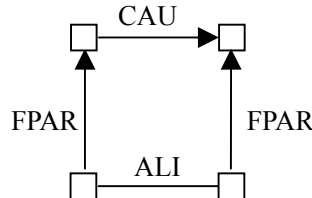
Example

按照计划,我们写文章。 An4zhao4 ji4hua4, wo3men2 xie3 wen2zhang1 = According to the plan, we write the paper.

Analysis: Something, part of “写, xie3”, write, is influenced by “计划, ji4hua4”, plan. As “写, xie3”, write, is a complex frame this something is chosen in a FPAR-relationship to “写, xie3”, write. The plan acts as a condition on the writing of the paper.



The something should show similarity with the plan, be “in accordance with” the plan, which is nicely expressed by “copy”. This leads to using the ALI-link. So we can express both the causality and the similarity accept by



and take this as the word graph for “按照, an4zhao4”, according to.

2. 依据, yi1ju4 = depend reason = in the light of

Example:

依据实际情况我们制定计划。Yi1ju4 shi2ji4 qing2kuang4 wo3men2 zhi4ding4 ji4hua4 = In the light of actual conditions we make the plan.

Analysis: This is an operator that is clearly similar to “按照, an4zhao4”, according to. A synonym is “依照, yi1zhao4” = depend copy. We would choose the same word graph for “依照, yi1zhao4” as for “按照, an4zhao4”, i.e. we consider them to be synonym, although the very choice of a different word suggests that there are different connotations.

3. 根据, gen1ju4 = root reason = due to

Example:

根据天气预报,我们不能出游。Gen1ju4 tian1qi4 yu4bao4, wo3men2 bu4neng2 chu1you2 = Due to the weather forecast, we cannot go out.

Analysis: The process of going out is conditioned by the weather forecast. If the forecasted weather is good then we go, if bad then we cannot go. So good and bad is function as a two-valued variable. A simple CAU-link may therefore be chosen for this operator: □—CAU→□, but notice the discussion of case 8.

4. 凭, ping2 = use, depend on = by use of

Example:

他凭经验办事。Ta1 ping2 jing1yan4 ban4 shi4 = He by use of his experience does this thing.

Analysis: Here the experience supports the doing, the using aspect stands central. Again a simple CAU-link may suffice for representing this CAU-word in a word graph lexicon.

5. 依靠, yi1kao4 = depend, depend = depend on

Example:

我们依靠工资生活。Wo3men2 yi1kao4 gong1zi1 sheng1huo2 = We depending on our wages live.

Analysis: A similar situation as in 4. Wages support the living. Again we may suffice with a simple CAU-link, but dependence may of course have various, often complicated, forms.

For these examples, of operators connecting a noun with a verb, we may restrict ourselves to two different word graphs. Yet it will be clear that some undertones of the words are not captured. For the development of expert systems, this need not have serious consequences.

Group 2

Noun relates to noun

6. 暗含着, an4han2zhe1 = darkness include = imply

Example:

黑暗暗含着危险。Hei1an4 an4han2zhe1 wei1xian3 = Darkness implies danger.

Analysis: There seems to be no direct causal relationship between “黑暗, he1an4”, darkness. and “危险, wei1xian3”, danger. Here we may try to give an extensive expansion of both concepts in order to locate any causal link between parts of the expanded graphs. For the development of expert systems it seems better to delete “暗含着, an4han2zhe1” from the list of CAU-operators. Note, by the way, the use of “暗, an4 = darkness” also in the operator. “Darkness include” is a way of description that does not suggest that the speaker know precisely where the causality is located!

7. 导致, dao3zhi4 = lead to, to form = cause

Example:

虚弱导致疾病。Xu1ruo4 dao3zhi4 ji1bing4 = Weakness causes illness.

Analysis: Seemingly a simple CAU-link, but again there is no direct causal relationship between weakness and illness. What is meant is that a weak body cannot fight the real potential cause of the illness. So, directly speaking, the example sentence is not correct. However, that is not the point. If the operator “导致, dao3zhi4” is used the description of a CAU-link is intended. Note the use of “致, zhi4 = to form”, which describes a pure ORD-link.

8. 来源于, lai2yuan2yu3 = come root from = root in

Example:

知识来源于实践。Zhi1shi lai2yuan2yu3 shi2jian4 = Knowledge root in practice.

Analysis: Here the word “root” expresses a similarity with a growing tree or plant, and is used to express the relationship between “knowledge” and “practice”. Like the earth is necessary for the growing of the tree, practice is necessary to obtain knowledge. As a condition we describe it by a simple CAU-link. However, it is illustrative to investigate this metaphorical use of language somewhat further.

For “tree roots in earth” we may construct the following knowledge graph

Note that “源, yuan2” is the word for “root” and “来…于, lai2…yu3” means “come from”. “Coming” describes a change in location that is basically described by an ORD-link. This ordering is implicit in the CAU-link. In natural language “come from” is sometimes used instead of “is caused by”. Our analysis led to the conclusion that the causal process is not directly from “源, yuan2”, root, toward “知识, zhi1shi1”, knowledge, but from something that belongs to the frame of “实践, shi2jian4”, practice.

We choose a simple CAU-link, but in principle all different CAU-operators should have different word graphs, that may show considerable complexity, similar to what we find here.

9. 酿成, niang3cheng2 = make wine success = bring on.

Example:

大意酿成重大灾难。Da4yi4 niang4cheng2 zhong4da4 zai1nan4 = Carelessness brings on great calamity.

Analysis: Like in the case of “导致, dao3zhi4”, cause, in case 7, the example sentence is, strictly speaking, not correct. However, “make wine success” = “bring on” is a clear indicator of a causal relationship. “酿, niang3” is a word like “brew” in English. The picturesque use of “酿成, niang4cheng2” for a causal operator is quite typical for Chinese.

10. 起因于, qi3yin1yu3 = up reason from = arises from.

Example:

争斗起因于一些小事。Zheng1dou4 qi3yin1yu3 yi4xie1 xiao3shi4 = The fight arises from a mere trifle.

Analysis: Here “小事, xiao3shi4”, trifle, is the source of the “争斗, zheng1dou4”, fight. Again we encounter metaphorical use of language. “因, yin1” = reason explicitly refers to the fact that mental processes of people are involved.

11. 取决于, qu3jue2yu3 = take decision from = decided by.

Example:

成功取决于信念。Cheng2gong1 qu3jue2yu3 xin4nian4 = The success is decided by the belief.

Analysis: The CAU-operator describes a condition. In English the word “by” stands central, in Chinese the decision-making aspect is stressed.

12. 所致, suo3zhi4 = place to form = result of.

Example:

事故由疏忽所致。Shi4gu4 you2 shu1hu suo3zhi4 = The accident was the result of negligence.

Analysis: Here we meet a case where the focus is clearly indicated. It is the result and not the cause that is focused up. Consequently we may choose a word graph in which the second token of the CAU-link is indicated with a focus sign, say a cross, e.g. a simple CAU-link: □—CAU→⊗.

13. 招致, zhao1zhi4 = ask to form = incurs.

Example:

他的过失招致重大损失。 Ta1 de guo4shi zhao1zhi4 zhong4da4 sun3shi1 = His fault incurs a heavy loss.

