SOFTWARE AND LINGWARE FOR HUMAN AND MACHINE TRANSLATION

Nitaya Kanchanawan

I. INTRODUCTION

Translation is a rendering from one language into another. What the translator does is to capture the meaning in one language and express it in another language. It is understandable that this is a task of a human being who not only knows the two languages involved but also the content of the subject, as well as world knowledge and many other things.

Translation has been carried out by human beings since 2,000 B.C. In the last 40 years since the invention of computers there has been an attempt to use computers as translators. How can a computer do that since it is only a machine? To visualize this we have to imagine that a human being is also another kind of machine - a very sophisticated one. As human beings, when we start to translate something, we have to " feed " in the source language. Then analyze it to capture the meaning. In doing this we have to compare many things with the knowledge which is supposed to be stored in our brain. Finally, the meaning is expressed in an appropriate form in the target language.

How many steps we have to take, including the details of each step, have been crucial topics of discussion for a long time. Whatever they are, we have to follow some strategies. The problem is that sometimes we cannot specify what we did to get the translation.

That is what a human being is supposed to do when translating something. What about the computer?

The main difference between a human being and a computer is that the computer is a machine. It cannot do anything by itself. While a human translator " feeds " in the source language by himself, it has to be " fed " to the computer. While a human translator analyzes the text by himself, we have to tell the computer how to analyze it. While a human translator applies his knowledge, the computer must be told how to do this and what that knowledge is. While a human translator expresses the meaning in a target language, the computer must be told how to do this and in what form.

No matter how different they are, both human and machine translators have to deal with the how and the what. The " how " may be compared to software in the computer world, that is, how to translate. What is the process of translation?

The " what " may be compared to " lingware " or " linguisticware ". That is " what it is that we have to use or compare " in order to analyze the source language and to generate the target language.

II. SOFTWARE

Software means programs or instructions that control the operations of a computer in solving a given problem. The given problem in our case is translation.

Before writing programs or instructions telling the computer to translate, the translation process must be broken down into definite steps. What the set of definite steps of translation is may not be unanimously agreed upon, but one has to be chosen.

One of the sets of the translation process that is adopted in many Machine Translation (MT) systems is analysis, transfer, and generation. Analysis

may be divided into two phases: morphological analysis and structural analysis. Transfer may be divided into two phases: lexical transfer and structural transfer.

Finally, generation may be divided into two phases: structural generation and morphological generation.

Then programs are written telling the computer how to process the six steps mentioned above.

The ARIANE system of the Groupe d'Etudes pour la Traduction Automatique (GETA), Grenoble University, France, and the METAL system of the Linguistic Research Center, the University of Texas at Austin, Texas, USA, are two examples of such programs. Programs may be written in any computer language; this has nothing to do with the pair of languages to be translated from and to.

Compare these processes with those of the human translator. The human translator too has to tell himself how to do it. While the computer has to be told how to do it, the human translator can do it by his knowledge which may be obtained from experience or education. In other words, the human translator has this software in his head. Different human translators may have different software, just like the difference between the ARIANE system and the METAL system.

III. LINGWARE

" Lingware " or " linguisticware " is a new word coined to convey the involvement of computer application. We have already seen that in order to translate, not only do we have to know how to do it, but also we have to know what to compare and choose. In other words, " lingware " is dictionaries and sets of grammatical rules written in the form appropriate for computer application.

Given the six phases of translation, there are various kinds of dictionaries and sets of grammatical rules involved. For the morphological analysis phase monolingual dictionaries are needed. A set of grammatical rules is needed both in the morphological analysis and the structural analysis phases. Bilingual dictionaries are needed in the morphological transfer phase and another set of grammatical rules is needed in the structural transfer phase. Yet another set of grammatical rules is needed in the structural generation phase. Last, but not least, another set of grammatical rules and a monolingual dictionary are needed in the morphological phase.

While for most of the human translators various kinds of dictionaries may be frequently consulted, the application of grammatical rules may be taken for granted. For the computer, not only must every process be broken down into definite steps, but all the dictionaries and grammatical rules must be treated the same way. Thus the term " lingware " is used instead of just plain " linguistic data. "

Monolingual dictionaries in the morphological phase deal with the source language only. Bilingual dictionaries in the morphological transfer phase deal with the pair of languages to be translated from and to. The monolingual dictionary in the morphological phase deals with the target language alone. Since both the dictionaries and the sets of grammatical rules have to be written in special languages, just like computer languages, " lingware " is sometimes referred to as another kind of " software ".

