Chinese Generation from Interlingua

Abstract

This report presents the generation of Chinese text from interlingua representations used in the KANT knowledge based machine translation system. Chinese is very different from the European languages, and its flexible sentence structures give us more freedom to do the translation, but it also presents some difficulties in choosing the suitable translation. The strategy of selecting the suitable translation is based on the cost and the scope of the syntactic structures of our system. The first prototype system in Chinese generation in CMT was developed by Dr. Li, and it deals with Caterpillar domain documentation. Our system is developed in the CNBC domain, and the syntactic scopes of this two systems are very different. We propose and implemented a new strategy in our system, and we believe that this system is scalable and could cover larger syntactic structures.

1 Introduction

In knowledge based machine translation system, how to use the knowledge to generate target language, and how to do good translation are the main issues we are concerning about. In the KANT knowledge based machine translation system, we have GLR parser to get the Interlingua representation which is independent of any language, and also we have Genkit to generate the target language from the Interlingua. These tools give us the opportunity to concentrate on the knowledge gathering and representation of the system, and use the knowledge to get very good translation output. The way we do knowledge gathering and representation in target language generation system is :

(1) build lexicon to represent the lexicon mapping rules from Interlingua to the target language

(2) build grammar rules to represent the syntactic structures of the target language.

Our Interlingua is a English based semantic representation, which contains the concept structure of the language.

In the framework of a prototype system which was developed by Dr. Li Tangqiu, we continued the research project on Sentence Generation from interlingua to Chinese. In this paper, we are going to introduce the basic Chinese language phenomena which is related to our research, and then briefly illustrate the system architecture of Dr. Li's prototype system, his approach of doing Chinese generation from Interlingua, and finally we will present our approach to do whole sentence Chinese generation.

2. Chinese phrase and sentence structures

Chinese is very different from the English and other European languages. Besides the obvious difference in the number of elements of character set and their appearance, there are many differences in syntax, and in the relationship between syntax and semantics. First, Chinese is characterized as an isolating language, and it has no inflections which function as number markers on nouns, such as the plural morpheme "+s" in English; it has no inflection of verbs to signal difference in number, person, tense or aspect, such as the English forms "give, gives, given, giving" for the verb "give". Also there is no number and gender agreement between subject and verb. Chinese use open class words, and with the help of word order cues and some closed-class functional words.

2.1 Chinese sentence structure

It is not easy to characterize the word order in Chinese. It is hard to say it is a SVO, SOV or VSO language. Usually, different word order will convey different meaning of the sentences (Li, Nyberg, Carbonell, 1996). But Chinese do have some regular word order which can represent the normal meaning of the sentence. For example, SVO is the regular sentence structure which can reasonably represent the sentence meaning.

1. I am Terry Keenan.
   我是Terry Keenan.

2. Unocal will consider a stock repurchase plan.
   UNOCAL将考虑一个股票重购计划 .

In most of the situations, SVO can almost convey the complete meaning of the sentences. But sometimes in order to express emphasization, we use different word orders. For example, passive sentence structures are as follows:

<NP> 被 <NP> <verb>
<NP> 被 <verb>

2.2 Chinese phrase structures

For the noun phrase or verb phrase, although we do have these structures:

<verb> <verb modifier>
<noun> <noun modifier>

most of the case, the normal structures:

<verb modifier> <verb>
<noun modifier> <noun>

can almost represent the real meaning. This means most of the case we can use some very regular grammar rules to do the generation.

We are going to talk about this phenomena in detail in the following sessions.

2.3 Tense

We use open words to express the tense. For example:

(1) 正在 present
(2) 将要 future
(3) 已经 perfect
(4) 了 past

In the first three classes, we put the open words in front of the verb; for the last case, we put the open word after the verb. In most of the case, we can use this regular order to represent the sentence meaning.

Since we have no morphology translation in Chinese, and the knowledge of the language can almost be represented by the lexicon and the grammar rules, our system can do one step generation from Interlingua to Chinese.

3. The Chinese sentence generation in Catapillar domain

3.1 Concept based generation

The KANT interlingua is based on the notion of concept frames. Each concept frame represents a given unit of meaning along with its specific properties and/or its relationships with the other concepts reflected in the utterance. The basic concept types are objects, events and properties, which are the basic elements of our interlingua and typically represent nouns, verbs and adjectives. (Li, Nyber, Carbonell, 1996)

The task of sentence generation is to map the interlingua semantic representation to a target language sentence or phrase according to the concept type of the interlingua. Traditionally, this is done in three phases in KANT: lexical selection, f-structure creation, and syntactic generation.