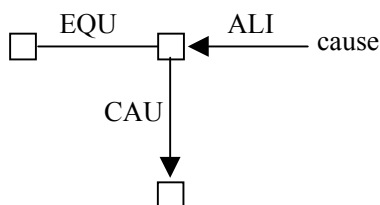
Analysis: There is no reason to represent this operator by more than a simple CAU-link, but in English we also say “a fault makes for a loss” and the explicit graph for “ask” is quite complex.

14. 在于, zai4yu2 = at from = lies in.

Example:

事物发展的根本原因在于其内部矛盾。 Shi4wu4 fa1zhan3 de gen1ben3 yuan2yin1 zai4yu2 qi2 nei4bu4 mao2dun4 = The fundamental cause of the development of a thing lies in its internal contradictions.

Analysis: The problem here is that in fact “cause lies in” is the operator. It is explicitly indicated what the cause is, in Chinese “在, zai4” = at is used, so we may choose:



15. 产生, chan3sheng1 = product born = produces.

Example:

摩擦产生热。 Mo2cha1 chan3sheng1 re4 = Friction produces heat.

Analysis: This operator expresses the process of production of heat by friction, so the basic CAU-link is described. The Chinese sentence graph would be:

mo2cha1 —ALI→□—CAU→□←ALI— re4

We will now discuss several ways this graph may be brought under words.

16. 产生于, chan3sheng2yu2 = produce born from = comes from.

Example:

热产生于摩擦。 Re4 chan3sheng1yu3 mo2cha1 = Heat comes from friction.

Analysis: The only different with case 15 is that the cause is focused upon. So, whereas it is a simple CAU-link, we should be aware of the fact that we may give $\boxed{\text{X}}\text{—CAU}\rightarrow\square$ as the word graph of this CAU-operator.

17. 原因是, yuan2yin1shi4 = basic reason is = reason of ... is.

Example:

热的原因是摩擦。 Re4 de yuan2yin1shi4 mo2cha1 = The reason of heating is friction.

Analysis: The sentence graph is of course the same, but the explicit mentioning of the reason lays a focus on the first token of the CAU-link. In Chinese this is stressed by the word “原, yuan2” = basic. An alternative operator is “理由是, li3you2shi4” = principle reason is.

18. 结果是, jie2guo3shi4 = gather fruit is = reason of ... is.

Example:

摩擦的结果是热。 Mo2cha1 de jie2guo3shi4 re4 = The result of friction is heating.

Analysis: Here the second token is focused upon: □—CAU→果. In Chinese this is expressed by “果, guo3” = fruit.

19. 起因是, qi3yin1shi4 = up reason is = cause of ... is.

Example:

热的起因是摩擦。 Re4 de qi3yin1shi4 mo2cha1 = The cause of heating is friction.

Analysis: This is more or less a synonym of “原因是, yuan2yin1shi4”.

The example is given here are rather frequently occurring and have the following equivalent formulations:

是...的原因, shi4 ... de yuan2yin1 = is ... of reason

是...的理由, shi4 ... de li3you2 = is ... of reason

是...的结果, shi4 ... de jie2guo3 = is ... of result

是...的起因, shi4 ... de qi3yin1 = is ... of cause

20. 致使, zhi4shi3 = to form make = cause

Example:

吸烟致使他死亡。 Xi1yan1 zhi4shi3 ta1 si3wang2 = Smoking caused his death.

Analysis: This is a rather pure CAU-operator. Again it is remarkable that “致, zhi4” = to form is used. A translation might be “death after smoking”.

We see that there are several metaphorically used CAU-operators. The examples concerning friction and heating, as we said, will be encountered most often in technical texts. The problem of determining the arguments is relatively easily solved in such cases. Metaphorical use already poses greater problems, as did the noun to verb CAU-operators in group 1. In the next group it may even be doubted whether there is always a CAU-link that can be explicitly determined.

Group 3

Statement to statement

This group involves, amongst others, the logic words, the word groups of which were studied by Hoede and Zhang [Hoede & Zhang, 2001], and that we discussed in Section 1, with respect to their interpretation as CAU-operators.

21. 从而, cong2er3 = from and = thus.

Example:

A 正确从而 B 正确。 A zheng4que4 cong2er3 B zheng4que4 = A is right, thus B is right.

Analysis: A statement like this is purely logical proposition and asks for a representation of “ $p \rightarrow q$ ” as basic representation, which was given in Figure 2.

Note the use of “而, er3” = and in Chinese.

In words “ $p \rightarrow q$ ” is usually expressed by “if p then q ”, whereas $p \leftarrow q$ is usually expressed by “ p only if q ”. In Chinese we have

22. 除非, chu2fei1 = divide not = only if.

Example:

除非我们有钱,我们才能旅游。 Chu2fei1 wo3men2 you3qian2, wo3men2 cai2neng2 lv3you2 = Only if we have money, we can travel.

23. 如果, ru2guo3 = example fruit = if.

Example:

如果天气晴朗, 我们可以出游。 Ru2guo3 tian1qi4 qing2lang3, wo3men2 ke2yi3 chu1you2 = If the weather is well, we can go out.

Analysis: Here no direct causality is expressed. The state of the weather does not influence the “going out”, but the decision to go out. This decision is not mentioned. *Ellipsis*, meaning something that is not mentioned, occurs rather often in language. It forms one of the major reasons why it is sometimes difficult to determine the arguments of a CAU-operator.

Now we ask in what way “从而, cong2er3” = thus differs from these two logic words. The essential difference lies in the determination of the truth value. For example, “A is right, thus B is right” includes the truth value of A and B. In Chinese the word “从, cong2” = from explicitly refers to an ordering involved. But in e.g. “we can travel only if we have money” or e.g. “if the weather is well, we can go out”, only the propositions are expressed. Nothing is stated about actual truth values.

24. 既然, ji4ran2 = result = since.

Example:

既然这种方法行不通, 我们换另一种方法。 Ji4ran2 zhe4zhong3 fang1fa xing2bu4tong1, wo3men2 huan4 ling4yi4zhong3 fang1fa = Since this method does not work, we try another one.

Analysis: Here too, like for “从而, cong2er3”, thus, in case 21, the truth value of the statements is fixed and in a sense, the rightness of one statement causes, result in, the rightness of the other statement. A representation by a CAU-link is defensible, especially when sentences like “the friction occurs, thus the heat increases” are met and changes or occurrences are mentioned explicitly. But then two nouns are explicitly mentioned. “除非, chu2fei1” = only if and “如果, ru2guo3” = if do NOT refer to causal relationships directly. As the example sentences show the

statement need not be elements that can be taken up in, say, a technical expert system.

25. 所以, suo3yi3 = place to = therefore.

因此, yin1ci3 = reason this = therefore.

Example:

他学习努力, 因此他取得很大进步。Ta1 xue2xi2 nu3li4, yin1ci3 ta1 qu3de2 hen3da4 jin4bu4 = He studied very hard, therefore he made great progress.

