

面向科技文献的机器翻译

摘 要

机器翻译与科技文献翻译有着非常密切的关系。早期的机器翻译研究就是为了翻译科学技术文献服务的。21 世纪是信息社会,科学技术的发展日新月异,科学技术文献的数量呈爆炸式增长。科学技术文献的翻译对中国科学技术的进步意义重大。虽然传统的手工翻译质量很高,但是速度太慢,难以满足科学技术文献日益增长的需要。对于科技翻译工作者来说,如何面对新的严峻挑战,跟上时代的发展步伐,这将是一个新课题。本文从转变翻译观念,改进传统翻译方法入手,提出机器翻译、网络翻译和软件翻译的概念。同时就实行机器翻译辅助翻译、翻译项目管理和共享翻译资源的问题提出一些自己的建议。

前言提出了信息时代的特征和一些现代化的翻译手段。第一章,主要介绍了科技文献的定义,科技文献的特征,包括语法特征,文体要求,以及科技文献的翻译原则,翻译方法,最后概括出科技文献翻译面临的挑战。第二章,本文在叙述机器翻译的定义、结构特征之后介绍了机器翻译发展初期的实用主义倾向和它的历史发展。接着提出机器翻译的类型,策略。在翻译市场一节中,描述了为科技文献设计机器辅助翻译环境的新思路。第三章,作者主要描述了网页翻译和软件翻译的定义、发展和现状。第四章,笔者介绍了机器辅助翻译的核心技术(TM)之后推荐了两种机器辅助翻译工具,TRADOS 和 Yaxin CAT,并对他们进行了比较。第五章中,首先在认真比较人工翻译和机器翻译的过程后,作者用实际的例子,分析了机器翻译产生的一些错误,进而提出可以从两个方面来应对科技文献翻译和机器翻译中出现的问题:翻译项目管理和建立资源共享的平台。最后是结论。

关键词: 科技文献, 机器翻译, 机器辅助翻译, 翻译项目管理

Machine Translation of Sci-tech Document

Abstract

Machine translation is closely-tied with the translation of Sci-tech document. The research of machine translation is to serve translations of Sci-tech documents in earlier days. The 21st century is the information times. The development of science and technology is quick. Quantities of Sci-tech document grow explosively. Translations of Sci-tech documents play a significant role in the progress of Chinese science and technology. Although the quality of traditional manual translations is very good, the speed of translation is too slow. Traditional manual translations can not meet the need of translations of Sci-tech documents day by day. It is a new topic to face the new stern challenge and follow the development's steps of the times for technical translators. The paper will discuss the transformation of translation idea, and suggest carrying on machine translation, web page translation and software translation. Finally, the author states her views on the implementation of computer assistant translation, the managements of translation projects and information-shared platform.

In the foreword, the author introduces the characteristics in the information age, and proposes some modern translation methods. In the first chapter, the author mainly introduces the essential concepts of the Sci-tech document, for instance, the definition of Sci-tech document, characteristics of the Sci-tech document, including the grammar characteristics, the literary style's request, some translation principles of Sci-tech document, the translation method, and finally points out the challenges of Sci-tech document which translators have to faces. In the second chapter, the author narrates the concepts of machine translation. After explaining structure characteristics of machine translation, the author mainly lays emphasis on early pragmatism and its historical development, introduces the types of machine translation, the strategies and translation market. In the end of this chapter, the author draws out computer assistance translation,

which is a kind of new mentality to help Sci-tech document translators. In the third chapter, the author mainly describes the homepage translation, including the definition, the development and the present situation. In the fourth chapter, the author compares the process of human translation and machine translation, with the solid examples and the errors of machine translation. Two suggestions are given, one is project management of Sci-tech document translations, the other is to establish resources_shared platform. Final part is the conclusion.

KEY WORDS: Sci-tech documents, MT (Machine translation), CAT (computer assistant translation), Management of translation project

Introduction

1. Translation characteristics of information age

The translation is a key of international communication. It is said, Western European civilization gives credit to the translator. Society's progress cannot be separated with the translation obviously. Perhaps in the future world, to be or not be of a cultural tribe will be decided by translating a word accurately. Therefore, some people call out in alarm "Translate or die." But the public is still quite unaware of the gravity of the problems. Only when the human relation is damaged, the strikes are created, disorder, the legal matters, fatal accidents appear, and sometimes even the wars break, will the public pay attention to the importance of translations. In 1970, several Americans ate one kind of virulent mushroom. Doctors did not know how to treat them, and the patients died within the days. When a European pharmacist heard the radio on this matter through broadcasting report, he was surprised. Because he knew the method of treatment which was utilized in 1959, and the method went public in 1963 successfully. Why American doctors did not know it after 7 years? There was an estimation that the method of treatment was merely published in the European continent but did not published in American magazines. This example shows that no prompt and correct translation will cause the enormous harm to the countries.

0.1 The times call for new translation method

Along with the arrival of the information times, Sci-tech documents grow increasingly in the world. The prompt translations of Sci-tech documents are the key of social production, the competitive power and development of science and technology. Traditional translation method characterized by the massive consumptions of human resource, physical resource and financial resource, does not keep the pace of the development. A new translation method based on reader's direct demand—MT and CAT system (computer-assistant translation) has already been proved practicable and suitable for need of the information times

0.2 The phenomenon of mixing uses of English and Chinese is a character information age

In recent years, numerous new English terms have appeared, especially in science and technology. There appeared the translations which absorbed English terminology directly. For

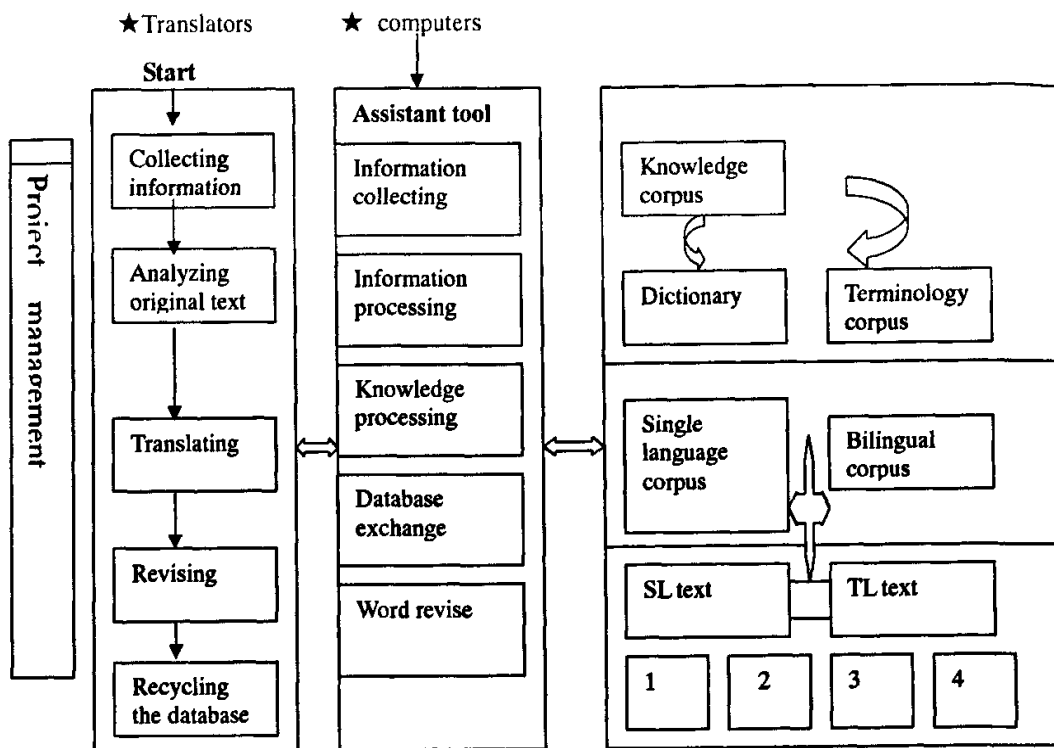
example, Internet, Homepage, Expro, and so on. All these computer's terminologies appear in the lines of Chinese texts. Besides, there are other translations like `WTO, MBA, etc. The reason of these phenomena is that popularity of English-speaking is becoming bigger and bigger in the world .Only English has surpassed to other any languages in the international contact. It has become the most commonly used instrument and the communicative language; Second, in information age, the new technical glossary appears too much and too quick. The people do not have time to consider coordination of each terminology and the ingenious translated names. Third, the reason is to avoid the misunderstanding from the translated text .Sometimes different translators in the social sciences and the natural sciences translate the identical terminology differently. People, who do not read the original text, can be muddled by these terminologies. For instance , "AIDS", in the mainland the translation is "艾滋病".But in Hong Kong and Taiwan, it is translated as "爱死病".Is it the way the target text can be avoided misunderstanding? Fourth, we may not find the coordinative names for the new terms in Chinese .The method of borrowing words has not only enriched this nationality's language, but also made the national language easily connect with other nationalities' languages, and exchange various cultures more conveniently.

2. Modern methods

Modern method means CAT (computer-assistant translation) and the management of translation project. It is reported that translator who is engaged in the Sic-tech translation reach the number of over 1 million in our country. But most of them are individuals who are in manual operations and in the small workshop management. The work way is extremely backward. Many people work with dictionaries, the pens and the manuscript papers, and do not make fully use of the advanced method of the modern science and technology. The way of the translation is not in time to absorb overseas' advanced technology at the high speed and the great breadth. To solve this problem and enhance the working efficiency, the working of Sic-tech translator must be fundamentally transformed.

CAT is a improved way of translation. The nature of CAT is as followed: In limited conditions (Automatic MTs can not made a great breakthrough in the quality),people make computers serve translation to guarantee a product and increase efficiency

Way of project management of translation



As the picture shows, people complete the tasks at the period of the time under the assistant of the computer. Computers offer an integration environment managing the kinds of database. People can collect information, analyze documents, storage parts of translation, edit, and complete translation in this condition. Computer shoulders the repetitive and mechanical work. Translators complete the work that computers are not able to do. The writer will recount it in detailed in the later chapter.

*1, database in the project; 2, original analyzing; 3, e-note; 4, edition management.

Chapter 1 Basic information of Sci-tech documents

1.1 What are Sci-tech documents

The Sci-tech documents are the records which carry information of science and technology. Documents can keep the database, the marks, the images, the audios, pictures, figures and so on. From this definition we can see, the documents have three basic attributes: be knowledgeable, be recorded and be material. The functions of the Sci-tech documents are knowledge-storaged, the transmission and the exchanges of Sci-tech information.