In lexicon selection, the most appropriate lexical item or items are selected for each frame in the interlingua; Then the interlingua and the set of appropriate candidate lexemes are analyzed to determine and produce a syntactic functional structure (f-structure) for the target utterance. Finally, the syntactic generation phase produces a properly inflected and ordered output sentence according to the target language generation grammar.

Since Chinese sentence is generated mostly according to the semantic meaning of each concept, syntactic elements sometimes do not convey the enough information to do generation. In Dr. Li's prototype system, he proposes to use two step generation: lexicon selection and sentence generation from interlingua. This will eliminate the semantic meaning loss in the f-structure creation step, and make this approach more suitable for Chinese generation.

3.2 Accurate generation

Since the sentence structure in Caterpillar domain is more technical oriented, and they use some limited syntactic structures in writing, the sentence structures are more regular. In order to do accurate translation for technical document, Dr. Li's approach is: writing some very specific grammar rules in order to achieve very accurate translation. For example:

```
(<simp-s> --> (<dec-act-s>)
        (((x0 root) =c "适用")
         ((x0 passive) =c +)
         ((x0 mood) =c dec)
         ((x1 cat) = v)
         ((x1 root) <= "是")
         (*EOR* (((x1 modal) == (x0 modal))) (((x0 modal) = *undefined*)))
         (*EOR* (((x1 tense) == (x0 tense))) (((x0 tense) = *undefined*)))
         (*EOR* (((x1 mood) == (x0 mood))) (((x0 mood) = *undefined*)))
         (*EOR* (((x1 tentative) == (x0 tentative))) (((x0 tentative) = *undefined*))
)
         (*EOR* (((x1 obligation) == (x0 obligation))) (((x0 obligation) = *undefined
*)))
         (*EOR* (((x1 compulsion) == (x0 compulsion))) (((x0 compulsion) = *undefined
*)))
         (*EOR* (((x1 negation) == (x0 negation))) (((x0 negation) = *undefined*)))
         (*EOR* (((x0 purpose) = *defined*))
                (((x0 location) = *defined*)))
         (*EOR* (((x0 theme) = *defined*)((x1 theme) == (x0 theme)))
                 (((x0 patient) = *defined*)((x1 theme) == (x0 patient))))
         ((x1 predicated-of-theme) <= (x0))))
```

In the sentence grammar rules, this rule is specifically used to deal with this kind of sentence:

1 diesel fuel is best suited for cold weather operation.
1#柴油燃料是最适用于冷天气的操作的.

Also, there are other same kind of verb or noun which need special attention in order to translate it accurately. Surely we can use some categorization approach to classify different groups of verb or noun according to their properties, but this still need a lot of efforts, and the most difficult problem is we don't know for sure when we are done. How many categories, and what kind of categorization are enought? For example, in this data set, some categorization according to some standards might be enough, but if we move to another data set, the categorization might be irrelevant or doesn't make too much sense. Because Caterpillar domain deals with heavy machinery documentation, its sentence structures are limited to some extent, I think this approach is suitable, and it can give us accurate translation if we can provide enough grammar rules to deal with different situations. But for some other domains, this could be very costly, and there is some trade-ofs between accuracy and cost. In the next section, I am going to discuss this issue in CNBC domain Chinese generation.

4. Special issues in Sentence generation in CNBC domain

4.1 Our philosophy

After doing sometime of research in Dr. Li’s prototype system, we started to working on the sentence generation in CNBC domain. We looked at the interlingua representation in this domain, and found there are some phenomena which is different from the old system, and the interlingua representation differs a lot. In this domain, the sentence structures are more free, and with more prepositional phrases, their attachments with the head (noun phrase or verb phrase) are more variable. For example:

(1) LET’S RUN DOWN THE FINAL NUMBERS NOW FROM WALL STREET.

```
(*A-RUN-DOWN (PUNCTUATION PERIOD)
             (FORM IMPERATIVE)
             (TENSE PRESENT)
             (MOOD IMPERATIVE)
             (ARGUMENT-CLASS AGENT+THEME)
             (AGENT (*PRON-WE (REFERENCE NO-REFERENCE)
                              (PERSON FIRST)))
             (THEME
               (*O-NUMBER (NUMBER PLURAL)
                          (REFERENCE DEFINITE)
                          (UNIT -)
                          (PERSON THIRD)
                          (ATTRIBUTE (*P-FINAL (DEGREE POSITIVE)))))
             (Q-MODIFIER
               (*K-FROM
                 (OBJECT
                   (*PROP-WALL-STREET (NUMBER MASS)
                                      (IMPLIED-REFERENCE +)
                                      (PERSON THIRD)))))
             (MANNER (*M-NOW (DEGREE POSITIVE))))
```

让我们现在从华尔街过一遍最后的数据 .