IV. COOPERATION OF SOFTWARE AND " LINGWARE "

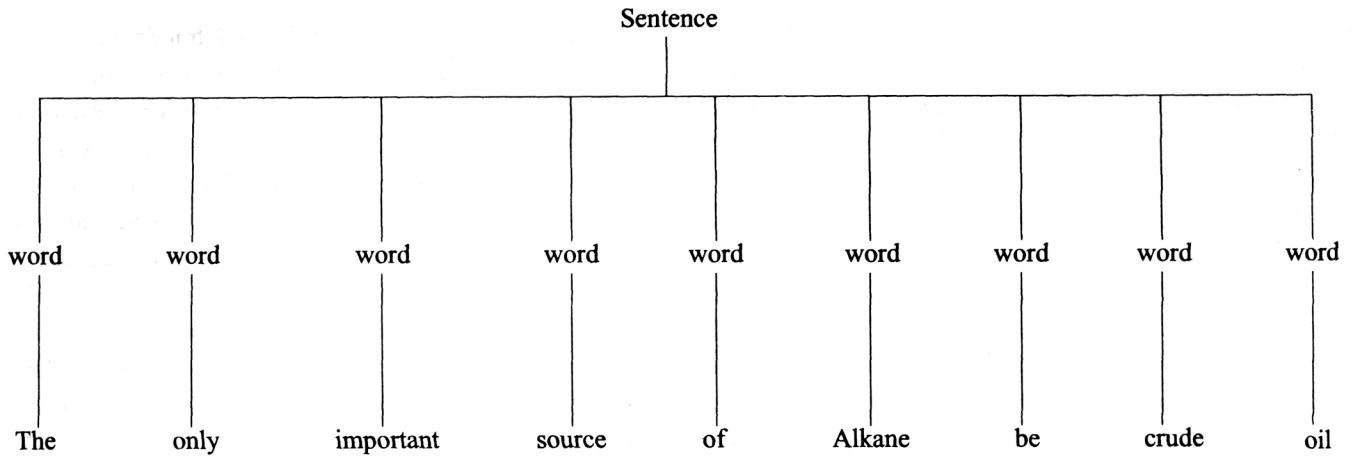
For human translation, what we need is a person who knows how to translate, the content of the text and world knowledge, and the source and the target languages. For machine translation what we need is hardware, software, and " lingware ". Hardware is the computer. This may be compared to the human translator himself. Software is the " know - how " of the translation process. This may be compared to the human knowledge of how to translate. " Lingware " is the dictionaries and grammatical rules to be operated by software. This may be compared to ordinary dictionaries, both in the head of the human translators and on the shelves, and the grammatical knowledge of the human translator.

To illustrate how software and " lingware " work in machine translation an example from the ARIANE system will be used. We are going to translate the following sentence:

" The only important source of Alkanes is crude oil " into Thai.

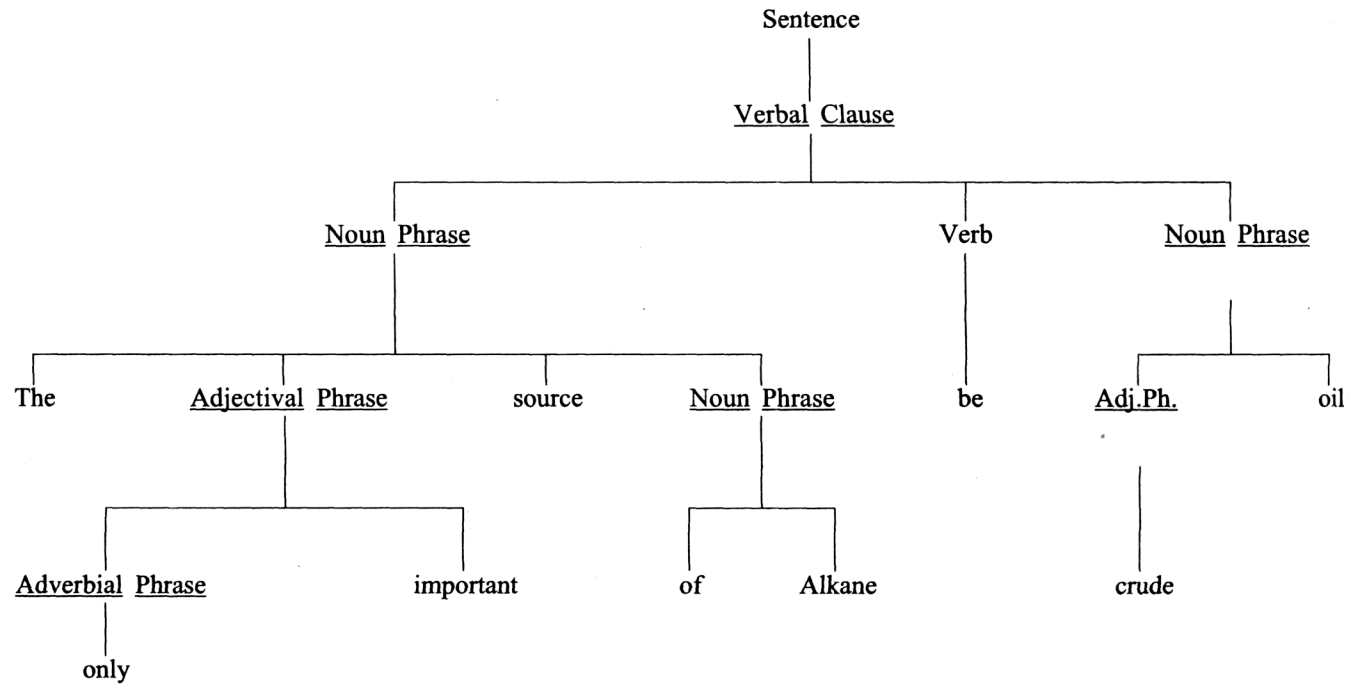
After this sentence is fed into the computer the first phase to be operated is morphological analysis.