Analysis: Here no condition is expressed. “Therefore” refers directly to a cause, in the example sentence it is “study”. In Chinese “所, suo3” = place, respectively “因, yin1” = reason, pinpoint one of the two tokens. A CAU-link description is justified here. Determining the precise arguments is not trivial. The statements describe processes, which themselves may be seen as the arguments. There is here the possibility to use nouns like “study” and “progress” as arguments, as there is always the possibility to substantiate verbs, expressing them by nouns.

The more advanced analysis would expand the concept “study” and “progress”. If “study” means “increasing knowledge” and “progress” means “increasing something” we shed some new light on the interpretation of the example sentence. Instead of “therefore” we then would possibly have used, “i.e.”, id est, to express the fact that “if knowledge increases, then something increases”. This would bring the sentence back into a logical setting.

26. 因为, yin1wei2 = reason for = because.

Example:

因为我是学生,我必须学习。 Yin1wei2 wo3 shi4 xue2sheng1, wo3 bi4xu1 xue2xi2 = Because I am a student, I have to study.

Analysis: Here too expansion of “student”, somebody who studies, sheds some new light on the seemingly causal relationship. Being a student implies studying, by definition. Of course, it is not this literal interpretation that is expressed. Just as in the foregoing example for “therefore”, there is an undertone in the statement that justifies the use of the word “because”. Calling oneself a student, meaning somebody who studies, “leads to” the *obligation* to study indeed. It is only on this level of interpretation that the causality becomes clear.

In the discussion we will come back to this special group of CAU-operators. The most important remark to make is that causal relationships CAN, but must not necessarily, be present in statements in which these CAU-words are used.

Group 4

Noun to statement

27. 鉴于, jian4yu2 = ability from = in view of.

Example:

鉴于这些事实,你可以离开。 Jian4yu2 zhe4xie shi4shi2, ni3 ke2yi3 li2kai1 = In view of these facts, you may go away.

Analysis: Here the allowance to go is what is influenced by the fact, which can be represented by a simple CAU-link.

28. 凭借, ping2jie4 = use borrow = rely on.

Example:

他凭借自己的力量取得进步。Ta1 ping2jie4 zi4ji3 de li4liang4 qu3de2 jin4bu4 = He relied on his own strength to made progress.

Analysis: As we saw before, “progress” is “increase of something” and that something is what is influenced by the “strength”, so again a simple CAU-link, in principle again between two nouns.

29. 使得, shi3de2 = make = make.

Example:

大雪使得交通瘫痪。Da4xue3 shi3de2 jiao1tong1 tan1huan4 = The strong snow made the traffic be out of order.

Analysis: The CAU-link from “snow” is directly oriented towards “traffic” here.

30. 因为, yin1wei2 = reason for = because of.

Example:

因为天黑,他打开了灯。Yin1wei2 tian1 hei1, ta1 da3kai1 le deng1 = Because of the darkness, he turned on the light.

Analysis: In this second example for “因为, yin1wei2” the *decision* to turn on the light is influenced by the darkness. So the real effect, of the darkness in this case, is again not mentioned. “因, yin1” = reason explicitly refers to mental processes involved.

31. 由于, you2yu2 = root from = due to.

Example:

由于大雾,航班被取消。You2yu2 da4 wu4, hang2ban1 bei4 qu3xiao1 = Due to the heavy fog, the flight was cancelled.

Analysis: Here too the heavy fog does not influence the flight directly, but rather the decision to cancel. The metaphorical of “root” has been discussed extensively before, see case 8.

In this group the statements contained nouns that were not mentioned, but were the concepts that were really involved in the CAU-links.

Group 5

Miscellaneous

32. 引起, yin2qi3 = guide up = set off.

Example:

他的话引起哄堂大笑。Ta1 de hua4 yin2qi3 hong1tang2da4xiao4 = His words set off roars of

laughter.

Analysis: The laughing is by people that hear the words and react to these words. The CAU-link is between “words” and “people”.

32. 为了, wei4le = for = in order to.

Example:

为了方便使用计算机, 我们研究自然语言处理。Wei4le fang4bian4 shi3yong4 ji4suan4ji1, wo3men2 yan2jiu1 zi4ran2 yu3yan2 chu4li3 = In order to use the computer easily, we study natural language processing.

Analysis: “In order to” or “for” is not really a CAU-operator, as the basic meaning is that using the computer is the goal that is to be reached by studying. The literal meaning of “为了, wei4le” is “for”, which does not express a causality, but an intention. However, an intention can be seen as a reason.

4 Extracting CAU relationships with their arguments from an example text

The extracting process is aimed at coding scientific text with “causal” relationships. In Section 3, we picked out some Chinese CAU-words, discussed their features, and classified them into some sub-groups. But one might wonder how such a process works out in practice. In order to show how to use these CAU-operators in “causal” knowledge extracting, a part of an article with many “causal” relationships was selected, and Chinese CAU relationships with their arguments in the text were extracted from this given example text.

4.1 The chosen part of the text

We have chosen the following material from the book of Xu [Xu, 1998]. Let us now show the text. We represent each sentence in three lines. The first line is the sentence in Chinese characters, the second one is the sentence in Chinese spelling, and the third line gives the corresponding English words. For each sentence we will underline the CAU-operators only, whether their corresponding arguments are nouns or not. We also indicate its number as occurring in Section 3.

众所周知, 故障 的 发生 与 发展 过程,
Zhong4suo3zhou1zhi1, Gu4zhang4 de falsheng1 yu3 falzhan3 guo4cheng2,
As everyone know, fault de occurrence and development process,

主要 是 由于⁽³¹⁾ 元件 的 内在 因素 所 决定 的。
zhu3yao4 shi4 you2yu2⁽³¹⁾ yuan2jian4 de nei4zai4 yin1su4 suo3 jue4ding4 de.
mainly be due to⁽³¹⁾ component de intrinsic factor suo3 determine de.

但是 元件 工作 的 外部 条件 可以 加速 或
Dan4shi3 yuan2jian4 gong1zuo4 de wai4bu4 tiao2jian4 ke2yi3 jia1su4 huo4
But component working de exterior condition could speed or
延迟 故障 的 发生,
yan2chi2 gu4zhang4 de falsheng1.
delay fault de occurrence.

有时 甚至 会 招致⁽¹³⁾ 元件 的 严重 损坏。
You3shi2 shen4zhi4 hui4 zhao1zhi4⁽¹³⁾ yuan2jian4 de yan4zhong4 sun3huai4.
Sometimes even could incurs⁽¹³⁾ component de serious shatter.

所以, 故障 的 发生 与 发展 的 机制,
Suo3yi3, gu4zhang4 de falsheng1 yu3 falzhan3 de ji1zhi4
Therefore, fault de occurrence and development de mechanism,

应当 是⁽¹⁹⁾ 外部 条件 和 内在 因素 综合
ying1dang1 shi4⁽¹⁹⁾ wai4bu4 tiao2jian4 he nei4zai4 yin1su4 zong1he1
should be⁽¹⁹⁾ exterior condition and intrinsic factor joint
作用 的 结果⁽¹⁹⁾。
zuo4yong4 de jie2guo3⁽¹⁹⁾.
function de result⁽¹⁹⁾.