(2) TOSCO SEALED A DEAL TO BUY THE WEST COAST OPERATIONS OF UNOCAL ALSO KNOWN AS "76 PRODUCTS" COMPANY FOR ABOUT $1.4 BILLION.

```
(*A-SEAL
   (PUNCTUATION PERIOD)
   (FORM FINITE)
   (TENSE PAST)
   (MOOD DECLARATIVE)
   (ARGUMENT-CLASS AGENT+THEME)
   (AGENT
      (*PROP-TOSCO
         (NUMBER MASS)
         (IMPLIED-REFERENCE +)
         (PERSON THIRD)))
   (THEME
```

```
(*O-DEAL
   (NUMBER SINGULAR)
   (REFERENCE INDEFINITE)
   (UNIT -)
   (PERSON THIRD)
   (COMPLEMENT
      (*A-BUY
         (FORM TOFORM)
         (TENSE PRESENT)
         (MOOD DECLARATIVE)
         (ARGUMENT-CLASS AGENT+THEME)
         (Q-MODIFIER
            (*K-FOR
               (OBJECT
                  (*U-DOLLAR
                     (ABBREV +)
                     (NUMBER PLURAL)
                     (REFERENCE NO-REFERENCE)
                     (UNIT +)
                     (NUMBER-UNIT
                        (*O-BILLION
                           (NUMBER SINGULAR)
                           (REFERENCE NO-REFERENCE)
                           (PERSON THIRD)
                           (UNIT -)
                           (QUANTITY
                              (*C-DECIMAL-NUMBER
                                 (NUMBER-FORM NUMERIC)
                                 (NUMBER-TYPE CARDINAL)
                                 (DECIMAL "4")
                                 (INTEGER "1")
                                 (MANNER
                                    (*M-ABOUT
                                       (DEGREE POSITIVE)))))))))))
         (THEME
            (*O-WEST-COAST-OPERATION
               (NUMBER PLURAL)
               (REFERENCE DEFINITE)
               (UNIT -)
               (PERSON THIRD)
               (Q-MODIFIER
                  (*K-OF
                     (OBJECT
                        (*PROP-UNOCAL
                           (NUMBER MASS)
                           (IMPLIED-REFERENCE +)
                           (PERSON THIRD)
                           (REL-QUAL
                              (*G-QUALIFYING-EVENT
                                 (EXTENT NONE)
                                 (EVENT
                                    (*A-BE
```

```
                                        (FORM FINITE)
                                        (TENSE PRESENT)
                                        (MOOD DECLARATIVE)
                                        (PREDICATE
                                           (*P-KNOWN
                                              (DEGREE POSITIVE)
                                              (Q-MODIFIER
                                                 (*K-AS
                                                    (OBJECT
                                                       (*PROP-76-PRODUCTS-CO
MPANY
                                                          (NUMBER MASS)
                                                          (IMPLIED-REFERENCE
 +)
                                                          (PERSON THIRD)))))
                                              (MANNER
                                                 (*M-ALSO
                                                    (DEGREE POSITIVE)))))
                                        (THEME
                                           (*PROP-UNOCAL
                                              (GAPPED +)
                                              (NUMBER MASS)
                                              (IMPLIED-REFERENCE +)
                                              (PERSON THIRD)))
                                        (IGNORE
                                           (*G-GAPPED-ARGUMENT
                                              (GAPPED +)))))))))))))
                    (AGENT
                       (*G-GAPPED-ARGUMENT
                          (GAPPED +))))))))
```