In the morphological analysis phase the software will check every word in the sentence comparing them to monolingual dictionary entries together with the application of grammatical rules. The result will be a tree structure which looks like the following:



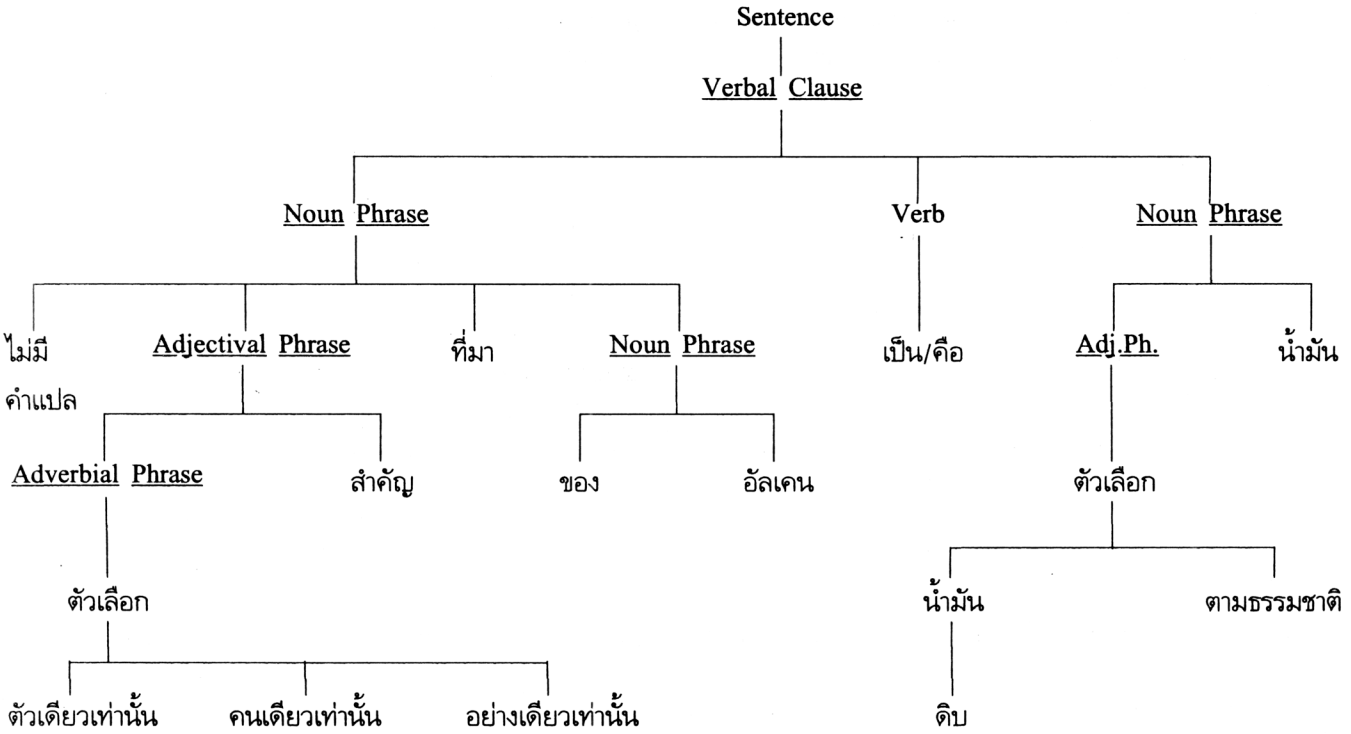
All the linguistic information on each word is kept in the computer memory for future reference. Notice that " Alkanes " is stripped down to " Alkane ". The " plural " feature of this word is stored in the memory. " Is " is manifested as " be ". The " present tense " and " singular " features are also stored.