1.2 Classifications of Sci-tech documents

There are 4 kinds of documents according to nature of the material information storaged :

- Printed form
- Microform ,Microfilm microfiche
- Audio-Visual form ,Compact audio/video form
- Machine Readable form and Electronic Publication : floppy, Magnetic tape, CD-ROM, online, network ,multi-media.

There are 3 ranks of documents according to the information research:

- Primary Literature/ Info ,for example,

Journal literature, Sci-tech report, Proceedings
Patent specification, Paper, Thesis, Dissertation,
- Secondary literature /info , for example,

entry reference citation record
Abstract ,Index ,Catalog, Bibliography
- Tertiary literature/ info , for example,

Reference book,

There are 7 kinds of documents according to the readers

- Sci-Tech book ,for example,
Monograph, Anthology, Textbook, Reference book,
- Sci-Tech Journal (magazine, serial, periodical,)
- Sci-Tech Report
- Patent Document: Patent specification/bulletin/aids,
- Conference paper:

Proceedings, compilation, collected paper, records
Meeting: workshop, seminar, colloquium, symposium, conference,
- Dissertation, Thesis
- Standard document, Product material; Technical file; Government publication

Texts may be classified in the following categories according to the usage:

- Presentation of new knowledge and descriptions of its application in practice: papers in the proceedings of learned societies and technical institutions, University theses.
- Integrations and reviews of existing knowledge and experience: articles in scientific and technical journals, separate pamphlets and reports, reference manuals.
- Educational material: syllabuses, certificates and diplomas, textbooks, popular science publications.
- Documents relating to engineering and industrial applications: contracts and specifications for works, reports on tests and analyses, trade catalogues, publicity and directions, patent specifications. National standards and international standards (UNESCO 1957).

Example

Gong, Victor, and Norman Rudnick, Eds. AIDS: Facts and Issues. New Brunswick: Rutgers UP, 1986.

Example 2

Schware, R. "Climate Change in Water Supply: How Sensitive is the Northeast?" Climate Change and Water Supply. Washington, D. C.: National Academy of Sciences, 1986: 82-98

Example 3

Chun, Lu, and K. Y. Lan, "Dynamic Analysis of Clamped Laminated Curved Panels." Composite Structures. 30.4 (1995): 389-396

ISBN: International Standard Book Number 0-471-81086-x

ISSN: International Standard Serial Number 1003-3513

CN11-2856/G2

Example 4

Wang, C. M., et al, Software for Performing Gray-Scale Measurements of Optical Fiber End Faces, Nat. Inst. Stand. & Technol, Report NIST/TN-1370, 1994.

Example 5

Gough, P. A., et al, "Electrothermal Simulation of Power Semiconductor Devices." Proc. of the 3rd International Symposium Power Semiconductor Devices & ICS. Ed. Shbib, M. A. New York: IEEE, 1991: 89-94.

Example 6

Zaccarin, Andre and Liu Bede, Method and Apparatus for Deterring Motion Vectors for Image Sequences. U. S. Patent 5210605, 1993.

Example 7

Dong, Siyi Terry, The Modeling Analysis and Synthesis of Communication Protocols. Diss. Michigan U, 1991, Ann Arbor: UMI, 1992. 9228854.

1.3 Characters of Sci-tech document translations

Texts require varied translations not only as regards types of documents, but also as regards their special language characters. The Sci-tech documents listed above are classified in principles. Though each type calls for more or less different treatments. There are some general characters of Sci-tech documents

Sci-tech documents are rigorous in the structure, well-organized, good logic to expound laws of development, discovery and invention of things. Besides, Sci-tech documents require

succinct words, clear content. Sometimes in them, there are figures, charts and formula to express accurate content of the paper. The unity and coherence are required for Sci-tech documents, too.

1.3.1 Language characters of Sci-tech documents

We knew that, the language has three major constitutions, the pronunciation, the glossary and the grammar. The glossary is the foundation, the grammar is a rule, and pronunciation is an outer covering of the language. Here, the article will mainly discuss the grammar characters

1) Word meanings

Now, the technical develops unexpectedly, the renewal changes each new day, each kind of English name, English guide, English screen demonstrated in front of people. It is advantageous to grasp the glossary characteristic of the English Sci-tech document. Understanding the original meaning, studying and translating the academic information, translators give their contributions to society's development

①Specialized ordinary glossary

In the Sci-tech documents, the noun is the most important constituent. Different scientific domain has its sets of technical expressions. An ordinary word can has the specific meaning .we can give examples in computer science: storage 储藏储存——存储器, instruction 教育, 指导——指令, processor 加工者——处理机(器), interfact 交界面, 相互关系——接口, terminal 末端, 终点——(终端)and so on. The words and expressions which constitutes with two individual words or parts of them, come from the general glossary and have own meaning, but differ in the different specialized applications, like molecule, hydraulics. Specialized ordinary glossary has the single meaning. The concepts of expression are clear. There are also the exceptions, for example, "the level", means "主平巷" in mining, but in water conservation means "水位", in load means "负荷" in the machinery, in the water conservation then means "泥沙".

②Massively borrowed glossary in the neighboring discipline

Computer technology borrows the glossary in the neighboring discipline, like electron and radio technology, automation, mathematics .The meanings of the word do not change obviously, for example: input 输入; output 输出; transmission 传送、传输; control 控制; process 过程, 处理 and so on.

③Utilized compound word a lot

The computer glossary utilizes massive compound words widely. The new words are free in the order. For example:

flip-flop 双稳态触发器, flowchart 流程图, keyboard 键盘. offline 脱线, general-purpose terminat 通用终端, high-resolution graphics 高分辨率图形 and so on.

④ Utilized abbreviation widely

Abbreviation is simple and easy to remember. They can not only reduce the length, but also speeds up the transmission, and frugal store up spatially: CPU (Central Processing Unit) 中央处理机(器, 单元); ALU (Arithmetic-Logical Unit) 算术逻辑单元(运算器); RAM (Random Access Memory) 随机存取存储器; PROM (Programmable PROM) 可擦除的可编程只读存储器; EPROM (Electrically PROM) 可擦除的可编程只读存储器; MOS (Metal Oxide Semiconductor) 金属氧化物半导体 and so on.

2) Nominalization

A Grammar of Contemporary tells us that Nominalizations are used widely in the Sci-tech document, because of the requirements of the text, such as succinctness, object, accurate, more information. What the text is emphasis on is the fact, not the behavior.

Example 1,

Archimedes first discovered the principle of displacement of water by solid bodies.

阿基米德最先发展固体排水的原理。

Example 2,

The rotation of the earth on its own axis causes the change from day to night.

地球绕轴自转, 引起昼夜的变化。

“the rotation of the earth on its own axis “ changes the complex sentence into simple sentence.

Nominalization makes the concept more accurate and rigour

Example 3,

If you use firebricks round the walls of the boiler, the heat loss, can be considerably reduce

炉壁采用耐火砖可大大降低热耗。

Example 4,

Television is the transmission and reception of images of moving objects by radio waves.

电视通过无线电波发射和接受活动物体的图像。

“the transmission and reception of images of moving objects by radio waves “ emphasize the fact and predicate focus on the ability of transmission and receiving.

Because Sci-tech documents describe the objective law or principle, writer should tries to avoid the use of the first person and the second person, and put the important information forward.

3) Sentences

① Declarative sentence

When indicating the definition, the theory and the laws as well as elaborating phenomenon, describing experiments, "Copula-predicative structure" is commonly used.

For example, a computer is a machine whose function is to accept database and to process them into information.

② Imperative sentence

When it is not necessary to indicate the subject, we use imperative sentence to elaborate proposition, problems, parts of equipment and operating instructions, to introduce craft, and literature annotation, and so on.

③ Passive voice structure

In the Sci-tech documents, the passive voice structure is used a lot, because: (1) the passive structure has little subjective color, the multi- objectivities. It is suitable to describe objective things very much (2) the way to explain the object is in the subject position can draw more attention of the reads;(3) the structure which reveals succinctly, is easy to remember.

For example,

Attention must be paid to the working temperature of the machine.

应当注意机器的工作温度。

This steel alloy is believed to be the best available here .

人们认为这种合金钢是这里能提供的最好的合金钢。

④ Compound sentence and non- predicate verb

Coordination and long sentences as well as the non- predicate verb phrase are massively used in Sci-tech documents, because the complex sentence structure can contain more information.

For examples:

A direct current is a current flowing always in the same direction.

直流电是一种总是沿同一方向流动的电流。

Radiating from the earth, heat causes air currents to rise.

热量由地球辐射出来时,使得气流上升。

A body can move more uniformly and in a straight line, there being no cause to change that motion.

如果没有改变物体运动的原因,那么物体将作匀速直线运动。

Vibrating objects produce sound waves, each vibration producing one sound wave.

振动着的物体产生声波,每一次振动产生一个声波。

⑤ long sentence

In order to indicate a complex concept, and make its logic strict and its structure compact,

long sentences are often used, for example,

The efforts that have been made to explain optical phenomena by means of the hypothesis of a medium having the same physical character as an elastic solid body led, in the first instance, to the understanding of a concrete example of a medium which can transmit transverse vibrations, and at a later stage to the definite conclusion that there is no luminiferous medium having the physical character assumed in the hypothesis.

为了解释光学现象，人们曾试图假定有一种具有与弹性固体相同的物理性质的介质。这种尝试的结果，最初曾使人们了解到一种能传输横向振动的具有上述假定所以认为的那种物理性质的发光介质。

⑥ Common sentence pattern

In the Sci-tech documents, several specific sentence patterns are used frequently. They form the technical style which distinguishes from other texts. Such as "that sentence", participial phrase pattern, abbreviate sentence pattern and so on, for examples:

It is evident that a well lubricated bearing turns more easily than a dry one.

显然，润滑好的轴承，比不润滑的轴承容易转动。

It seems that these two branches of science are mutually dependent and interacting.

看来这两个科学分支是相互依存，相互作用的。

The switching time of the new-type transistor is shortened three times.

新型晶体管的开关时间缩短了三分之二。（或---缩短为三分之一。）

Microcomputers are very small in size, as is shown in Fig.5.