TOSCO关闭了一笔*GAP*以大约1．4billion美元买下*GAP**GAP*而且作为76个产品公司闻名的UNOCAL的西海岸机构的生意．

(3) IT WILL BUY LENDERS BAGELS FROM KRAFT FOR ABOUT $455 MILLION.

```
   (*A-BUY
      (PUNCTUATION PERIOD)
      (FORM FINITE)
      (TENSE FUTURE)
      (MOOD DECLARATIVE)
      (ARGUMENT-CLASS AGENT+THEME)
      (AGENT
         (*PRON-IT
            (REFERENCE NO-REFERENCE)
            (PERSON THIRD)
            (ANAPHOR +)
            (GENDER NEUTER)))
      (Q-MODIFIER
```

```
     (*G-COORDINATION
        (CONJUNCTION NULL)
        (CONJUNCTS
           (:MULTIPLE
              (*K-FROM
                 (OBJECT
                    (*PROP-KRAFT
                       (NUMBER MASS)
                       (IMPLIED-REFERENCE +)
                       (PERSON THIRD))))
              (*K-FOR
                 (OBJECT
                    (*U-DOLLAR
                       (ABBREV +)
                       (NUMBER PLURAL)
                       (REFERENCE NO-REFERENCE)
                       (UNIT +)
                       (NUMBER-UNIT
                          (*O-MILLION
                             (NUMBER PLURAL)
                             (REFERENCE NO-REFERENCE)
                             (PERSON THIRD)
                             (UNIT -)
                             (QUANTITY
                                (*C-DECIMAL-NUMBER
                                   (NUMBER-FORM NUMERIC)
                                   (NUMBER-TYPE CARDINAL)
                                   (INTEGER "455")
                                   (MANNER
                                      (*M-ABOUT
                                         (DEGREE POSITIVE))))))))))))))
  (THEME
     (*PROP-LENDERS-BAGELS
        (NUMBER MASS)
        (IMPLIED-REFERENCE +)
        (PERSON THIRD))))
```

它[这]从KRAFT以大约455million美元将买下LENDERS-BAGELS .

From these sentences, we can see some of the structures are not so regular, the inter lingua are more complex, and there are a lot of PP attachments which need special con sideration in the translation.

I still think Dr. Li's approach doing generation directly from the semantic represent ation is good for Chinese, because the intermediate syntactic representation will cau se some information loss. But do we still need those specific grammar rules in order to deal with some specific verb or noun? Our approach is:

1. try to use the minimal, universal grammar rules to cover most of the sentence stru

ctures, if the translation is reasonably well.

2. If we can deal with the specific situation in the lexicon, we never do it in the grammar rules. Our grammar rules should be as general as possible.

We do not expect the best translation, because in the CNBC domain, the sentence structures will not be so limited and regular, and there are more various situation will appear, also high accuracy is not a big concern in this domain, sentence coverage is more important. If we can achieve reasonably good translation, we keep the grammar rules as simple as possible. If this is not enough, we use lexicon to mark some items' specific properties, and use some general grammar rules to translate these specialties into the translation.

4.2 Implementation

Our prototype system in CNBC domain first deal with 50 sentences. Our system architectures are as follows:

```
generator.lisp  ---  load everything we need :  load kant
                                                compile the grammar file
                                                 load cnbcfun.lisp file
                                                load cnbc-sys.lisp file

cnbc-sys.lisp   ---  build hashtable for lexicon (cnbclexicon.chinese)
                                 and interlingua (cnbcworking.ir)
                     get Fstructure from the interlingua
                     use generator function to do generation

cnbcworking.ir  ---  the interlingua file

cnbc.gra --- grammar file for interlingua to Chinese generation

cnbcfun.lisp --- lisp files for handling the mutual impact between PP and its head

cnbclexicon.chinese  ---  mapping from interlingua lexicon to Chinese
                          1. for countable noun, we specify its unit
                          2. for adjective, we specify a feature NO-DE
                          3. use feature SUBCAT to classify the lexicons
                             which are under the same category according
                             to their characteristics
```

4.2.1 lexicon selection

We still use two steps approach for Chinese generation. The first step, we do lexicon selection for every interlingua concept head, and replace the interlingua head with the Chinese head.

Example 1:

MATTEL BUYING TYCO TOYS FOR $755 MILLION.

Before we do lexicon selection, the interlingua is:

```
(*A-BUY
   (PUNCTUATION PERIOD)
   (FORM PRESPART)
   (TENSE PRESENT)
   (MOOD DECLARATIVE)
   (ARGUMENT-CLASS AGENT+THEME)
   (AGENT
      (*PROP-MATTEL
         (NUMBER MASS)
         (IMPLIED-REFERENCE +)
         (PERSON THIRD)))
   (Q-MODIFIER
      (*K-FOR
         (OBJECT
            (*U-DOLLAR
               (NUMBER PLURAL)
               (REFERENCE NO-REFERENCE)
               (UNIT +)
               (ABBREV +)
               (NUMBER-UNIT
                  (*O-MILLION
                     (NUMBER PLURAL)
                     (REFERENCE NO-REFERENCE)
                     (PERSON THIRD)
                     (UNIT -)
                     (QUANTITY
                        (*C-DECIMAL-NUMBER
                           (NUMBER-FORM NUMERIC)
                           (NUMBER-TYPE CARDINAL)
                           (INTEGER "755")))))))))
   (THEME
      (*PROP-TYCO-TOYS
         (NUMBER MASS)
         (IMPLIED-REFERENCE +)
         (PERSON THIRD))))
```