内在 因素 表现 在 两 个 方面,
Nei4zai4 yin1su4 biao3xian4 zai4 liang3 ge4 fang1mian4,
Intrinsic factor exhibit at two ge4 aspect,

一方面 是 正常 运行 的 自然 过程;
yi4fang1mian4 shi4 zheng4chang2 yun4xiang2 de zi4ran2 guo4cheng1;
one side be normal function de natural process;

另一方面 是 材质、 设计、 制造 或 装配
ling4yi4fang1mian4 shi4 cai2zhi4, she4ji4, zhi4zao4 huo4 zhuang1pei4
the other side be material, design, manufacture or assemble
方面 的 缺陷,
fang1mian4 de que1xian4,
aspect de limitation,

导致⁽⁷⁾ 成为 加快 故障 形成 与 发展 的 因素。
dao3zhi4⁽⁷⁾ cheng2wei2 jia1kuai4 gu4zhang4 xing2cheng2 yu3 falzhan3 de yin1su4.
cause⁽⁷⁾ become expedite fault form and development de factor.

元件 损坏 具有 各种 形式,
Yuan2jian4 sun3huai4 ju4you3 ge4zhong3 xing2shi4,
Component shatter have each form,

如 元 件 长 期 与 介 质 流 作 用 而 导 致⁽⁷⁾ 损 坏,
ru2 yuan2jian4 chang2qi1 yu3 jie4zhi1liu2 zuo4yong4 er3 dao3zhi4⁽⁷⁾ sun3huai4,
as component long time with medium-stream effect and induce⁽⁷⁾ shatter,

又 如 两 个 元 件 表 面 相 互 接 触 而
you4 ru2 liang3 ge4 yuan2jian4 biao3mian4 xiang1hu4 jie1chu4 er3
again as two ge4 component surface each other touch and
造 成⁽⁷⁾ 的 损 坏 等。
zao4cheng2⁽⁷⁾ de sun3huai4 deng3.
make⁽⁷⁾ de shatter etc.

如 果 表 面 没 有 相 对 运 动,
Ru2guo3 biao3mian4 mei2you3 xiang1dui4 yun4dong4,
If surface no relative movement,

一 般 只 能 造 成⁽⁷⁾ 表 面 挤 压 损 伤。
yi1ban1 zhi3neng2 zao4cheng2⁽⁷⁾ biao3mian4 ji3ya1 sun3shang1.
commonly only make⁽⁷⁾ surface extrusion scathe.

当 初 期 接 触 形 式 为 线 接 触 或 点 接 触 时,
Dang1 chu1qi1 jie1chu4 xing2shi4 wei2 xian4 jie1chu4 huo4 dian3 jie1chu4 shi2,
When initial stages touch form be line touch or point touch time,

对 应 的 相 对 运 动 形 式 有 多 种。
dui4ying4 de xiang1dui4 yun4dong4 xing2shi4 you3 duo2 zhong3.
corresponding de relatively movement form have many kinds.

配 合 副 间 作 无 滑 动 的 相 对 滚 动,
Pei4he2fu4 jian1 zuo4 wu2hua2dong4 de xiang1dui4 gun3dong4,
Cooperating component among make slipless de relative roll,

通 常 会 使⁽⁷⁾ 表 面 产 生⁽¹⁵⁾ 疲 劳。
tong1chang2 hui4 shi⁽⁷⁾ biao3mian4 chan3sheng1⁽¹⁵⁾ pi2lao2.
commonly could cause⁽⁷⁾ surface produce⁽¹⁵⁾ fatigue.

表 现 形 式 为 金 属 微 粒 从 接 触 表 面 上 脱 落,
Biao3xian4 xing2shi4 wei2 jin1shu3 wei1li4 cong2 jie1chu4 biao3mian4 shang4 tuo1luo4,
Exhibition form be metal mote from touch surface at fall off,

如 滚 动 轴 承 和 凸 轮 机 构 的 滚 动 就 是 这 样。
ru2 gun3dong4 zhou2cheng2 he2 tu1lun2 ji1gou4 de gun3dong4 jiu4 shi4 zhe4yang4.
as roll axletree and cam framework de roll right on be this matter.

当 材料 硬度 太 低 或 比压 太 高 时,
Dang1 cai2liao4 ying4du4 tai4 di1 huo4 bi3ya1 tai4 gao1 shi2,
When material rigidity much low or relative press much high time,

也 会 出现^(?) 挤压 损伤。
ye3 hui4 chu1xian4^(?) ji3ya1 Sun3shang1.
also could appear^(?) extrusion scathe.

同时 存在 滚动 和 相对 滑动 时,
Tong2shi2 cun2zai4 gun3dong4 he2 xiang1dui4 hua2dong4 shi2,
At the same time occurrence roll and relative glide time,

则 产生⁽¹⁵⁾ 磨损 和 疲劳。
ze2 chan3sheng1⁽¹⁵⁾ mo2sun3 he2 pi2lao2.
so produce⁽¹⁵⁾ abrasion and fatigue.

但 在 许多 情况 下 也 有 可能 产生⁽¹⁵⁾
Dan4 zai4 xu3duo1 qing2kuang4 xia4 ye3 you3 ke3neng2 chan3sheng1⁽¹⁵⁾
But at many instance under also have possible produce⁽¹⁵⁾
表面 挤压 损伤。
biao3mian4 ji3ya1 sun3shang1.
surface extrusion scathe.

疲劳 往往 发生 在 相对 滑动 量
Pi2lao2 wang3wang3 fa1sheng1 zai4 xiang1dui4 hua2dong4 liang4
Fatigue sometimes occurrence at relative glide quantity
最小 或 等于 零 的 区域,
zui4xiao3 huo4 deng3yu2 ling2 de qu1yu4,
least or equal to zero de area,

而 磨损 最 严重 的 部位 则 经常 位于
er3 mo2sun3 zui4 yan2zhong4 de bu4wei4 ze3 jing1chang2 wei4yu3
and abrasion most grievous de part but often locate at
相对 滑动 量 最大 的 区域。
xiang1dui4 hua2dong4 liang4 zui4da4 de qu1yu4.
relative glide quantity maximum de area.

Remarks: The CAU-words, “使, shi3, cause”, “造成, zao4cheng2, make” and “出现, chu1xian4, appear” (marked by (?) in the sentence), do not occur in the collected CAU-word set in Section 3. But they are really Chinese CAU-words. The reasons are the following two. The first reason is that, in Section 3, we picked out *some* Chinese CAU-words, not all. The word “出现, chu1xian4, appear” and “造成, zao4cheng2, make” are ones that were not collected in our set. The second reason is that there are many words have the same meaning, i.e. synonymy. The word “使, shi3, cause” is the synonym with the word “致使, zhi4shi3, cause”. In order to extract CAU relations

from texts, we should make a complete CAU-operator set.

We see that there are twelve CAU-operators in this article. Our goal is, not only to find the CAU-operators, included in the text, but also to find the corresponding arguments they connect, and to check whether these corresponding arguments are nouns.

4.2 The sentences with CAU relationships

We know that to find CAU-operators is not very difficult. But it is very difficult to check their corresponding arguments. First of all, the problem is the segmentation of Chinese words. We do not discuss that here. After we complete the word segmentation, the next step is to keep only the CAU-operators we found and their corresponding arguments. This step is still not easy. We should detect which words are the arguments of the found CAU-operators. After we finish these two steps, we get the following extraction results.