The human translator may flip through this part without notice. This structure is fed into the second phase, the structural analysis. A set of grammatical rules helps determine that this sentence has the following structure:



A set of grammatical rules to determine English structures may be taken for granted by the human translator at this point. All the linguistic information, i.e., structural features, semantic features, and logical relations, are stored for future reference. This structure then is fed into the third phase, the lexical transfer. This is when the software consults various kinds of bilingual dictionaries: idiom dictionaries, general term dictionaries, and specific dictionaries.

The translated word or group of words is selected from the dictionaries within the limits of the software. The result will look like the following:

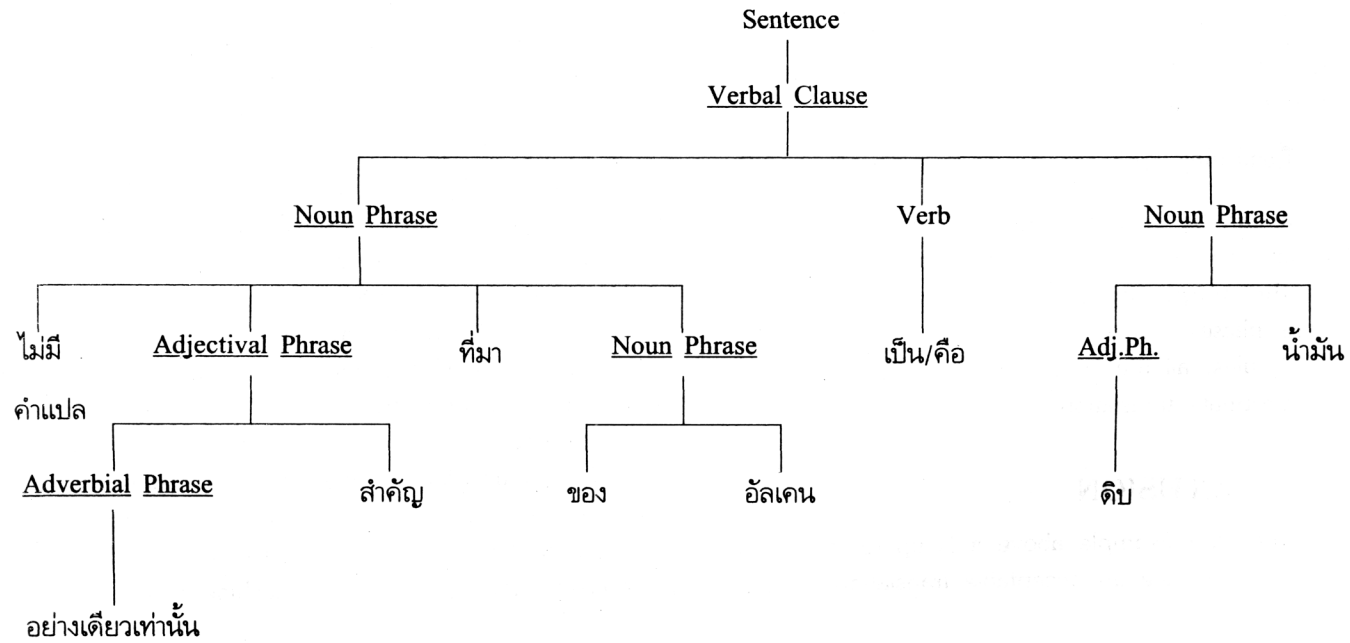


Notice that the computer, just like the human translator, cannot choose all the translations at this phrase. It is a well - known fact among human translators that not all translations can be determined at the morphological level. So it is left to the fourth phase to solve this problem.