The main lexicon head is:

(*A-BUY (CAT V) (ROOT "买下") (FOR ((ROOT "以"))))

And also we notice, the default translation of preposition "for" is:

(*K-FOR (CAT PREP) (ORG FOR) (ROOT "对于"))

So after the lexicon selection step, we get:

```
((CAT V) (ROOT "买下")
         (FOR ((ROOT "以")))
         (PUNCTUATION PERIOD)
         (FORM PRESPART)
         (TENSE PRESENT)
         (MOOD DECLARATIVE)
         (ARGUMENT-CLASS AGENT+THEME)
         (AGENT
           ((CAT N) (SUBCAT PROP-NOUN)
                    (ROOT "MATTEL")
                    (NUMBER MASS)
                    (IMPLIED-REFERENCE +)
                    (PERSON THIRD)))
         (Q-MODIFIER
           ((CAT PREP) (ORG FOR)
                       (ROOT "对于")
                       (OBJECT
                         ((CAT N) (SUBCAT CURRENCY)
                                  (ROOT "美元")
                                  (NUMBER PLURAL)
                                  (REFERENCE NO-REFERENCE)
                                  (UNIT +)
                                  (ABBREV +)
                                  (NUMBER-UNIT
                                    ((CAT N) (ROOT "million")
                                             (NUMBER PLURAL)
                                             (REFERENCE NO-REFERENCE)
                                             (PERSON THIRD)
                                             (UNIT -)
                                             (QUANTITY
                                               ((CAT NUMBER)
                                                 (NUMBER-FORM NUMERIC)
                                                 (NUMBER-TYPE CARDINAL)
                                                 (INTEGER "755")))))))))
         (THEME
           ((CAT N) (SUBCAT PROP-NOUN)
                    (ROOT "TYCO TOYS")
                    (NUMBER MASS)
                    (IMPLIED-REFERENCE +)
                    (PERSON THIRD))))
```

Why do we need two definitions for the preposition "for"? This is a special phenomena in Chinese. Because of the free sentence style and the use of various preposition phrase attachments in the CNBC domain, one translation for a specific preposition is not enough any more. In Chinese, one English preposition could have a lot of translation in different situation, and mostly this difference is caused by the head (a noun or verb) to which the preposition is attached. We use the lexicon to represent this knowledge, and there is no specific rules for dealing with this in grammar file except we do some condition check and do some replacement in the grammar file.

Example 2:

(*A-LOOK-AROUND (CAT V) (ROOT "四处找寻") (FOR ((phrase +) (root "*GAP*"))))

LOCTITE SAYS IT IS LOOKING AROUND FOR OTHER BUYERS.
LOCTITE说它[这]正在四处找寻*GAP*另外买家 .

"look around for sth." is translated as a verb phrase in Chinese, because without "for", "look around" can not be directly followed by a noun phrase; and without "look around", "for other buyers" do not have a complete meaning. So in this situation, if "look around" is followed by "for", there is no special translation for the preposition "for". We use some features in the lexicon to represent this knowledge.

Example 3:

(*A-SEE-AS (CAT V) (ROOT "看") (AS ((ROOT "作") (ba +))))

ANALYSTS SEE THE DEAL AS A PERFECT FIT FOR KELLOGG’S FAST-GROWING CONVENIENCE FOODS BUSINESSES.
分析家们把这生意看作对于KELLOGG的快速增长的方便食品产业的完美的适合 .

Usually, the English sentence:

see A as B

is translated in Chinese as:

把 A 看作 B

Which is different from the usual PP translation in Chinese, we will discuss the normal sentence order in the following section.

Example 4:

(*O-DAY (CAT N) (ROOT "天") (OF ((root "*GAP*") (headroot " 日子"))))

IT WAS A SCHIZOPHRENIC KIND OF DAY OF TRADING ALL DAY LONG:
它[这]一整天是SCHIZOPHRENIC类型的*GAP**GAP*交易的日子 :

Another interesting phenomena in Chinese translation for the PP attachment is:

A special preposition could require to use a different translation for the head (a noun or a verb). We notice the default translation for "day" is: 天, but if we select this default translation if "of" is attached to it, the translation is kind of weird, people usually never say in that way.