如图 5 所示，微型计算机体积很小。

In water sound travels nearly five times as fast as in air.

声音在水中的传播速度几乎是在空气中传播速度的五倍。

1.3.2 Standardization of the modern Sci-tech documents

The standardization of Sci-tech document became the objective request and the inevitable trend with science and technology development. In information times, computer becomes popular, and takes more roles in our daily life. How to research, storage, exchange and study Sci-tech documents is a question. The way is to Standardize Sci-tech documents ,which became important method in the scientific research .American "Dialogue" , European "European space agency" and ESDAC "(European space database center) have stored the massive standardized Sci-tech documents now . They are significant for the world.

Standardization of the modern Sci-tech documents consists of five -parts: 1)topic and outline;2)introduction;3)body;4) conclusion;5)notes and bibliography.

1) Topic and outline

A topic is the focus of a document. It sets forth the proposition or the main point of view of the document is going to prove or maintain, and is usually in one sentence. The document statement should be descriptive, and not have an argumentative edge. Our country's technical publication request that the topic is not to surpass 20 characters, generally noun phrase constituted. Therefore, topic must reflect the research center. The word must be concise. Writers refuse long words and expressions even complex sentence structure to express it. Topics are significant to retrieval.

The topic should be followed by an outline of the paper. The topic and the outline will guide the writer in writing and will enable the future reader to grasp the main points of the paper at a glance.

2) Introduction

The introduction usually explains the writer's motive, intention or purpose in discussing his or her topic, and its scope and focus. It also provides the background or the situation the reader may need. In other word, the introduction should make clear why the writer chooses writing a paper and how much he or she intends to cover. It may raise some questions the paper is going to answer, or put forward the view or views it is going to elaborate. It may also make clear the method or methods of research the writer intends to adopt in the paper.

Generally speaking, the introduction should present some basic ideas of the paper and arouse the future readers' interest, It does need too long. It is about 600 to 1200 printed characters.

3)Body

The body is the bulk, the biggest part of the paper. This is where the writer elaborates his or her ideas in detail. It is advisable to divide the body into several sections with or without headings or subheading. Usually, paper on linguistics tends to use headings. But the writer should be careful not to use too many headings or different kinds of numerals before sections and paragraphs for they may confuse rather than help the reader.

4) Conclusion

The conclusion should in some way be connected with the introduction. It is a summary or restatement of the point of the view put forward, or an answer to the question posed, in the

introduction. .

5) Notes and bibliography

Notes may be placed at the bottom of the page on which the references occur (footnote), or given in a consecutive series at the end of the paper (endnotes). Obviously; the former is more convenient to the reader, while the latter makes the typing easier.

The bibliography should be placed after the endnotes on a separate sheet or sheet.

1. 4 Criteria of Sci-tech document translation

Criteria of Sci-tech document translation are the principles translators must comply with. Translation of Sci-tech documents proves itself to be a hard and challenging work due to their unique characteristics including serious content, compact style of writing, highly practical knowledge of science and technology. In this part the author intends to focus on the process, the standard and method of the translation of the Sci-tech documents.

1) Requirement of faithfulness and expressiveness

According to Yan Fu's translation criterion including faithfulness, expressiveness and elegance, faithfulness means that translator should correctly understand and express the thought, content and style of the original; expressiveness means that translator should use as expressive language as the original and elegance means translator should use refined words to make target text graceful. Sci-tech documents are more professional and accurate than other texts. Therefore, translator should render Sci-tech documents by the way of literal translation and try his best to keep source text's language forms including words, sentences structures and rhetoric, and meanwhile to achieve the purpose of making TL smooth and clear.

2) Requirement of resemblance

In a Sci-tech document, resemblance of the words and the sentences is the performance of systematic and rhetoric. In different Sci-tech documents, resemblance of the words and the sentences shows the need of regulation.

3) Technical terms

The language of science and technology has its own characteristic vocabulary. There are a lot

of technical terms used for given scientific disciplines, as well as semi-technical words common to all scientific disciplines. New terms continuously emerge and are introduced into English and other languages in large quantities.

The standardization and unification of translating Sci-tech terms is essential to scientific research and development, and scholar's cooperation and exchange. Newmark says that technical translation is primarily distinguished from other forms of translation by terminology, although terminology usually only makes up about 5-10% of a text.

Early in the 1870s, the Translation Office attached to the Jiangnan Machine-Building Plant in Shanghai prescribed the following methods for translating technical terms:

- a) Establishment of a technical term by use of the name available in Chinese. If the relevant name has already appeared in Chinese but has no entry in available English- Chinese dictionaries, one can consult books published in China about science and technology by Chinese scholars and foreign missionaries, or visit Chinese businessmen or manufacturers who know the name.
- b) Establishment of a new technical term. If the term hasn't appeared in Chinese and must be established, the following methods can be used. First, a Chinese character component (most of Chinese characters are composed of two or more components) is added to a common noun to form a new term with the same pronunciation as the common noun. Second, two or more characters are used to explain the matter; thus, these characters constitute the term. Third, transliteration is undertaken.
- c) Compilation of English-Chinese glossaries. Translators should write down newly translated words and compile glossaries or dictionaries so that other translators can compare their own work with them. In addition, translators can also consult Westerners.

Chinese sci-tech translators have always attached great importance to the standardization and unification of technical terms. At the beginning of this century, an organization was established for the examination and approval of scientific and Sci-tech translation and its research in technological terms consisting mainly of translated words. Since the founding of the People's Republic of China in 1949, China has strengthened terminological study and formulated principles and policies concerning the unification of scientific and technological terms. The year 1985 saw the foundation of the China National Committee for Natural Scientific Terms, an authoritative organization to examine, approve, promulgate and supervise scientific and technological terms on behalf of the Chinese government. The organization has so far promulgated and published terms in 30 subjects such as astronomy, physics, biochemistry, electronics, agronomy, and medical science. Overseas editions in eight subjects have also been published, printed in the original complex form of Chinese characters. The promulgation and publication of the terms of these subjects play an important part in Sci-tech translation, scientific research and academic exchange.

To translate Sci-tech terms in a standardized, precise manner, Chinese technical translators usually adopt the Sci-tech terms promulgated by the China National Committee for National Scientific Terms.

1.5 Types of Sci-tech document translation

There are types of translation for different Sci-tech purposes. In technical translation the following five general types can be distinguished:

a) Complete Translation: in such a translation, the entire SL text, whether a scientific report of hundreds of pages or a short abstract of several lines is translated, usually sentence by sentence and paragraph after paragraph. Nothing is intentionally omitted.

Translation scholars often discuss problems on the basis of the complete translation. It is considered to be faithful to the original and wholly expressive of its content.

b) Selective Translation: in this type of translation, only part of the SL text is selected and translated to produce the TL (TL) text. According to its purpose, the TL text may consist of paragraphs or sentences from the SL text. For example, when translating an original research paper, the description of the experiment process and the results may be translated and the rest omitted. In patent specifications, the summary and description of the invention may be translated, the other parts not taken into consideration.

c) Condensed Translation: in this type of translation, the TL text represents a systematic abridgement of a SL text. It retains basic information and deletes materials which have little reference value for scientists and technicians. In a condensed translation, sentences or paragraphs of the SL text may be inverted or rearranged.

d) Summary Translation: this type of translation is a summary of the SL text. It retains the key words and a few sentences capturing the main points of the SL text. In a summary translation, translators must reorganize the information of the SL text.

e) Composite Translation: this type of translation is related to two or more SL texts in the same subject. A composite translation may be completed cooperatively by two or more translators if the SL texts are written in two or more languages. Before starting research on a new subject, scientists usually want to know what methods are current in the field chosen. A composite translation is actually part of these feasibility studies.

The above can be regarded as a basic taxonomy of translation types, although these types may overlap one another.

may overlap one another.

1.6 Methods of Sci-tech document translation

When translating Sci-tech documents, we should comply with the general criterion of translation. Actually, there are 6 methods of Sci-tech document translation we can adopt. The writer will explain them with examples one by one.

1) Find out the accurate meaning based on specialized knowledge

With the rapid development of science and technology, Sci-tech document uses different kinds of words of disciplines, for instance, : “Sophistication”, its original meaning is “世故的人”, now means “复杂程度” eg: sophistication of the microprocessor(微处理器的复杂程度)

The dihydrate is partially dewatered to hemihydrate under the influence of the increased temperature in the cementmills.

石膏在水泥磨中受到高温影响,部分脱水成半水石膏.

“dihydrate” means “二水化合物”, according to the context and common sense, it is translated as “石膏”.

2) Transfer the nature of the word

- a) The instrument design by this company is of precision.该公司设计的仪器是很精确
- b) Herer, the verb “revision” “transfer to an adj
- c) The number of devices or terminals that use the line is determined by the experts who design the communication or network, system.

设计通信网络或系统的专家确定使用这种线路设备或终端的数目

In original language, “The number of devices or terminals that use the line” is the subject. when the sentence is translated into TL, it becomes the object.

3) Add more information

The heat applied is controlled by the standard time temperature curve, with the fire resistance ratinn (in hours) determined by the time before failure criteria are met.

按照时间某一温度标准曲线进行加热试验的持续时间从受到火的作用起自到失去支持能力或完整性被破坏或失去隔火作用等任一条件出现, 即到了耐火极限(小时)

Here adding “破坏条件” in Chinese translation is to express the original meaning more accurate.

4) Translate into the antonym of the word

It is possible to maintain an accuracy of test result of $\pm 0.1\%$.

测试结果的误差可能在 $\pm 0.1\%$ 范围内。

" accuracy ",its original meaning is“精确度”。“ $\pm 0.1\%$ ”means“误差范围” .In Chinese traditional texts ,the two words can not appear in the one sentence, so, " accuracy " is translated into its antonym.

5) Summary first then explain one by one

A fire protection engineer shall be a graduate of an engineer curriculum of accepted study and shall have completed not less than 4 years of full-time engineer work, indicative of growth in engineering competency and achievement, three of which shall have been in responsible characters of fire protection engineer work.

负责防火工程设计的工程师应具备三项条件:(1)有得到承认的防火工程设计专业的大学学历;(2)至少有4年的防火设计工作经验;(3)取得一定的工作成绩

During translation, we summary the information and explain later, so as to make text clear, simply.

6) Adjust sentence order

Because the air and material forces are balanced throughout the classification zone, precise separation takes place which results in average capacity increases of 30%and reductions of 20 to 30`% in specific power consumption when producing ordinary cement.