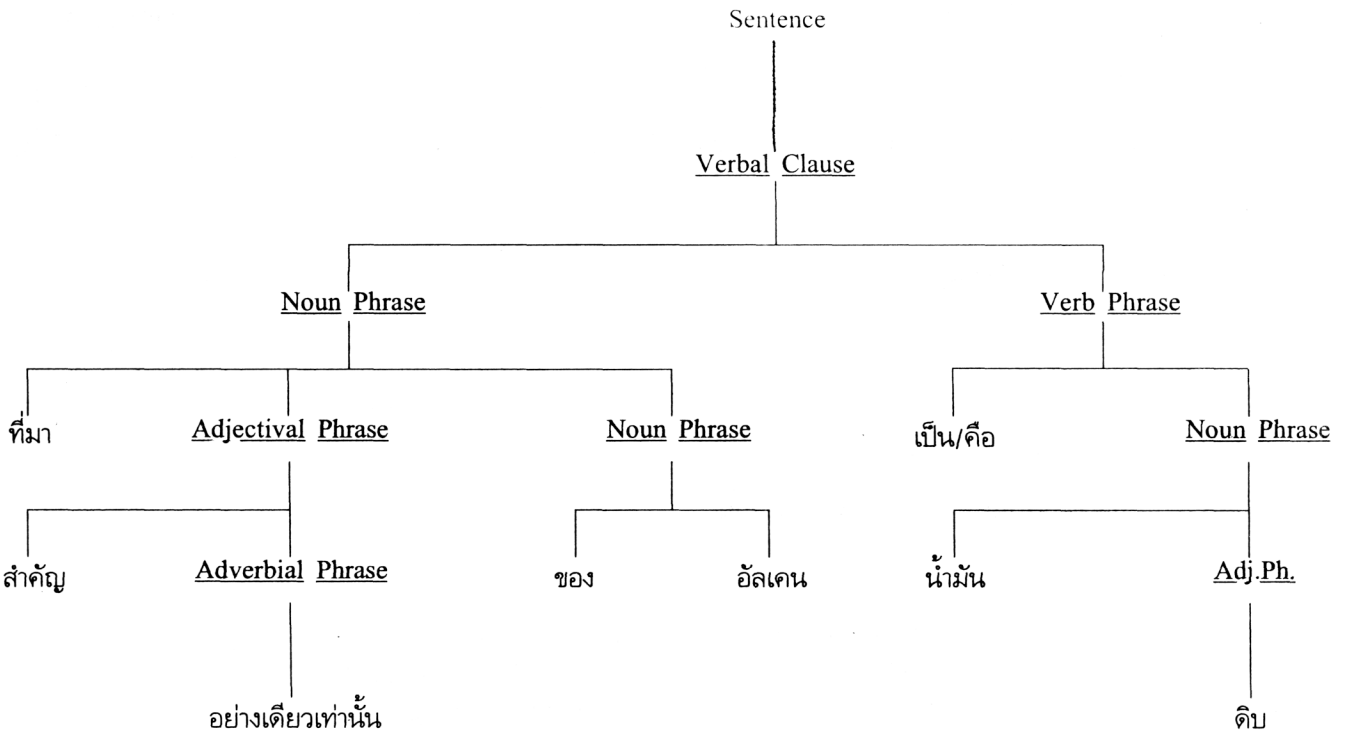
A set of grammatical rules at this point tells the software that if " only " modifies " human, " " **คนเดียวเท่านั้น** " will be selected; if it modifies all other " animates, " " **ตัวเดียวเท่านั้น** " will be selected; otherwise " **อย่างเดียวเท่านั้น** " will be selected.

The grammatical rules also state that if " crude " modifies " oil " it will be directly translated as " **ดิบ** " otherwise " **ตามธรรมชาติ.** "

Thus the result will look like the following:



Notice that the structure is still that of the source language, English. Grammatical rules in the fifth phase, the structural generation, will convert this structure into the target language structure, Thai, which looks like the following:



Then comes the last phase, the morphological generation. This phase is very important for inflectional languages. The surface form of each word will be generated here. Luckily Thai is an isolating language in that there is usually no change in the word form. Thus this phase only converts the tree structure into linear form:

ที่มาสำคัญอย่างเดียวเท่านั้นของอัลเคนเป็น/ คือน้ำมันดิบ

It can be seen that the program or software depends heavily on the linguistic information obtained from the morphological analysis and the structural analysis phases. Together with dictionaries and grammatical rules, all the linguistic information help select an appropriate translation.