4.2.2 Sentence generation in Chinese

As we mentioned before, although Chinese is hard to determine its word order, we do have some regular grammar rules which can represent the sentence meaning in most of the cases. In order to cover more sentences in CNBC domain, we do not expect perfect translation, we prefer large coverage with reasonable translation. Now the problem is: What is a reasonable translation? It could have a lot of standards to define this concept, but in our system, we define this as: a translation which is grammartical and understandable, but it needn't be perfect.

In most of the cases, our general sentence grammar rule is: SVO.

For VP and NP, there are some variations:

a. VP PP (English) --> PP VP (Chinese) [normal case]

THEY CLOSED (VP) AT 59 7/8 (PP). -->
他们 在59 7／8(PP) 结束了 (VP).

b. VP PP (English) --> VP PP (Chinese) [special case]

LOCTITE SAYS IT IS LOOKING AROUND (VP) FOR OTHER BUYERS (PP). -->
LOCTITE说它[这]正在 四处找寻(VP) 另外买家(PP).

In this cases, "look around for sth." is a verb phrase, it doesn't make sense if we seperate the "for sth" apart from the verb.

c. NP PP (English) --> PP de NP (Chinese) [normal case]

AND BILLIONAIRE MARVIN DAVIS HAS SWEETENED HIS TAKEOVER OFFER (VP)
FOR CARTER-WALLACE (PP).
而且亿万富翁的marvin davis已经更加优惠 给予CARTER-WALLACE(PP) 的(de) 他的接管提供(NP).

d. NP PP (English) --> NP de PP (Chinese) [special case]

TEXT "IT WAS A SCHIZOPHRENIC KIND (NP) OF DAY (PP) OF TRADING ALL DAY LONG:
它[这]一整天是SCHIZOPHRENIC类型(NP) 的 (de) 交易的日子(PP):

4.3 Results:

See Appendix A.

5. Some issues:

In order to do translation directly from the semantic meaning of the interlingua, we determine the general syntactic structures according to some important concepts in the interlingua. For example: If the sentence head is "be", we know it must have "theme

" and "predicate". Another important feature is ARGUMENT-CLASS. We check this feature to predict the syntax of the Chinese. For example:

```
((:NUMBER 5) (:TYPE :SENTENCE) (:TEXT "ON WALL STREET THE DOW INDUSTRIALS MISSED "RE
CORD NUMBER NINE".")
 (:INTERLINGUA
    (*A-MISS
       (PUNCTUATION PERIOD)
       (FORM FINITE)
       (TENSE PAST)
       (MOOD DECLARATIVE)
       (ARGUMENT-CLASS AGENT+GOAL)
       (AGENT
          (*PROP-DOW-INDUSTRIALS
             (NUMBER MASS)
             (REFERENCE DEFINITE)
             (PERSON THIRD)
             (UNIT -)))
       (Q-MODIFIER
          (*K-ON
             (TOPIC +)
             (OBJECT
                         (*PROP-WALL-STREET
                            (NUMBER MASS)
                            (IMPLIED-REFERENCE +)
                            (PERSON THIRD)))))
       (GOAL
          (*O-RECORD-NUMBER-NINE
             (QUOTED +)
             (NUMBER MASS)
             (REFERENCE NO-REFERENCE)
             (PERSON THIRD)
             (UNIT -))))))
```

From the content: (ARGUMENT-CLASS AGENT+GOAL), we know the sentence structure in Chinese would be: AGENT VERB GOAL. But the problem is we don’t know how many values the ARGUMENT-CLASS could have. We believe we must limit the scope to some extent, so if the GLR parsing grammar can handle the limited scope of sentences, the interlingua structure will be limited as well, so our system definitely can handle all the phenomena in the interlingua.

6. Conclusion and future work

We have presented a system which generates Chinese sentences from interlingua representation. Our approach use two step generation: lexicon selection and syntactic generation, and based on different domain, translation coverage and accuracy requirements, we use different approachs to define and use the knowledge to select the suitable translation. Our implementation in Chinese generation proves that Genkit provides a very good tool for target language generation, and the interlingua representation is sufficient for Chinese generation.

Our system is scalable and extendable. In order to build a practical machine translation system, our lexicon and grammar rules need further extension, but our prototype system proves our approach is feasible, and the results are very promising. Within this 53 sentences, we can do good translation for 50 sentences, and the remaining sentences are due to the misrepresentation in the interlingua.