在生产普通水泥时,整个选分区的风力和物料力平衡,选分精确,因此平均生产力可提高30%,单位耗电量可减少20-30%.

We put the adverbial clause “when producing ordinary cement.” in front of the TL to have an emphasis.

1.7 Problems of Sci-tech document translation

Let us take a review of both Sci-tech translation practice and theory in ancient china. Translation has a 3000-year long history in China and is instrumental in the development of the Chinese national culture. Although translations carried out during the Ming Dynasty were mainly on science and technology: mathematics, astronomy, medicine, hydrology etc. Technical translations during the Ming Dynasty facilitated the scientific and technological development of ancient China So successful were the Ming translators as pioneers on technical translation, that some of the translated technical terms are still in use today. However, translation practice was

during the Ming Dynasty was not so influential in terms of the history of translation in China. During the Tang Dynasty, there was translation practice accompanied by a quest for systematic translation theories, while during the Ming Dynasty, the main purpose of translation was to introduce western technical knowledge.

In 1978, two years after the end of disastrous "cultural revolution," China adopted its "open door and reform" policy. A new and dynamic atmosphere for translation emerged. With the booming economy, all sorts of institutions, especially those in the developed coastal areas, where the central government, greatly spurring the development of foreign trade, economic and diplomatic relations have granted preferential policies, need more and more qualified translators and interpreters. Although literary translation remains a very important part of translation practice, translations concerning international trade, foreign affairs, technology, information science etc. are the main tasks of the 500,000 translators. The urgent need of qualified translators and interpreters has prompted higher educational institutions with foreign languages departments or foreign languages educational institutions to establish translation as a degree course. Some students receive training so as to be a good translator, who should study not only special knowledge and but also common knowledge. Management of terms and the process of the translation are also required to learn. Actually, the situation of training is unmet demands of the market.

Some translation companies comprehensively launches at the historic moment, in the late of 90's. They have the quite bright market and kinds of orientations, as followed:

Technical translations. Their customers are mainly the domestic companies with large-scale technical projects, as well as foreign corporations, which are ready to, or have already entered the Chinese market;

Legal, economics and trade translation. Their terminal customers are individuals who go abroad to study, visits friends and relatives, and who are engaged in the economic and trade activities in some Chinese and foreign companies;

IT translation. Because this kind of company's key point is the website localization, the website content translation, and the E-business depending on the website operations. These are possible the biggest new variable factor of growth in the translation industry.

Generally speaking, along with the development, translation companies met with some problems. Interior management of these translation companies is irregular. Translators of the companies are employed from the society. Moreover, most of their work way is the part-time jobs. These translators who have grasped some foreign languages are numerous, but few of them

received systemic and specialized translation training before entering the translation companies. Though there are advantages of these translators to open up and further the translators' specialized background, and serve the socialization and the marketability. The negative influence is also obvious, that is, making the chaotic translation markets go further and lose the faith of the buyers because of the low quality. Besides, there is malignant competition between the companies.

Chapter 2 Machine Translation

2.1 The times thirst for MT

According to Bible, the Babylonians wanted to make a name by building grand Babel Tower, "With its top in the heavens.". Raged by their bold act, God disrupted Babylonians' work by confusing the language of the workers so that they could no longer communicate with each other. The unfinished tower was later called Babel, which, of course, was no more than a fairy story, but it really showed, on one perspective, that man has never ceased to find a way that can communicate with each other without language barrier.

From translation to Esperanto, from Esperanto to machine translation (MT), human beings tried a variety of ways to overcome language obstacles. Although someone once foresaw that we, at last, will return to the original state before the Babel and speak only one language, before that, human beings will go on to revolt against god's will and find ways to clear languages barrier

In translation history, it has been a dream to make use of machine, later computer to carry out translation task more quickly, more accurately. When history strides into the 21st century, we embrace Information Ages, which will boost the development of MT into a new era. By means of Internet, the latest achievements of world civilization, including the science and technology, economy and politics, culture and art, get a chance of spanning over geographical obstacle, becoming the common wealth of human beings.

If the twentieth century has been called as the translation or reproduction age (Benjamin, 1923), perhaps the 21st century will be an age of MT. We are seeing an exponential increase in development of science and technology, but any attempt to introduce and assimilate advanced culture from foreign country will go through language conversion process-translation. For example, though the EEC (European Economic Community) employs as many as 1600 translators, it does not seem to meet the requirement of translation volume. In 1967, 80,000 scientific journals were being translated annually. Now, demand to translation has far surpassed supply, which is restricted by the workload of translators. Under this circumstance, the MT's status quo and prospect are naturally attracting more and more public attention.

After long-term research, the results of MT have been a great progress, though still far from ideality. Since the birth of MT, the discussion of the preconditions, capability, and limitations of MT has become a controversial issue in linguistic circles. In view of the complexity of the subject

matter, it was quite natural that in the course of this discussion, a wide spectrum of opinions emerged, ranging from profound skepticism to exaggerated optimism'

I will list some popular misconceptions, including exaggerated optimism and profound optimism, about MT, and the contemplation to these misconceptions will be the preface of the whole paper:

- 1) MT is a waste of time because none of marketable MT system can translate literary works.
- 2) Generally speaking, the quality of translation you can get from an MT system is very low, which makes them useless in practice.
- 3) The emergence of MT will threaten the jobs of translators.
- 4) The aim of MT must be fully automatic systems producing translations at least as good as Those made by human translators.

2.2 Definition and categorization of MT

MT is the use of computational techniques to translate text from one natural language into another; in short, it means translation by using computer. In broad sense, MT can be seen to include such computer applications as compilers and compression programs, etc., which convert a file in one computer language into a file in another computer language. In narrow or general sense, we refer MT as one kind of natural language processing (NLP).

As far as NLP is concerned, there are basically two types of translation, one of which is the type of human translation; the other is the type of MT.

1) Human translation. As we all know, a human translator performs all the steps in the translation process by using his knowledge, creativity, and experience.

2) According to the varying degree of human intervention, MT, in research field, can be further subcategorized into different types, such as

- ① **Machine-Assisted Human Translation (MART)**. A human translator performs the translation, but he/she uses the MT system as an assistant tool to improve or speed up the translation process.
- ② **Human-Assisted MT (HAMT)**, The SL text is revised or standardized by a human translator before (check), during (proofread), or after (post-editing) .it is translated by the MT system.
- ③ **Fully Automatic MT (FAMT)**, The SL text is input into a MT system as a file, and the computer produces a translation automatically without any human intervention.
- ④ **Fully Automatic High quality Translation (FAHQT)** can produce reasonable draft translations without the need of check, proofreading and post-editing, which is the dreamboat of

translations without the need of check, proofreading and post-editing, which is the dreamboat of human beings.

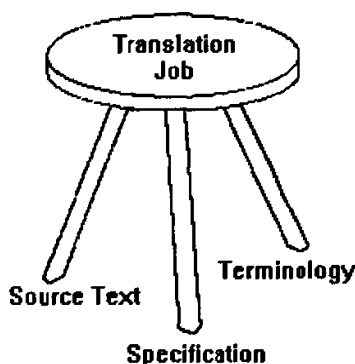
When people hear the word MT, they usually think the last two types-FAMT and FAHQT. In those cases, the SL text is input into the computer as a file, and the computer produces a translation automatically without any human intervention.

Obviously, the ultimate goal of MT we seek is to realize FAQMT, although current performance of commercial MT has made someone think that goal much ridiculous, why? The reason I will explain here after.

For a long time, automatic translation between human languages (MT) is the most fascinating but, at the same time, the most difficult domain in science history, a long-term scientific dream with enormous social, political, and scientific importance. Since the invention of computer, MT has been one of the earliest applications. However, turning dream into reality has turned out to be a much harder, and in many ways, a more interesting task than we first expected. Even today there remain many fundamental problems in Mt still unsolved, to some degree, MT has become a daily work.

2.3 Structure of MT

A translation project can be thought of as sitting on a tripod whose three legs are the source text, the specifications, and the terminology. If any of the three legs is removed, the project falls down.



Picture 2.1.1

2.3.1 Source text

Obviously, no translation can be done without a source text (i.e. the document to be translated). But for MT, an additional basic requirement is that the source text be available in machine-readable form. That is, it must come on diskette or cartridge or tape or by modem and end up as a text file on your disk. A fax of the source text is not considered to be in machine-readable form, even if it is in a computer file. A fax in a computer file is only a graphical image of the text, and the computer does not know which dots compose the letter "a" or the letter "b". The conversional source text on paper or in a graphical image file isn't a machine-readable form. The text using imaged character recognition (ICR) is not usually accurate enough to be used without human editing. Human editing is expensive, and add an unacceptable cost component to the total cost of MT. Thus, for MT to be appropriate, it is usually necessary to obtain the word processing or desktop publishing file from the organization that created the source text. But this is only one of many requirements.

2.3.2 Specifications

All translations projects have specifications. The problem is that they are seldom written down. Specifications tell how the source text is to be translated. One specification that is always given is what language to translate into. But that is insufficient. Should the format of the target text (i.e. the translation) be the same as that of the source text or different? Who is the intended audience for the target text? Does the level of language need to be adjusted? In technical translation, perhaps the most important specification is what equivalents to use for technical terms. Are there other target texts with which this translation should be consistent? What is the purpose of the translation? If the purpose is just to get a general idea of the content of the source text, then the specifications would include "indicative translation only." An indicative translation is usually for the benefit of one person rather than for publication and need not be a high-quality translation. Thus, publication-quality translations are high-quality translations (and are usually the result of human translation), while indicative translations are low-quality translations (and are usually the result of MT). These two types of translation are not normally in competition with each other, since a requester of translation will typically want one type or the other for a given document and a given set of specifications. Sometimes, the two types are complementary, such as when an indicative translation is used to decide whether or not to request a high-quality translation of a particular document. In this environment, an indicative translation may be requested for a number of documents, and, using the indicative translations, the requester may select one or two documents for publication quality translation.