(1) 故障.....由于.....内在 因素
gu4zhang4 you2yu2 nei4zai4 yin1su4
fault due to intrinsic factor

CAU operator: 由于=root from = due to

Argument 1: 内在因素=intrinsic factor

Argument 2: 故障=fault

Representation: intrinsic factor —CAU→ fault

(2) 外部 条件.....招致.....损坏
wai4bu4 tiao2jian4 zhao1zhi4 sun3huai4
exterior condition beget shatter

CAU operator: 招致=provoke arrive = beget

Argument 1: (外部) 条件=(exterior) condition

Argument 2: 损坏=shatter

Representation: exterior condition —CAU→ shatter

(3) 故障.....是 外部 条件 和 内在 因素.....的 结果
gu4zhang4 shi4 wai4bu4 tiao2jian4 he nei4zai4 yin1su4 de jie2guo3
fault be exterior condition and intrinsic factor de result

CAU operator: 是.....的结果=be..... de result = be the result of

Argument 1: 外部条件和内在因素=exterior condition and intrinsic factor

Argument 2: 故障=fault

Representation: exterior condition and intrinsic factor —CAU→ fault

(4) 作用.....导致.....损坏
zuo4yong4 dao3zhi4 Sun3huai4
effect cause shatter

CAU operator: 导致=lead to, to form = cause

Argument 1: no noun argument extracted.

Argument 2: 损坏=shatter

Analysis: Because Argument 1 is a sentence. So, here, no noun argument is extracted.

(5) 接触.....造成.....损坏
jie1chu4 zao4cheng2 sun3huai4
touch make shatter

CAU operator: 造成=create accomplish = make

Argument 1: 接触=touch.

Argument 2: 损坏=shatter

Representation: touch —CAU→ shatter

(6) 没有 相对 运动.....造成.....挤压 损伤
mei2you3 xiang1dui4 yun4dong4 zao4cheng2 ji3ya1 sun3shang1.
no relative movement make extrusion scathe

CAU operator: 造成=create accomplish = make

Argument 1: (没有相对) 运动=(no relative) movement

Argument 2: (挤压) 损伤=(extrusion) scathe

Representation: [NOT: relative movement] —CAU→ extrusion scathe

(7) 无滑动 的 相对 滚动.....使.....产生 疲劳
wu2hua2dong4 de xiang1dui4 gun3dong4 shi3 chan3sheng1 pi2lao2
slipless de relative roll cause produce fatigue

CAU operator: 产生=product born = produce

Argument 1: (无滑动的相对) 滚动=(no slipless relative) roll

Argument 2: 疲劳=fatigue

Representation: [relative roll without slip] —CAU→ fatigue

(8) 表面 产生 疲劳
biao3mian4 chan3sheng1 pi2lao2
surfave produce fatigue

CAU operator: 产生=product born = produce

Argument 1: 表面=surface

Argument 2: 疲劳=fatigue

Representation: no extracted CAU relations.

Analysis: Although two noun arguments are extracted, they do not express “causal” relation here.

(9) 硬度 太低 或 比压 太高.....出现 挤压 损伤
ying4du4 tai4 di1 huo4 bi3ya1 tai4 gao1 chu1xian4 ji3ya1 sun3shang1
rigidity much low or relative press much high appear extrusion scathe

CAU operator: 出现=grpw up become visible = appear

Argument 1: (硬度太) 低或(比压太) 高=

(rigidity much) low or (relative press much) high

Argument 2: (挤压) 损伤=(extrusion) scathe

Representation: [(rigidity much) low] —CAU→ extrusion scathe

[(relative press much) high] —CAU→ extrusion scathe

Analysis: Because the sentence is a parataxis, there exist two CAU relations.

(10) 滚动 和 相对 滑动 产生 磨损 和 疲劳
gun3dong4 he2 xiang1dui4 hua2dong4 chan3sheng1 mo2sun3 he2 pi2lao2
roll and relative glide produce abrasion and fatigue

CAU operator: 产生=product born = produce

Argument 1: 滚动和（相对）滑动=roll and (relative) glide

Argument 2: 磨损和疲劳=abrasion and fatigue

Representation: roll —CAU→ abrasion

roll —CAU→ fatigue

(relative) glide —CAU→ abrasion

(relative) glide —CAU→ fatigue

Analysis: as in (9), there exist four CAU relations.

(11) 产生 挤压 损伤
chan3sheng2 ji3ya1 sun3shang1
produce extrusion scathe

CAU operator: 产生=product born = produce

Argument 1: 滚动和（相对）滑动=roll and (relative) glide

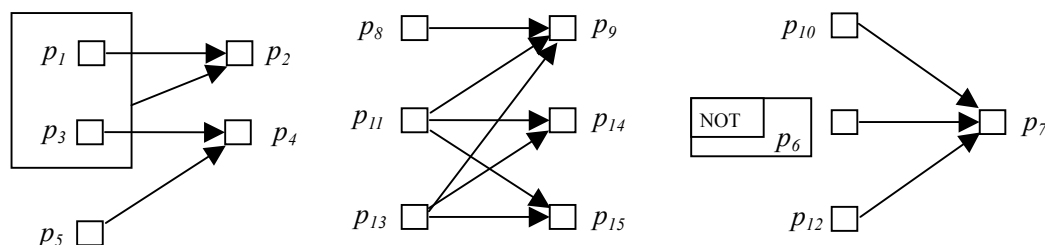
Argument 2: 损伤=scathe

Representation: roll —CAU→ scathe

(relative) glide —CAU→ scathe

Analysis: Because this sentence is not a complete sentence, it is another part of the above sentence. These two sentences have the same SUBJECT. So, the argument 1 is also the same.

After we include these representations in a combined representation, we get the following.



We introduce some symbols here in order to represent the graph clearly. They indicate:

X	$f(X)$	X	$f(X)$
intrinsic factor	p_1	fault	p_2
exterior condition	p_3	shatter	p_4
touch	p_5	relative movement	p_6
(extrusion) scathe	p_7	relative roll without slip	p_8
fatigue	p_9	(rigidity much) low	p_{10}
roll	p_{11}	(relative press much) high	p_{12}
(relative) glide	p_{13}	abrasion	p_{14}
scathe	p_{15}		

From the combined representation we can see that the article consists of three main themes. One is on p_2 and p_4 , i.e. fault and shatter. The second one is on p_7 , i.e. (extrusion) scathe. The third one is on p_9 , p_{14} and p_{15} , i.e., fatigue, abrasion and scathe.

4.3 Aspects of the extraction process

In Section 4.2, we picked out some causal relations, i.e. Chinese CAU-operators and corresponding arguments, from the given example text by hand. How do we pick them out? A very important point is that we use inference to do it. In this section, we will discuss this and get some information from extracting CAU relationships by hand. This is very useful to help us to build an automatic extracting system.

As we discussed above, it is not difficult to choose CAU-operators, because the number of these CAU-words is finite. But it is very difficult to determine the corresponding noun arguments.

