Project Report

---- the chinese generation system(CGS)

Wei Fang
房 蔚

This semester I have read and analysed the chinese generation system which was completed by Doctor Tangqiu Li. This system translates from interlingua to chinese, and successfully implements the translation of 215 interlingua sentences. The output of most of these sentences is very good. Here I am going to analyse this system in detail.

First I will talk about the system itself. The 215 interlingua sentences surely doesn't cover all the phenomenon of the interlingua system, but it is still very representative. This CGS has covered all the sentence structures which appear in the "ilt" file, and most of its grammar rule for chinese generation is general. For dealing with some specific sentence structures, it uses some specific grammar rules in order to get the best output for those sentences. I will analyse this system from general to specific in section 1, and then I will talk about the "ilt" which is used here in section 2. In section 3, I am going to suggest what we are going to do next to improve this system.

Section 1. CGS system

At first this system builds a *sentence-table* for the 215 interlingua sentences, and builds a *lexicon* for the mapping from interlingua to chinese. Second it use the sentence number index to get the interlingua, use function lexicon-lookup to translate the interlingua to the Fstructure which replaces the lexicon in interlingua with chinese lexicon. At last, those grammar rules for generation take effect now.

I think most of the grammar rules for generating chinese sentence from the interlingua is general.

1. The sentence coverage is as follows:

(1) label sentence

label : sentence.

for example:

Note: The heat content rating per unit volume ... --->
注意: ...

(2) hyphen sentence

--sentence.

for example:

= Anticipate the average outside temperature for the area of operation.--->
--...

(3) bullet sentence

bullet. sentence

for example:

1. Operate the engine until the engine reaches normal operating temperature.

(4) heading sentence

sentence (without punctuation)

for example:

Effect of Cold Weather on Fuel

(5) imperative sentence

for example:

Avoid sharp angles.

(6) discourse-cohesion sentence

discourse-cohesion, sentence.

for example:

However, the filter should not be too fine. --->
然而, ...

(7) pre-condition sentence

if pre-condition, sentence.

for example:

Only add the coolant conditioner precharge if you are not using Caterpillar antifreez

e.

A if B --->
如果 B, A

(8) reason sentence

because reason, sentence.

for example:

The new dipsticks are different because they have a "FULL RANGE" mark.

A because B --->
因为 B, A

(9) when-event sentence

for example:

Remember that the capacity of an eyebolt decreases when the angle between the support
ing members and the object becomes less than 90 *.

A when B --->
当 B 时, A

(10) although-event sentence

for example:

Although the PEEC system dissipates heat into the fuel, fuel heaters are still necess
ary.

Although B, A --->
尽管 B, A

(11) during-event sentence

for example:

You can use No. 2 diesel fuel in diesel engines during cold weather.

A during B --->
当 B(root is verb), A
在 B(root is noun) 期间 A

(12) until-event sentence

for example:

Operate the engine until the engine reaches normal operating temperature.

A until B --->
A 直到 B

(13) but-event sentence

for example:

The fuel heater should be mechanically simple, but the heater must be adequate for th
e application.

A, but B --->
A, 但是 B

(14) sequential-event sentence

for example:

Drain the oil and change the oil filter.

A and B --->
A, B

(15) or-event sentence

for example:

Add liquid conditioner or replace the coolant conditioner element.

A or B --->
A 或 B

(16) purpose sentence

for example:

Slowly loosen the radiator filler cap in order to relieve pressure.

A in order to B --->
为了 B(root is noun) A
为了 B(root is verb), A

All the above sentence structures are translated into chinese correctly,
and without any lack of generality.

2. Special tense markers

In chinese, we use some specific word as the tense marker in English. For example, we use "在" or "正在" as the progre tense; use "将" or "将要" as the future tense. Also we use "已经" as the perfect tense. These are correctly implemented in this system.

3. Case marker

No case marker in chinese. That means "he" and "him" is the same in Chinese.

3. Number and person

Generally speaking, there are no number and person agreement in chinese. But "I" is different from "we", "she" is different from "he", and "this" is different from "that", etc. All these difference have been implemented in this system.

4. Some limitations

In the generation grammar rules, we can find some rules are trying to evaluate the root. These kind of grammar rules are specific for some particular sentences or words. In order to achieve high accuracy translation, these kind of rules are necessary sometimes.

For example:
if the root is "design"(设计) and it is a passive sentence, the translation should be:

theme 是 为 purpose 设计的.

This kind of translation is the best. So we need some very specific rules to get this output. One way to make these kind of rules more general is to put a class of words together, and try to evaluate the class instead of the root. This is also the things need improving. How to find a good criterior to put the class of words together is what I should do next.

Section 2. The interlingua

Our CGS just uses a interlingua with 215 sentences. This interlingua covers:

*O- noun
*P- adjective
*E- verb
& adverb

Also besides all the annotation we mentioned above, a lot of other annotation are used here, such as:
negation, modal, event-frequency, predicated-of-event, and so on.

These annotations are just a part of the real interlingua system. To improve our system, we need to try to cover some other representative interlingua structures. This is the task of my next step.

Section 3. What I should do next semester

I have mentioned all the system's limitation and interlingua's limitation. First I think I should take some more representative sentence structures from the real interlingua system, and add some more rules into this system to cover all the phenomenon. And I also should make those rules more general to cover more sentence structures.