As previously mentioned, indicative translations are usually done using MT and high-quality translations are usually done using human translation. This fact reveals a basic difference between humans and computers. Humans, with proper study and practice, are good at producing high-quality translations but typically can only translate a few hundred words an hour to approximately a thousand words an hour, depending on such factors as the difficulty of the source text. Even with very familiar material, human translators are limited by how fast they can type or dictate their translations. Computers are good at producing low-quality translations very quickly. Some MT systems can translate tens of thousands of words an hour. But as they are "trained" by adding to their dictionaries and grammars, they reach a plateau where the quality of the output does not improve. By upgrading to a more powerful computer, the speed of translation improves but not the quality. By upgrading to a "more powerful" human translator, the quality of translation improves but not necessarily the speed. Here we have a classic case of a trade-off. You can have high speed or high quality but not both.

Indicative translation (high speed, low cost, but low quality) represents a new and growing market but does not substantially overlap with the existing market for publication quality translation. The existing market, variously estimated at 10,000,000,000 to 20,000,000,000 US dollars worldwide per year, is primarily for high-quality technical translation. If, on the one hand, your specifications include low quality (barely understandable) translation, then MT is for you, and you can stop reading right here. If, on the more likely hand, your specifications include high-quality translation, then it is not obvious that MT is appropriate for your current translation job. Here quality would be measured by whether the target text is grammatical, accurate, understandable, readable, and usable. Usability can be measured by selecting tasks, such as maintenance operations, which can be accomplished by a source-language reader with the help of the source text and seeing whether those same tasks can be performed by a target-language reader with the help of the target text. Such measurements are notoriously expensive, but a skilled reviewer can accurately predict usability simply by studying the source and target texts. A target-language monolingual person can measure grammaticality, and understandability, and readability, which are progressively more stringent requirements. But accuracy requires the assistance of a skilled bilingual person who examines both the source and target texts.

2.3.3 Terminology

The treatment of terminology could have been included solely under specifications. But terminology is so important that the actual terminological database (also called a "term base") supplied with a source text has been listed as a third essential component of a translation job. The

aspect of terminology that does fit under specifications is the requirement that the translation job use a certain term base into order to achieve desired consistency. Let me explain what I mean by consistency. Translation requesters typically want the terminology in their translated documents to mesh closely with terminology in related documents. For example, a software company will want all revisions of a software manual to use the same terms as the original, to avoid confusing readers. Translation requesters should track all terminology relevant to a given document and deliver that terminology to the translation provider along with specifications and source text. The specification component of the job tells what appropriate term base to use and, as is all too common, tells what to do if a source-text term is missing from the term base. The terminology component of the job contains the term base itself.

Now we can define an appropriate translation job (for a human or for a computer) as one that sits on a stable tripod. It must include a source text (in machine-readable form if for MT); it must include well-defined follow the specifications; and it must include any specified term base. In addition, we can define an appropriate translation as a translation that combines the source text and the term base in a way that matches the specifications. Note that I said "appropriate" translation, not "good" translation. A poor (low-quality) translation may be appropriate if the specifications include a requirement for a fast, indicative translation.

2.4 History and present condition of MT

The idea of using computers to translate or help to translate human languages is almost as old as the computer itself. In general, MT is one of the oldest non-numeric applications of computers. Its history has been colorful and eventful and has been influenced by the politics, science, and economics of different periods of modern history. Here is a brief summary of its development.

Pre-computer: This period was not the beginning of the advent of MT, but an essential period for the idea of MT. Some of the ideas that had influenced MT were already current or at least existed in the pre-computer period. Since at least the 17th century scholars and philosophers had proposed the use of language-neutral representations of meaning in order to overcome linguistic barriers. More recently, a mechanical procedure for carrying out translation was proposed by the Russian in 1946.

Early pragmatism: We can first see the proposals for the use of numerical techniques in MT as early as in 1947, just after the invention of computer. At that time, computers had just been successfully employed in deciphering encryption methods during the Second World War. A memo from Warren Wear proposed some specific strategies for using computers to translate natural languages. This memo was the beginning of MT research in the USA and in the rest of the world,

with the first public demonstration of a Russian-English prototype MT system around the world. In the early 1950s, research of MT was necessarily modest in its aims, as it was constrained by the limitations of hardware, in particular, inadequate memories and slow access to storage, and the unavailability of high-level programming language. Apart from these, research was done inadequately without necessary assistance from the language experts in those fields of syntax and semantics. Syntax was a relatively neglected area of linguistic study and semantics was virtually ignored in the United States thanks to the behaviorist inclinations of the leading scholars. It was therefore not surprising that the first MT researchers turned initially to crude dictionary based approaches, i.e. predominantly word-for-word translation, and to the application of statistical methods. Warren Weaver himself, in the 1949 memorandum, which effectively launched MT research, had advocated statistical methods alongside cryptography, which was soon recognized as being irrelevant, and more futuristically the investigation of universal interlinguas.

With such limitations, early researchers set out with modest aims. They knew that whatever systems they could develop would produce low quality results, and consequently they suggested the major involvement of human translators both for the pre-editing of input texts and for the post-editing of the output, and they proposed the development of controlled languages and the restriction of systems to specific domains. Above all, they proposed that MT systems could progress by the cyclical improvement of imperfect approaches, i.e. an application of the engineering feedback mechanism with which they were familiar. In this atmosphere the first demonstration systems were developed, notably the collaboration between IBM and the Georgetown University in 1954.

The outcome of these early demonstrations was, however, that the general public and potential sponsors of MT research were led to believe that good quality output was achievable within a matter of a few years. The belief was strengthened by the emergence of greatly improved computer hardware, the first programming languages, and above all by developments in syntactic analysis. It was not clear which methods would prove most successful in the long run, so US agencies were encouraged to support a large number of projects.

Early researchers also suggested the major involvement of human translator the pre-editing of input texts and for the post-editing of the output. They advocated the development of controlled languages and the restriction of systems to specific domains.

Perfectionism: At the end of the 1950s, researchers in the United States, Russia, and Western Europe began to see things in an optimistic way. They were confident that high-quality MT of scientific and technical documents would be possible within a very few years. This might also be a response to the emergence of great-improved computer hardware, the first programming language, and above all developments in syntactic analysis. Since it was not clear which method would prove

most successful in the long run, US Agencies had to support a large number of projects. As a result of enthusiasm for MT spread for theories and methods for the achievement of “perfect” translation.

Of course, one could still hear voices, which sounded different from the dominant “perfectionism”. For instance, researchers at Georgetown University and IBM accepted the long-term limitations of MT in the production of usable translations. In 1960, Bar-Hillel was strongly critical of the theory-based projects, particularly those investigating Interlingua approaches and the non-feasibility of FULLY AUTOMATIC HIGH QUALITY TRANSLATION (FAHQT).

The “quiet” period: In 1996, the ALPAC report, which was supported by government sponsors of MT, the National Academy of Sciences of the United States, demonstrated that MT was not effective for the government because it is too expensive comparing with its efficiency. This highlighted the “failure” of MT research gave an end to the initial over-optimism in MT in the USA. After the publication of the report, funding in MT was transferred mainly to the research of Artificial Intelligence (AI) and Computational Linguistics. At this time, only few groups in USA survived. For some years after this report, research continued on a much-reduced scale, and resources were redirected toward more fundamental questions of language processing that would have to be answered before any translation machine could be built. This period lasted until about 1975.

The 1970s and Operational MT: Dismayed by the perfectionism approach, researchers began to look for sophisticated translation tools, that is, translation workstations, which can make their work more productive.

Some successes could be showed in three main strands: computational tools for translators, operational MT systems involving human assistance in various ways, and “pure” theories research towards the improvement of MT methods.

In the early 1970s, operational systems came into being for the continued effort in MT. SYSTRAN began Russian-English translations for the US Air Force in 1970, while Mateo began translating weather reports in 1976. Also in 1976, the Commission of the European Union (now called European Community) installed an English-French version of SYSTRAN. The development of translation tools was made possible due to the fact that real-time interactive computer environments were available since 1960, word-processing appeared in the 1970s, and microcomputers together with networking and large storage capacities were mass-produced in the 1980s.

The late 1970s and early 1980s witnessed an increase in interest in MT. the Aurora project from the European Community began in 1982, which was influenced by the work done at Grenoble and Saarbrücken since the 1960s and 1970s. Similarly, Japan started the project in 1982, and the USA began its Knowledge-Based MT in earnest in 1983. Some commercial systems also began to

appear at this time. A number of companies, especially some large Japanese electronics manufactures, began to develop MT software for workstations. A number of products appeared for personal computers, and various MAT tools such as translation memory began to be commonly used. "Translation Memory" enables the information stored, reuse and revise later as sources of example translations. It has also been recognized that all current commercial and operational systems produce output must be edited or revised if it is to be of publishable quality.

This period work on speech translation began to be taken into consideration by many researchers. At this time people first used statistical approaches to MT.

Late 1990s and MAT: At the end of the last ten years in 20th century, research on MT was invigorated by the coming of corpus-based method, notably the introduction of statistical methods and of example-based translation. Statistical techniques have brought liberation from the increasingly evident limitations and inadequacies of previous exclusively rule-based (often syntax-oriented) approaches. Problems of disambiguation, anaphora resolution and more idiomatic generation have become more tractable with corpus-based techniques. On their own, statistic methods are no more the answer than rule-based methods have been, but there are now prospects of improved output quality, which did not seem attainable a decade ago. As many observers have indicated, the most promising approaches will probably integrated rule-based and corpus-based methods. Even outside research environments integration is already evident: we can see many powerful translation engines on personal computers, translation on the Internet, widespread use of translation memory and translator's work benches, multimedia and software localization, as well as an increased interest in Example-based MT. This period was the most promising ever from the birth of computer.

The above we have briefly introduced different stages of MT development. MT was the first computer-based application related to natural language. Ten years ago, it was large organizations that were the typical users of MT, such as the European Community, the US government, etc. Today, there are over 60 major companies all over the world who produce and sell computer-aided translation system. The most distinguished example is the METEO system, which provides French translation for the weather reports used by airlines, shipping companies, and others. Some manufacturers find it possible to translate, largely automatically, maintenance manuals used within their organizations (not by their customers) by asking the technical writers to use only certain words and only in a carefully prescribed ways. Few small companies and free translators also use MT system, although translation tools such as online dictionaries were becoming more and more popular. However, ongoing commercial success in Europe, Asia, and North America continued to illustrate that, despite the imperfection, the quality of fully automatic MT (FAMT) and human-assisted MT (HAMT) system really meet some users' practical needs. At present, translation

systems are produced and sold by companies around the world (Fujitsu, NEC and others in Japan, Germans and others in Europe, SYSTRAN in North America and Power Translator and Kingsoft in China.)