In Appendix, we have analyzed the situation that after CAU-operators are obtained, the corresponding noun arguments are difficult to determine. But how do we pick out these CAU relationships (including CAU-operators and their corresponding arguments) by hand in Section 4.2? We know that we determine the CAU relations according to their definition strictly.

In order to determine the arguments corresponding to some CAU-operators, let us remember first what “A CAU B” means. From the definition we know that “A CAU B” means “An occurrence of or a change in A can cause an occurrence of or a change in B.” Now CAU-words or CAU-operators play the role like “cause” in the definition. Actually, CAU-operators are very important in determining CAU relationships, although some sentences with CAU meaning do not contain any CAU-operators. Almost all CAU relations are connected by CAU-operators. What about “an occurrence of or a change in” A or B? Mainly we determine it by the following two methods.

(1) to find change indicators or occurrence indicators

In order to find noun arguments connected by the CAU-operators in a sentence, we would check which part is a noun, and to check whether its change or occurrence can “cause” (expressed by the CAU-operators) the change or occurrence of another noun. So we would have to check *change indicators* or *occurrence indicators*. Analyzing a sentence on change or occurrence indicates we have two possibilities. First, we may actually find that something is changing or occurring. Second, some words, by their very meaning, imply a change or occurrence. This second possibility asks for a semantic analysis of words, i.e., more elaborate word graphs for these concepts. Here we essentially use expansion of concepts. For example,

- 如 元件 长期 与 介质流 作用 而 导致⁽⁷⁾ 损坏,
- Ru2 yuan2jian4 chang2qi1 yu3 jie4zhi1liu2 zuo4yong4 er3 dao3zhi4⁽⁷⁾ sun3huai4,
- As component long time with medium-stream effect and induce⁽⁷⁾ shatter.

In this sentence the CAU-operator is “导致, dao3zhi4”, induce, which is just recognized and which is an operator pointing from the first part of the sentence to the second part, which is “损坏,

sun3huai4”, shatter. This, in fact, already indicates that “shatter” is induced. An expansion of “shatter”, by using a dictionary may indeed confirm a change of state from “whole” to “broken”.

In the first part of the sentence the word “作用, zuo4yong4”, effect, expresses a change or an occurrence. The effect is described by “component long time with medium-stream”. Like in the case of “high blood pressure (may) cause heart attack”, here “long time with medium-stream” describes the occurrence of a *state*. No noun argument is extracted. That is to say, if the effect (between medium-stream and component) occurs, then shatter occurs. Because “shatter” is something changing, the sentence describes that if something (effect) occurs then something else changes.

- 硬度 太低 或 比压 太高.....出现 挤压 损伤
- Ying4du4 tai4 di1 huo4 bi3ya tai4 gao1 chu1xian4 ji3ya1 sun3shang1
- Rigidity much low or relative-pressure much high appear extrusion scathe.

In this example, although “出现, Chu1xian4, appear” has a “causal” meaning more or less, its “causal” meaning is not strong. We did not pick it out and add it to our CAU-words set in Section 3. But, apart from this word, we find an “occurrence indicator” and a “change indicator” in the sentence. “Much low” and “much high” indicate that something changes, they are change indicators. “Extrusion scathe” is still something changing, like “shatter” in our first example, it is a change indicator too. So, to determine “occurrence indicator” or “change indicator” of nouns is an important method to find noun arguments.

(2) to find noun arguments from expansion of word graphs.

Another method to find noun arguments, connected by the CAU-operators in a sentence, is from expansion of word graphs, although there are no sentences in the given text that need expansion of word graphs to find noun arguments. Because such an inference is a very important technique for finding noun arguments by hand, using expansion of word graphs is another method to determine noun arguments. For example, in Chinese there is a proverb:

- 燕子 低 飞 蛇 过 道, 大雨 马上 就 来到。
- Yan4zi di1 fei1 she2 guo4 dao4, da4yu3 ma3shang4 jiu4 lai1dao4.
- Swallow low fly snake pass way, big rain immediately soon come.

In Chinese villages, people even now do weather forecast according to this proverb. Actually, there are many proverbs used in weather forecast. We know that whether it will rain does not depend on swallow or snake. But the foreboding of raining can influence swallows and snakes. In order to obtain this CAU relationship, we should expand word graphs, showing possible causes for their behavior of “flying low” respectively “passing way”. This makes case of considerable background knowledge.

4.4 Concluding remarks

Unlike from the sentences in everyday life, from the sentences in scientific texts or technical

reports with CAU-operators we can, in most cases, easily extract CAU relations. In the chosen article, there are twelve CAU-operators in eleven sentences. From nine sentences of them, we can extract CAU relations, including CAU-operators and their corresponding noun arguments. After we picked out the CAU relations in scientific texts, we obtain the material to establish the knowledge base in expert systems. This paper is only showing the method to extract CAU relations using CAU-operators by means of knowledge graphs. In order to build an automatic extracting system, there are still many things we have to do.

(1) Word segmentation

In order to segment words from unmarked sentences correctly, we think a better method is to build a good Chinese coding system containing much information. At least we should change the current Chinese input and storage system based on Chinese characters to a new one that is based on Chinese words.

(2) Checking parts of speech

After we get words from texts, we still should know the parts of speech. Nouns have to be located as well as change and occurrence indicators.

(3) Choosing noun arguments by CAU-operators connected

When we get the words and parts of speech, the next work is to determine noun arguments connected with the found CAU-operators or even in the case where no such operators occur.

(4) Concept similarity measure and concept identification

For the case of similarity measures and concept identification, we can refer to [de Vries, 1989].

5 Discussion

The Chinese CAU-words considered exhibit several features that are of importance for actual extraction of causal relationships from text for the establishment of expert systems.

The first feature is the difference between logical implication and causal relationship. A sentence like “if $2 \times 2 = 4$ then Beijing is a big city” or “if $2 \times 2 = 5$ then Beijing is a small city” does not convey any causal relationship. Both sentences are true from the logical point of view, but nobody will extract a CAU-link from sentences like these.

In natural language, however, words are not always well chosen. Let us consider the examples on friction and heat(ing). “Friction causes heat” is a wording that would immediately be used to establish a CAU-link. “If friction occurs then heating occurs” is a wording in a completely logical setting. There is a causal relationship behind this statement and that would be missed in the extraction when “if”, “only if” and “if then” would be excluded from the set of CAU-operators.

The use of such, not so well chosen, operators to describe CAU-links forces us to consider the linked propositions “friction occurs” and “heating occurs”. When friction and heating can be seen as both playing a role in a process, this would justify establishing a CAU-link between them on basis of the chosen wording. The conclusion is that statement to statement operators are to be handled with care. If a CAU-link is present it will be recognizable by the occurrence of words like

“occur” or “change” and in general will be a noun to noun link.

The second feature is the use of several operators that express intentions by explicitly mentioning “reason” or by use of a word like “for”. The focus in sentences containing such CAU-words lies on the reasoning process of people. Nevertheless, at the level of arguments of the operator there may be a causal relationship present, describing the way the goal, usually expressed in the second argument, can be achieved. Like for IF-THEN-statements, this causal link between the arguments can be discovered by asking whether there are nouns exhibiting change or occurrence, so that the standard sentence for the recognition of a causal relationship can be formulated with the same meaning.