V. CONCLUSION

From the example above it is up to the " lingware " to produce an acceptable translation. There are still many aspects that cannot be defined as a rule or rules at this point. In the example shown there is yet to be a rule or rules to determine when " be " is to be translated either as " เป็น " or " คือ " At the moment this is left to the posteditor to decide. Human translation needs to be refined the same way.

It is true that linguistic data alone is not enough for good translation. This is a limitation of machine translation. Only text in which the content is more important than the style is suitable for machine translation and there are a lot of these kinds of text now in this information - oriented society.

Whatever limitations there may be, both human and machine translation use two common devices. Both make use of software and " lingware " in the same manner.

Human translation may be more sophisticated, but machine translation is quicker and more consistent in some respects. Thus, leave some of the job for humans and some for the machine to do.

REFERENCES

Bennett, Wilfield S., and Jonathan Slocum. 1985. The LRC Machine Translation System, *Computational Linguistics*. Vol. 11, No. 2, April - June.

Boitet, C. 1982. ARIANE-78-VERSION 5. G.E.T.A.,

_____. 1984. Research and Development on MT and Related Techniques at Grenoble University (G.E.T.A.), Lugano Tutorial on Machine Translation, April.

Boitet, Guillaume, Quezel-Ambrunaz. ARIANE-78: An Integrated Environment for Automated Translation and Human Revision. Proc. of COLLING-82.

Kanchanawan, Nitaya. 1983. A Preliminary Linguistic Data Model for English - Thai MT by ARIANE System. Ramkhamhaeng University.

_____. 1984. A Preliminary Model for Thai Static Grammar and a Supplement to Transfer Phrase. Ramkhamhaeng University.

_____. 1985. A Model for Thai Static Grammar and Dynamic Grammar for Syntactic Generation. Ramkhamhaeng University.

Liu, Jocelyn, and Joseph Liro., 1987. The METAL English-to-German System: First Progress Report. *Computers and Translation*, Vol. 2, No. 4, October - December.

Slocum, Jonathan., 1985. A Survey of Machine Translation: Its History, Current Status, and Future Prospects. *Computational Linguistics*, Vol. 11, No. 1, January - March.

Vauquois, B. 1979. Aspects of Automatic Translation in 1979. IBM - Japan, Scientific Program.

_____. 1983. Automatic Translation. Proc. of the Summer School. The Computer and the Arabic Language. Ch. 9, Rabat.

Vauquois, B. and Boitet, C., 1985. Automated Translation at Grenoble University. *Computational Linguistics*, Vol. 11, No. 1, January - March.

APPENDIX A

ESSAI ESI 9 OCTOBER 1986 15H 33MN 465

SOURCE AND TARGET LANGUAGES IN2 - TH2

— INPUT TEXT —

THE ONLY IMPORTANT SOURCE OF ALKANES IS CRUDE OIL. IF AIR WHICH HAS BEEN HIGHLY COMPRESSED IS ALLOWED TO EXPAND, IT BECOMES COLDER. WE HAVE SEEN THAT FERTILIZERS ARE PRODUCED FROM NITROGEN COMPOUNDS.

VM/SP CONVERSATIONAL MONITOR SYSTEM

ทมาสาคญอยางเดยวเทานนของ อลเคน (เปน คอ) นามนดบ ถาอากาศซงถูกอดอยางสูง แลวถูกปลอยใหขยาย* อากาศจะเยน

(กวา ขน ลง) เรา (พบเหน) แลววาปุยผลตจากสารประกอบไนโตรเจน.

APPENDIX B

ESSAI ES2 13 MARCH 1987 15H 49MN 32S

SOURCE AND TARGET LANGUAGES IN2-TH2

--INPUT TEXT--

DURING THE LAST 30 YEARS THERE HAVE BEEN ENORMOUS DEVELOPMENTS IN THE USE OF OIL AND IT IS NOW THE PRIMARY SOURCE OF AN EXTREMELY EXTENSIVE RANGE OF ORGANIC CHEMICALS.

THE TARGET TEXT IN THAI CHARACTER

*ในระหว่างตลอดเวลา (30) ปีที่แล้วมานี้ การพัฒนามากมาย (มีปรากฏ) แล้ว ในการใช้น้ำมัน และมัน (เป็น คือ อยู่) บัดนี้ ที่มา (ขั้นต้น ปฐมภูมิ) ของ (ระยะ ขอบเขต แนว) กว้างขวางอย่างที่สุดของเคมีภัณฑ์อินทรีย์.