2.5 Methodology of MT

After the introduction of history of MT, We will have a look at the development of MT methodology. Perhaps the flaw in methodology of MT can partially explain the deficiency of MT

MT research method can be represented from two perspectives, one being from the application of linguistic theories, another from what has been actually practiced by MT research.

1) The linguistic approach

MT research has been regarded as a field in which new linguistic formalism or new computational techniques can be tried out. In other words MT has been seen as a test-bed for linguistic theories, because non-experts can judge the quality of MT and translation.

The relevant theories were information theory, categorical grammar, transformational-generative grammar, dependency grammar, and stratification grammar in the 1950s and 1960s; artificial intelligence, non-linguistic knowledge bases; formalism such as lexical –foundational grammar, generalized phrase structure grammar, head –derived phrase structure, Montague semantics in the 1970s and 1980s. neural networks, connectionism, parallel processing, and statistical methods and others in the 1990s.

It has been found out later that these new theories, which were successful in their initial trials on small samples, have turned out to be problematic in the end.

- The practical approaches

These approaches can be further divided into 3 strands.

① Direct Systems

The first MT systems were Direct Systems. They simply translated the SL (SL) text to the Target (TL) directly using a bilingual dictionary as reference. Think of the hand-held travel dictionary that converts phrases that travelers use most frequently. Direct MT systems do not analyze or disambiguate the SL text. They translate only simple text, or text with a very low level of ambiguity (some are capable of finding entries for past participles, gerunds, noun plural forms, and adjective forms). If a SL word has more than one meaning, this approach could produce the wrong results in the TL. For example the translation of the “*I am fine*” response to “*How are you?*” could conceivably become “*I am a traffic ticket*” in the TL. Mistakes such as this will amuse your hosts in the visiting country, who will still appreciate your effort to communicate with

them in their language, but they are unacceptable in a formal translation.

② Indirect Systems

a. The transfer approach

According to the majority transfer view of MT, a certain amount of analysis of the source text is done in the context of the SL alone and a certain amount of work on the translated text is done in the context language of the text language, but the bulk of the work relies on comparative information about the specific pair languages. This is argued for on the view that translation is, by its very nature, an exercise in comparative linguistics. The massive European Union participated, was a TRANSFER system. The Japanese share the general perception that the transfer approach offers the best chance for early success.

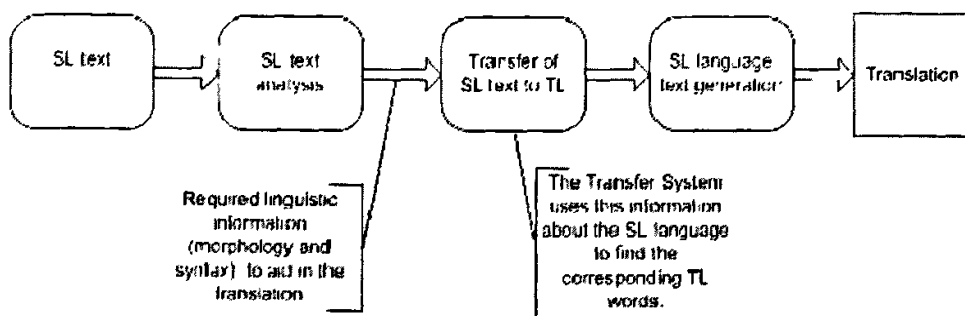
b. The interlingual approach

The method is taken as a move towards robustness and overall economy in that translation between all pairs of a set of language in principle requires only translation to and from the “INTERLINGUA” for each member of the set, if there are n languages, n components are therefore required to be translated into the interlingua and then into the TL.

In “The development and use of MT systems and computer-based translation tools”, John Hutchins explains the differences between these to MT systems:

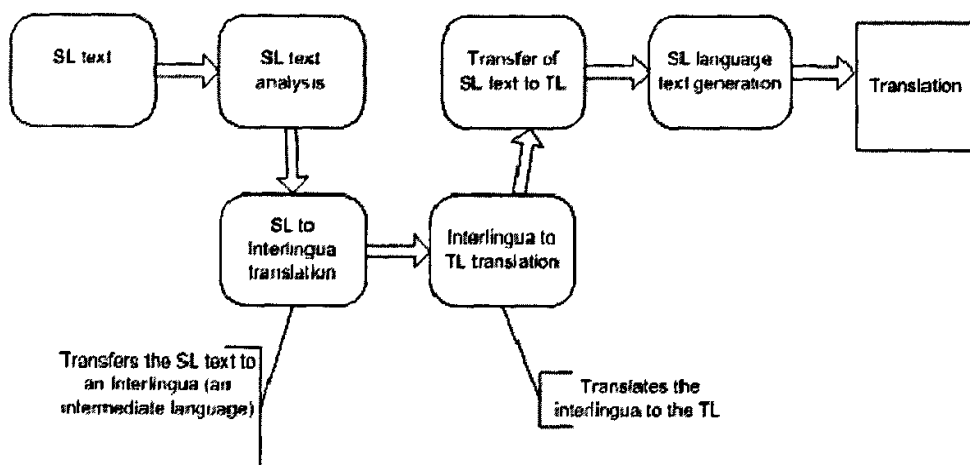
“There have in fact been two basic ‘indirect’ approaches. One abstract representation is designed to be a kind of language-independent ‘interlingua’, which can potentially serve as an intermediary between a large numbers of natural languages. Translation is therefore in two basic stages: from the SL into the Interlingua, and from the Interlingua into the TL. Other indirect approach (in fact, more common approach) the representation is converted first into an equivalent representation for the TL. Thus there are three basic stages: analysis of the input text into an abstract source representation, transfer to an abstract target representation, and generation into the output language.”

c. Transfer Based Indirect MT System



A Transfer Based Indirect MT System includes analysis and parsing of the SL text, independently of the TL text. The results of this analysis are used during the transfer to TL process to find the corresponding words in the TL.

Interlingua MT Systems



The Interlingua system adds two additional steps to the process: Translation to an intermediate language (interlingua) before transferring the SL text to the TL.

According to Hutchins & Somers (1992), an Interlingua is the intermediate representation of meaning that “includes all information necessary for the generation of the target text without ‘looking back’ to the original text. The representation is thus a projection from the source text and at the same time acts as the basis for the generation of the target text; it is an abstract representation of the target text as well as a representation of the source text.”

Some researchers used an artificial language, like Esperanto as the Interlingua because artificial languages are considered to be more regular and consistent in their morphology and syntax.

Commercially available MT systems fit the three basic system types (‘transfer’, ‘direct’ and

'interlingua'). Some best known of the systems in the industry, such as Systran, Logos and the Fujitsu Atlas systems are based on 'direct translation', but they are vastly improved versions of their predecessors. They are highly modular, easily modifiable and extendable. Systran powers the popular Babel Fish Translation that is available in the Internet. Systran was originally developed in the 1970 to translate Russian to English. Today, it offers a large number of language pairs, including most European as well as some Asian languages. The Commission of European Communities bought an English-French version of this system even though the quality of the translation. After a considerable investment in time and effort by the Commission's evaluators and lexical coding specialists, the quality of the translation had improved enough to do post level editing in multiple language pairs.

There have been also a number of 'transfer' systems. One of these is METAL, which was supported by Siemens, Germany during the 1980's. The system became available commercially in the 1990's but sales were disappointing, and Siemens transferred the rights to METAL to GMS and LANT. The most famous systems based on the 'transfer' process were Ariane at GETA in Grenoble, France _a project that began in the 1960's - and Eurotra - a project funded by the Commission of the European Communities. Neither system met the desired expectations. In the late 1980's, Japanese Government agencies sponsored and cooperate with researchers from China, Thailand, Malaysia and Indonesia to study Interlingua' system of Asian languages. After more than a decade of effort, this system has not produced the desired results.

Our own experience with 'untrained' MT systems is that translators prefer to start 'from scratch' rather post-edit an automated translation. However, even 'untrained' MT systems are helpful in many ways. For example, Alta Vista Babel Fish makes it possible to get the idea of what a website in a language we do not know is about - we call this 'gusting'.

③ knowledge-based approach

A transfer system takes on much of the flavor of an Interlingua system while not making the commitment in linguistics universality that may be seen as the hallmark of the Interlingua approach. Such semantic transfer systems are attracting quite a lot of attention. This can also be seen to some extent as a compromise between the mainly linguistically based approaches and the so-called KNOWLEDGE-BASED systems. Translation relies heavily on information and abilities that are not specifically linguistics. It is only because we take their common sense and knowledge of the everyday world for granted.

The only major new lines of investigation that have emerged in recent years have involved the use of existing translations as a prime source of information for the production new ones. EXAMPLE-BASED MT, a system of otherwise fairly conventional design, is able to refer to a collection of existing translations. A much more radical approach, championed by IBM, is the one

in which virtually the entire body of knowledge that the system uses is acquired automatically from statistical properties of a very large body of existing translation.

2.6 Analysis of the methods

After looking back the history of MT and methodology, let us analysis of the methods.

2.6.1 Weave's mistakes in the design of methodology of MT

In the early stage of MT development, Weaver apparently distorted the process of Interlingua transfer. In his mind, it is possible to simplify translation process to the simple, linear substitution of SL words or words combination by TL words or words combinations". Weaver wrote as follows

"One naturally wonders if the problem of translation could conceivably be treated as problem in cryptography. When I look at an article in Russian, I say: this is really written in English, but it has been coded in some strange symbols. I will now proceed to decode it ".

So, MT is no more than a design to make the Interlingua transfer process, it is a combination of SL-decoding and TL encoding operations just like the sending and receiving of telegram. According to Weaver's plan, translation can be carried out by a something like a database processing machine. From this, I think Weaver's idea was overly influenced by his communication theory and, At the same time neglected linguistic and translation theory; besides, he was not a professional Translator. In other words, the fundamental object of MT is to ascertain and to clearly illustrate the Interlingua transfer process, so that the computer and software can simulate transfer operations. Then, the engine or algorithmic principles will seem to carry out the task of simulation.

But, we have had a clear idea that translation will never be as simple as word substitution. So, the idea of simulating human translation in the form of Interlingua transfer processes is more easily said than done, especially when the research of artificial intelligence (AI) still remain underdevelopment. Restricted by low AI, the optimal solution is, of course, not yet well understood. So, it is not at all surprising that MT has been unable to live up to expectations, the attempt of perfect simulation ended in vain.