The third feature is the metaphoric use of language. Usually quite complicated statements are linked by such operators as “来源于, lai2yuan2yu3” = “come root from” or “酿成, niang3cheng2” = “make wine success”. This kind of language use partly from the fact that the Chinese way of making words favors such constructions, a historical aspect probably based on the fact that the original, vary old, words describing every day life things, had to be used to “bring under words” relationships like we consider. Partly metaphorical use also stems from the fact that the speaker cannot or does not want to make a more precise description and trusts that the listener can infer the meaning from the vaguer metaphorical description by “seeing” the analogy.

For our extraction process it may again help to localize two nouns in the arguments that allow a standard sentence formulation with (about) the same meaning.

In all these three cases the recipe for extraction of CAU-links is the same. *Locate nouns in the arguments that change or occur and such that the meaning of the sentence is more or less maintained.*

The second part of this prescription may be omitted but then one is more prone to mistake. A more systematic, but much more intensive way of discovering a causal relationship involved is to *expand* the sentence. This means that the complete sentence graph has to be constructed and then the words in it are to be expanded, which leads to an even larger graph. This expansion procedure is to be continued until the hidden, but present, CAU-link comes forward.

At the time of writing this expansion process, which asks for a large word graph lexicon, cannot be carried out automatically yet. As a first application of our results, to establish an expert system from a text which is a graph with only CAU-links, the above mentioned recipe can be used.

References

- [Ce, 1988] Editorial Division, *A modern Chinese-English dictionary*, Foreign Language Teaching and Research Press, 1988.
- [de Vries, 1989] P. H. de Vries, *Representation of Scientific Texts in Knowledge Graphs*, Ph.D. Thesis, Rijksuniversiteit Groningen, 1989.
- [Hoede & Li, 1996] C. Hoede and X. Li, *Knowledge graphs: the first set*, in Proceedings of the 4th International Conference on Conceptual Structures, ICCS'96, August 1996, Australia.
- [Hoede & Liu, 1998] C. Hoede and X. Liu, *Knowledge graphs: the second set*, in Proceedings of

the 6th International Conference on Conceptual Structures, ICCS'98, August 1998, France.

[Hoede & Zhang, 2001] C. Hoede and L. Zhang, *Word Graphs: The Third Set*, in: *Conceptual Structures: Broadening the Base* (H. S. Delugach and G. Stumme, eds.), Proceedings of the 9th International Conference on Conceptual Structures, Stanford University, California, U. S. A. Springer Lecture Notes in Artificial Intelligence, Vol. 2120, ISBN 3-540-42344-3, (2001) 15-28.

[Huang, 1988] K. Huang, *Introduction to Expert Systems*, Southeastern University Press, 1988.

[James, 1991] P. James, *Knowledge graphs*, Memorandum no. 945, University of Twente, Enschede, The Netherlands, 1991.

[Waterman, 1986] D. A. Waterman, *A Guide to Expert Systems*, Reading, Mass., Addison-wesley, 1986.

[Willims, 1993] M. Willims, *Chemistry of language: a graph-theoretical study of linguistic semantics*, Ph.D. thesis, University of Twente, Enschede, The Netherlands, 1993.

Appendix

We go through the example sentences focusing on the arguments and try out the recipe of finding nouns in the arguments. The CAU-words are described by the literal Chinese translation. The primary argument words are underlined.

1. CAU-word : 按照, an4zhao4 = puch copy
Argument 1 : the plan
Argument 2 : we write papers
Analysis : Something should change, suggesting the papers as the writing influences the state of the paper. The changing aspect of papers is according to the plan. As a causational like to a part implies a causational link to the whole we chose papers as the second noun.
Result : plan —ALI→□—CAU→□←ALI— papers
2. CAU-word : 依据, yi1ju4 = depend reason
Argument 1 : actual conditions
Argument 2 : we make the plan
Analysis : The word “reason” indicates that the mental process is focused upon. The making of the plan is triggered by the conditions. However, other conditions would leads to a different plan, so on the background, as it were, we have
Result : conditions —ALI→□—CAU→□←ALI— plan
3. CAU-word : 根据, gen1ju4 = root reason
Argument 1 : the weather forecast
Argument 2 : we cannot go out
Analysis : As going out does not describe anything else but an act, no noun is discernable here, or it should be “going”. This is, however, not causally related to the weather forecast.
Result : No CAU-link extracted.
4. CAU-word : 凭, ping2 = use, depend on
Argument 1 : his experience
Argument 2 : he does this thing
Analysis : Like in case 1, we might link “experience” and “thing”, but although the first clearly influences the second, by the doing, there is no explicit relationship that can be indicated.
Result : No CAU-link extracted.
5. CAU-word : 依靠, yi1kao4 = depend-depend
Argument 1 : our wages
Argument 2 : we live

- Analysis : The influence of wage on living is clearly that of wage on some aspect of living, mainly “possessions”. This is not explicitly mentioned and “live” is such a complex concept that no specific aspect presents itself in the expansion of the concept “live”.
- Result : No CAU-link extracted.
6. CAU-word : 暗含着, an4han2zhe1 = darkness include
- Argument 1 : darkness
- Argument 2 : danger
- Analysis : Although two nouns are explicitly mentioned, “dark include” is an operator that might be paraphrased as “may include”. There is no real causal relationship between the two concepts “darkness” and “danger”.
- Result : No CAU-link extracted.
7. CAU-word : 导致, dao3zhi4 = lead to-to form
- Argument 1 : weakness
- Argument 2 : illness
- Analysis : This is a clear CAU-operator. The discussion of the example sentence showed only an indirect relationship between the arguments. Weakness as a condition for the occurrence of the illness, that is caused by something else.
- Result : No CAU-link extracted, unless indirect causal relationships are accepted.
8. CAU-word : 来源于, lai2yuan2yu3 = come root from
- Argument 1 : knowledge
- Argument 2 : practice
- Analysis : The discussion of the example sentence showed that metaphorical use of language tends to hide actual causal relationships.
- Result : No CAU-link extracted.
9. CAU-word : 酿成, niang3cheng2 = make wine success
- Argument 1 : carelessness
- Argument 2 : calamity
- Analysis : Again the “real” causal relationship is hidden, due to the metaphorical use of language.
- Result : No CAU-link extracted.
10. CAU-word : 起因于, qi3yin1yu3 = up reason from
- Argument 1 : trifle
- Argument 2 : fight
- Analysis : The causality is via the mental process of people.
- Result : No CAU-link extracted.