2.6.2 Development of MT methodology

As for MT methodology, it has experienced the transition from first generation to second and Third, each generation of MT methodology has its own character and focus, such as first with its focus on lexical orientation; second with its focus on purely morpheme-syntax orientation, the third with its focus on syntax-semantic approach. Generally, the object of research was lexically wider,

from lexical orientation to morpheme-syntax; to syntax-semantic, approach was more complex. gradually, researchers have been aware of the complexity of translation process, far from being simple like the description of decoding and encoding.

Experienced human translators use their extensive knowledge, sometimes subconsciously, too resolve syntactical semantic and cultural ambiguities, which are beyond computers computing power. In the process of MT system, ambiguity has to be left unresolved, or resolve by some rigid rules, which will often result in a wrong even unintelligible translation. From these ambiguities, researchers realized that Interlingua structural difference {syntactic barrier) has a more important role than word-for-word substitution. In order to overcome the syntactic barrier, the second generation of MT research appears on the stage, with the focus on the establishment of grammatical rule systems for the surface structure analysis of texts. However, after several years' quest for grammatical rule -based models, the researchers found that satisfactory results were hard to obtain without syntactic disambiguation programs and semantic information.

Till now, researchers have realized that one can translate a text only if one has understood it beforehand. Translation performance requires a certain degree of intelligence, so dose MT system. but the computer has been lacked of intelligence of any kind. If MT wants to make a big breakthrough, it is necessary to make some kind of artificial intelligence integrate into MT systems. but, don't forget that the development of MT can't go alone; it has to follow the pace of computer technology. When computer expert has not resolved the difficulty of artificial intelligence, then it will be difficult for MT to make a huge leap.

To understand a text, a translator should be provided with three types of knowledge, real work knowledge; situational knowledge, and text-internal knowledge. This conclusion provides explanation of the fact that in many texts, more is meant in simple words than actually said all.

But, up till now, the MT system has no integrated model of reality knowledge. Its task is only transferring text meaning rather than understanding text meaning. Lacking of creativity is MT System's biggest defect; computer can practice of a text-internal approach by using logically explicit instructions. The human translator, on the other hand, in going form SL to TL, has the possibility to combine a text-internal with a text-external knowledge, to carry out a perfect translation process. Therefore, if without integrated AI model, MT can at best approximate the processes used by humans in comprehending and translating language.

When researchers direct their focus on syntactic and semantic approach, methodology of MT has come to third generation. At a time, various attempts were made to design MT procedures, which combine syntactic and semantic approach? At that time, Chomsky's concept of syntactic deep structure has attracted attention. According to the Chomsky linguistic theory, the Deep-structural level languages are more similar than on the surface-structural level. This opened

the vista for the development of MT systems, which are no longer, limited to a specific language pair, but could also be applied to other language pairs. But two sentences with different deep-structure may have similar surface -structure, like 'it is a shame he never wins' and 'that he never wins is a shame'. Vice versa, such as 'it is a shame to know you' and 'it is a shame to drink you'. The diversity of syntax structure brought MT methodology again into corner.

Recently, a fourth generation of MT has been under discussion. The fourth MT generation consists of ideas to build MT systems on the basis of artificial intelligence. What is the definition of artificial intelligence? Simply put, it is the attempt to emulate human patterns of thinking and behavior by using computer models. This approach is deeply influenced by the fact that the expressions of a natural language cannot be determined or decided by any set of rules or storage of information that former three generations implicitly supposed can be done.

2.6.3 The-AI's role in the translating and an example

In order to have a better understanding of AI, let's take a comparison with what humans do with, for instance, calculation. There are actually 3 ways (and maybe more) of performing calculation:

1) When asked the result for $145+133$, we actually break down the calculation into smaller ones, perform the necessary calculation and give the answer.

2) When asked the result for 8×5 , we immediately respond with recalling a multiplier table that we learned at primary school.

3) When asked whether $1,450,000 \times 3,789$ is greater or smaller than 1, we give the answer immediately "greater," although we do not actually perform the calculation (a computer will not respond subconsciously-it will calculate first, then give a final answer, though the speed is the same with human).

Scientists can confirm that computers perform methods 1 and 2, with considerable speed but not step 3. We may say that method 3 is nice, and subconscious, something unique to human, which is object that computer try to emulate. Now, the IT engineers' task are concerned with what computers will do like the third method. If computer scientists succeed in attaining the target in the future, we will then see perfect MT systems in our computer.

Apart from conceived artificial intelligence-oriented MT research, MT projects can presently be classified either under second-generation or third generation MT research. Until the end of the 1980x, the dominant framework of MT research was based on linguistic rules of various kinds: rules for syntactic analysis, lexical rules, rules for lexical transfer, rules for syntactic generation, rules for morphology, etc. till now, the rule-based approach was most obvious in the most popular

transfer systems in the market.

Rules-based MT system focuses on the analysis, judgment, and choice to the SLs. In these systems, texts of the SL are firstly analyzed into abstract representations of meaning; secondly computer programs (rules) for identifying word structure (morphology) and sentence structure (syntax) and for resolving problems of ambiguity (semantics). After the rearrangement of the word order as the previous rules, a computer produces the TL.

2.7 Types of MT

When giving any general overview of the development and use of MT (MT) systems and translation tools, it is important to distinguish four basic types of translation demand. The first, and traditional one, is the demand for translations of a quality normally expected from human translators, i.e. translations of publishable quality – whether actually printed and sold, or whether distributed internally within a company or organization. The second basic demand is for translations at a somewhat lower level of quality (and particularly in style), which are intended for users who want to find out the essential content of a particular document – and generally, as quickly as possible. The third type of demand is that for translation between participants in one-to-one communication (telephone or written correspondence) or of an unscripted presentation (e.g. diplomatic exchanges.) The fourth area of application is for translation within multilingual systems of information retrieval, information extraction, database access, etc.

The first type of demand illustrates the use of MT for dissemination. It has been satisfied, to some extent, by MT systems ever since they were first developed in the 1960s. However, MT systems produce output, which must invariably be revised or ‘post-edited’ by human translators if it is to reach the quality required. Sometimes such revision may be substantial, so that in effect the MT system is producing a ‘draft’ translation. As an alternative, the input text may be regularized (or ‘controlled’ in vocabulary and sentence structure) so that the MT system produces few errors, which have to be corrected. Some MT systems have, however, been developed to deal with a very narrow range of text content and language style, and these may require little or no preparation or revision of texts.

In recent years, the use of MT systems for dissemination purposes has been augmented by developments in translation tools (e.g. terminology databases and translation memories), integrated in authoring and publishing processes. These ‘translation workstations’ are more attractive to human translators. Whereas, with MT systems translators see themselves as subordinate to the machine, in so far as they edit, correct or re-translate the output from a computer, with translation workstations (or workbenches) the translators are in control of computer-based

facilities, which they can accept or reject as they wish.

The second type of demand – the use of MT for assimilation – has been met in the past as, in effect, a by-product of systems designed originally for the dissemination application. Since MT systems did not (and still cannot) produce high quality translations, some users have found that they can extract what they needed to know from the unedited output. They would rather have some translation, however poor, than no translation at all. With the coming of cheaper PC-based systems on the market, this type of use has grown rapidly and substantially.

With the third type – MT for interchange – the situation is changing quickly. The demand for translations of electronic texts on the Internet, such as Web pages, electronic mail and even electronic ‘chat’ lists, is developing rapidly. In this context, the possibility of human translation is out of the question. The need is for immediate translation in order to convey the basic content of messages, however poor the input. MT systems are finding a ‘natural’ role, since they can operate virtually or in fact in real-time and on-line and there has been little objection to the inevitable poor quality. Another context for MT in personal interchange is the focus of much research. This is the development of systems for spoken language translation, e.g. in telephone conversations and in business negotiations. The problems of integrating speech recognition and automatic translation are obviously formidable, but progress is nevertheless being made. In the future – still distant, perhaps – we may expect on-line MT systems for the translation of speech in highly restricted domains.

The fourth type of MT application – as components of information access systems – is the integration of translation software into: (i) systems for the search and retrieval of full texts of documents from databasebases (generally electronic versions of journal articles in science, medicine and technology), or for the retrieval of bibliographic information; (ii) systems for extracting information (e.g. product details) from texts, in particular from newspaper reports; (iii) systems for summarizing texts; and (iv) systems for interrogating non-textual databasebases. This field is the focus of a number of projects in Europe at the present time, which has the aim of widening access for all members of the European Union to sources of database and information whatever the SL.

2.8 Evaluate MT

While there is general agreement about the basic features of MT (MT) evaluation, there are no universally accepted and reliable methods and measures, and evaluation methodology has been the subject of much discussion in recent years

As in other areas of NLP, three types of evaluation are recognized: adequacy evaluation to determine the fitness of MT systems within a specified operational context; diagnostic evaluation

to identify limitations, errors and deficiencies, which may be corrected or improved (by the research team or by the developers); and performance evaluation to assess stages of system development or different technical implementations. Adequacy evaluation is typically performed by potential users and/or purchasers of systems (individuals, companies, or agencies); diagnostic evaluation is the concern mainly of researchers and developers; and performance evaluation may be undertaken by either researchers/developers or by potential users. In the case of production systems there are also assessments of marketability undertaken by or for MT system vendors.

MT evaluations typically include features not present in evaluations of other NLP systems: the quality of the raw (unedited) translations, e.g., intelligibility, accuracy, fidelity, appropriateness of style/register; the usability of facilities for creating and updating dictionaries, for post-editing texts, for controlling input language, for customization of documents, etc.; the extendibility to new language pairs and/or new subject domains; and cost-benefit comparisons with human translation performance. Adequacy evaluations by potential purchasers usually include the testing of systems with sets of typical documents. But these are necessarily restricted to specific domains, and for diagnostic and performance evaluation there is a need for more generally applicable and objective test suites; these are now under development.