11. CAU-word : 取决于, qu3jue2yu3 = take decision from
Argument 1 : belief
Argument 2 : success
Analysis : The causality is via the mental process of people.
Result : No CAU-link extracted.
12. CAU-word : 所致, suo3zhi4 = place to form
Argument 1 : negligence
Argument 2 : accident
Analysis : The accident was only indirectly caused by the negligence, which acts as a condition.
Result : No CAU-link extracted.
13. CAU-word : 招致, zhao1zhi4 = ask to form
Argument 1 : fault
Argument 2 : loss
Analysis : Not all faults incur losses. Between the two concepts “fault” and “loss” there is no direct causal relationship.
Result : No CAU-link extracted.
14. CAU-word : 在于, zai4yu2 = at from
Argument 1 : contradictions
Argument 2 : development
Analysis : The two arguments may be seen as causally related. Here we see very clearly what is coming forward from several examples so far. There is a causal relationship present and the CAU-operator is used in a correct way. Yet it is rather difficult to discover the precise arguments of the causality. The actual relationship between “contradictions” and “development” remains vague, even after expanding the concept “development”. Which statements or aspects in the description of development contradict each other? If we extract a CAU-link between the two noun concepts, we have no precise picture of the causality, which we feel may indeed be present.
Result : contradictions —ALI→□—CAU→□←ALI— development

So far the example sentences did not give many CAU-links that we would accept for an expert system. We have determined pairs of nouns and we could have taken these pairs as being causally related. But then, at the same time, we are aware of the fact that although the CAU-words used are rightfully called CAU-words, the arguments are not directly linked in most cases. The basic causal relationship is often not explicitly mentioned. It is only by thinking carefully about what is said, and that means expansion of concepts, that we find why a CAU-word is used in a certain sentence.

For the representation this means the following. The CAU-words may be represented by word graphs and in many cases just by a simple CAU-link, although it is evident that each CAU-

operator has its specific connotations and should have its specific word graph. Metaphorical use of language makes it extremely difficult to give these word graphs. A typical example is “酿成, niang3cheng2” = make wine success.

We are facing a decision on the representation. Either we give explicit different word graphs for all CAU-words, which may help in the process of determining the arguments, or we accept that a CAU-link usually is representing a more complex graph. We choose for the second option, because even if the word graphs in expanded form are known the use of CAU-words is such that still are arguments of the basic causality expressed are often not directly detectable. As for these argument we have seen that pairs of nouns can usually be determined, although there need not be a direct causal relationship between them. However, just as the CAU-operators themselves are complex in most cases and yet are represented by simple CAU-links, the arguments may only be indirectly related causally and yet be represented as the arguments of the simple CAU-link. The quality of such a result of extraction is, of course, rather poor.

For extracting technical knowledge from texts the described simplification may be acceptable. Even if the CAU-words are quite complex in nature their use in a technical setting may concern arguments that are quite simply identifiable. But usually the CAU-words used are restricted to a small set. The cases 15, 16, 17, 18, 19 and their equivalent formulations are all describing the technical causal relationship between “friction” and “heat”, arguments that are easily extracted. Result: friction $\text{—ALI}\rightarrow\Box\text{—CAU}\rightarrow\Box\leftarrow\text{ALI—}$ heat in all these cases.

20. CAU-word : 致使, zhi4shi3 = to form make
 Argument 1 : smoking
 Argument 2 : death
 Analysis : This is clearly a CAU-operator. The causal relationship is complex. “Smoking” is a process, described by a substantiation of the verb “to smoke”. “Death” is a fact a process too. In particular in examples like this the actual causality is hard to pinpoint. Yet in the establishment of medical expert systems we would like to extract information like this.
 Result : smoking $\text{—ALI}\rightarrow\Box\text{—CAU}\rightarrow\Box\leftarrow\text{ALI—}$ death
21. CAU-word : 从而, cong2er3 = from and
 Argument 1 : A
 Argument 2 : B
 Analysis : As A and B are arbitrary statements no causal relationship is expressed.
 Result : No CAU-link extracted.
22. CAU-word : 除非, chu2fei1 = divide not
 Argument 1 : money
 Argument 2 : travel
 Analysis : Two statements are concerned in which two nouns can be determined. But clearly this example is expressing a logic

- proposition like in case 21.
- Result : No CAU-link extracted.
23. CAU-word : 如果, ru2guo3 = example fruit
- Argument 1 : weather
- Argument 2 : go out
- Analysis : A purely logical sentence, and on the background only a relationship between “weather” and the “decision” to go out.
- Result : No CAU-link extracted.
24. CAU-word : 既然, ji4ran2 = result
- Argument 1 : method
- Argument 2 : another one
- Analysis : Even through two methods are put into relationship, there is no reason to extract a CAU-relationship.
- Result : No CAU-link extracted.
25. CAU-word : 因此, yin1ci3 = reason this
- Argument 1 : study
- Argument 2 : progress
- Analysis : As we discussed, in fact a tautological statement is involved. Study, increasing knowledge means increasing something, “therefore” progress.
- Result : No CAU-link extracted.
26. CAU-word : 因为, yin1wei2 = reason for
- Argument 1 : student
- Argument 2 : study
- Analysis : Analogous to case 25.
- Result : No CAU-link extracted.
27. CAU-word : 鉴于, jian4yu2 = ability from
- Argument 1 : facts
- Argument 2 : may go away
- Analysis : Instead of a “decision”, here an “allowance” is the second noun involved.
- Result : No CAU-link extracted.
28. CAU-word : 凭借, ping2jie4 = use borrow
- Argument 1 : strength
- Argument 2 : progress
- Analysis : The second noun involved is not “progress”, which is “increase of

- something”, but that “something” that is increasing.
- Result : strength —ALI→ □ —CAU→ □ ←ALI— where the second token is not typed.
29. CAU-word : 使得, shi3de2 = make
- Argument 1 : snow
- Argument 2 : traffic
- Analysis : A noun normal case of causal relationship
- Result : snow —ALI→ □ —CAU→ □ ←ALI— traffic
30. CAU-word : 因为, yin1wei2 = reason for
- Argument 1 : darkness
- Argument 2 : light
- Analysis : As can be seen from the Chinese use of the word “reason”, there is no direct causal relationship.
- Result : No CAU-link extracted.
31. CAU-word : 由于, you2yu2 = root from
- Argument 1 : fog
- Argument 2 : flight
- Analysis : Here the metaphorical use of language hints at the non-existence of a direct causal relationship.
- Result : No CAU-link extracted.
32. CAU-word : 引起, yin2qi3 = guide up
- Argument 1 : words
- Argument 2 : laughter
- Analysis : The listening people are intermediate.
- Result : No CAU-link extracted.
33. CAU-word : 为了, wei4le = for
- Argument 1 : study
- Argument 2 : use
- Analysis : The operator expresses an intention, not a causality.
- Result : No CAU-link extracted.

General conclusion

Almost all operators considered are CAU-words, or may lead to extraction of a CAU-link. However, the example sentence show that in everyday non-technical conversation or writing language is used in a way that asks for considerable interpretational power of the listener or reader. In more than half of the case we concluded that, strictly taken, no CAU-link could be extracted,

although the presence of a CAU-link in an expansion of the sentence graph was not to be excluded. Interesting in this connection are the CAU-operators that contain the word “reason”. A reason is in a sense a cause, but the effect is on the decision taken by people. The word “decision” is usually not mentioned in the sentence, as it is already implied by the word “reason”.

It turns out that even if the word graphs for the considered CAU-words would be extensively given, there still would exist a considerable problem with the arguments. Often the background knowledge needed for proper interpretation of a sentence is considerable. For automatic extraction of CAU-links this poses a tremendous problem. However, it may be that technical texts are structured in such a way that the needed use of background knowledge is much less. As mentioned before, an example text is studied and will be reported upon in a forthcoming paper.