Initially, MT evaluation was seen primarily in terms of comparisons of unedited MT output quality and human translations, e.g., the ALPAC evaluations and those of the original Logos system. Later, systems were assessed for quality of output and usefulness in operational contexts, e.g., the influential evaluations of Systran by the European Commission. Subsequently, many potential purchasers have conducted their own comparative evaluations of systems, often unpublished, and often without the benefit of previous evaluations. Valuable contributions to MT evaluation methodology have been made by in her study for the European Commission, and by the JEIDA committee which proposed evaluation tools for both system developers and potential users---described in more detail in section .The evaluation exercise by ARPA compared the unedited output of the three ARPA-supported experimental systems (Pangloss, Candied, Ling stat) with the output from 13 production systems from Globalink, PC-Translator, Microtac, Pivot, PAHO, Metal, Socatra XLT, Systran, and Winger. The initial intention to measure the productivity of systems for potential users was abandoned because it introduced too many variables. Evaluation, therefore, has concentrated on the performance of the core MT engines of systems, in comparison with human translations, using measures of adequacy (how well a text fragment conveys the information of the source), fluency (whether the output reads like good English, irrespective of accuracy), and comprehension or in formativeness (using SAT-like multiple choice tests covering the whole text).

With the rapid growth in sales of MT software and the increasing availability of MT services

over networks there is an urgent need for MT researchers, developers and vendors to agree and implement objective, reliable and publicly acceptable benchmarks, standards and evaluation metrics.

2.9 MT on market

From a global perspective, the translated market already takes shape, and it has the specialized market in the strong autonomy degree. According to a Intelligence Corporation's investigation of American Allied Business, including the artificial translation, the MT as well as the software and the website localization translation, market scale estimated will rise from 13 billion dollars in 2000 to 227 billion US dollars in 2005. Moreover, looked from the function, globalization and the localization is at the same time, whether the translation profession develops perfectly, whether the translation serve consummates, may become weighs to judge society construct whether it can be elastic and potential size of the development

Chapter 3 Web pages Translation and software translation

3.1 Basic information of web pages translation

Web pages are documents on the World Wide Web. Every Web page is identified by a unique URL (Uniform Resource Locator).

3.1.1 What is Web Translation

With the rapid development of AT (Automatic translation), some web-based AT (WBAT) programs appear on the Internet to meet the needs of translating a great number of web pages from one language into another.

Language is the fabric of the web. The rapid growth of the Internet/WWW and the emergence of the information society pose exciting new challenges to language technology. Although the new media combine text, graphics, sound and movies, the whole world of multimedia information can only be structured, indexed and navigated through language. For browsing, navigating, filtering and processing the information on the web, we need translation that can get general meaning at the contents of documents at first. Language technology for content is a necessary precondition for turning the wealth of digital information into collective knowledge. The increasing multilingualism of the web constitutes an additional challenge for our disciplines. The global web can only be mastered with the help of multilingual tools for indexing and navigating. Systems for information and knowledge translation will surmount language barriers for e-commerce, education, technology and international cooperation.

3.1.2 Types of web pages translation

There are four translation ways in the net.

1) Web page translation: Users type in the URL of a web page and will get the translated version of the web page in a TL.

Al-Alamiyah's Sikh company announces the availability of a Web-site to translate Web pages automatically from English into Arabic. The announcement specifies that the Web-site uses Sakhr's own MT System. It also specified that the result is just a draft quality. Translated version is not perfect. However, it may enable the user to decide whether a more 'professional' translation needed

or not.

2) Text translation: Users type in any text (from one word to a paragraph), submit the text and will get the translation.

3) File translation: Users upload a created file, submit it to the server and will get the translated version.

4) E-mail message translation: Users send an e-mail message to a server and will receive the translated message in their e-mail account

T-mail (<http://wwwt-mail.com>) is a system through which a user can send an E-mail message to another user. The system would take the E-mail message in the SL and translate it to the TL specified by the sender. Both the SL E-mail and the TL E-mail are mailed to the receiver specified by the user. At present, T-mail translates between English and Spanish, French, Italian, Portuguese and German. T-mail is advertised as a free E-mail translation service designed to be an aid to communication. It is not a precise translation service. As to the translation of the whole web page, T-mail does not fit for the job yet

3.1.3. How to Translate Web Pages

WorldLingo (http://www.worldlingo.com/zh/microsoft/computer_translation.html). It provides a free service that translates web pages (and other text) in many different translation pairs.

▼ To use them, go to http://www.worldlingo.com/zh/microsoft/computer_translation.html



- ▼ put in the URL of the page or the text you wish to translate (www.whitehouse.org)
- ▼ choose the language you want(from English to simple Chinese)
- ▼ click on the Translate button.

You can get the web page as the picture showed.



Other translations were also convenient, with Microsoft's consistent style. For instance, when using email translation service, we are also deferring to the ordinary the mail-written method, the form and so on. The different of e_mail translation is needs to choose the receiver's language as well as addressee's language.

电子邮件翻译

发件人信息:

姓名:

电子邮件:

语言:

[Text input field]

[Text input field]

English [v]

记住 ☐

收件人信息:

姓名:

电子邮件:

语言:

[Text input field]

[Text input field]

English [v]

记住 ☐

主题:

邮件:

[Text input field]

在此处键入或粘贴您的邮件

给我发送一个副本 ☒

翻译并发送电子邮件

3.2 The times thirst for web pages translation

At the time of September 1999, there were over an estimated 800 million documents on the World Wide Web (WWW) and the amount is increasing in a very fast speed. This means if we want to translate those documents on web pages by human translator, it will not an impossible mission, because the amount of information is too huge even at that time. Even if we limit the translation of the Web pages to of people access (rather than translate all the documents on the Internet), the cost of translation by humans is still very high. Many companies make advertisement of asking human to translate Web Pages on the Internet. The typical price is around \$0.19 per translated English word which makes the average cost per page around \$80.

Even if we do not care about the cost, Internet users usually want to receive the documents within a very short time after they click for them. It is reported that because of delays only one minute, the user may give up the request for the Web Page. Human translation is therefore not appropriate since it will take hours for a good human translator to concentrate all his attention in order to be able produce a translated Web Page that corresponds exactly to the source the user requested, even though some companies on the Internet have offered to perform Human translation within a minimum period of 24 hours. It is not a practical choice to ask human translators to perform such task because of the large amount of information available on the Internet, the growing race is great and people require the speed. Another reason for the impossibility of human translation of the Web pages is the manic nature of Web pages since a page is changed over and over again it may survive in a very short time. Many Web pages have been under construction for a long period before being checked.(Hutchins, 1999).

Definitely we cannot ask English people to publish information on the Internet in Chinese in order to make it accessible to Chinese people. It is the responsibility of our Chinese speaking readers to develop tools that will make the Internet more Chinese-friendly. Thus, an important tool to develop is a system to translate the web pages that are available on the Internet into Chinese. At the same time, it is also very important that these tools should also be developed to be able to translate pages from Chinese into other languages.

Today, in order to tackle the situation of the dominance of the English language on the Internet, we have one choice, i.e. to use computer technology to develop an English-Chinese MT system. With the advancements in MT technology and tools such as morphological analyzers and parsers, on-line language translation would definitely be a useful, and maybe an essential tool. If we could transfer the English written WebPages into Chinese, there will surely be a large more amount of Chinese people who can touch and receive more updated information from the Internet.

3.3 Present situation of Chinese Web Translation

With the expansion of global telecommunications (the Internet and World Wide Web has come the networking of translation services. Now they offer their systems as a service to individual or company customers. Texts can be sent on-line for immediate 'rough' translation with no post-editing, or for treatment in a more traditional manner with expert revision, editing and preparation for publication by the service. This form of networked MT is clearly a further development of familiar translation services. It is assumed that in future there will emerge various forms of networked 'translation brokerage' services which will advise customers on the most appropriate

MT service for their needs, e.g. in terms of costs, languages, speed, dictionary coverage, terminology control, overall translation quality, post-editing support, etc. Some of these 'translation brokers' may advocate themselves be automated, and undertake searches of the Web for particular client needs. As a consequence, we may well see the emergence of more specialized MT systems (for particular domains and language pairs), some of which will thrive and others which will fail in the global competitive market.

Even more significant for the future, however, is the appearance of systems for on-line and real-time translation of electronic mail messages.

As to MT to Chinese, many translation systems have already been programmed by some companies as a commercial use in China, like Transtar, Jinshan2001, and Yaxin CAT, etc. These systems use the direct approach to translate Web pages from English into Chinese. It divides the English sentences into different parts according to where an HTML tag occurs. Then it translates the part of the English sentence independently of others and inserts the translation between the HTML tags that are present in the source. These companies have an ambition to translate any Web page that can be obtained on the Internet. However, it still has many problems when some marked words merge with other words or are translated into more than one word

In September 1999, Language Force, a USA-based company, announced the new 'Language Force Universal Translator 2000' that instantly translates Web pages-mail and even spoken language in 40 different languages (<http://www.bgtv.com/translator/software/software.html>). This kind of software, to some extent, is capable for meeting some people's certain needs. However, it still has a long way to go as for a practically acceptable translation. Reverso Pro 5 is a product introduced by Softissimo to translate Web pages from English into Spanish, German and French. It uses the company's MT engine to translate any text within software that supports the drag and drop feature. It is thought by the company as a product that will "translate directly your text inside Internet Explorer or Netscape Navigator while keeping the structure of the source page and

the images and links on the source page"(http://www.fortissimo.com).

Also there is other software or systems that can translate web pages into Chinese to a certain degree, including IBM alpha Works and IBM developer Works, WorldLingo and some software in China as I have listed before. I here will not describe them any longer.

A system, which could be adapted to translating Web pages, is that with Corpus-based, Accurate Natural-language Translation. It is a set of software tools for automatic and interactive analysis of source text and generation of target text. This system is designed to take advantage of document markup such as Hyper Text Markup Language (HML) rather than stripping out the HTML codes or tags, and trying to re-insert the tags after translation. It analyses the tags directly as an integral art of the input text, and represents them in the transfer process, so that they can be properly generated in the correct position by the target generator (Fleming and Cohen, 2000).

3.4 Selected web wits of Web-based automatic translation services

AltaVista Babelfish by Altavista.com

URL: <http://babelfish.altavista.com/>

Languages:

English => Chinese, French, German, Italian, Korean, Portuguese, and Spanish

Chinese, French, German, Italian, Japanese, Korean, Portuguese, Spanish => English

French => German

German => French

Function: web page and text translation

Systranet by Systranet.com

URL: <http://www.systranet.com/systran>

Languages:

English=> French, German, Italian, Portuguese, Spanish, and Greek

French, German, Italian, Portuguese, Spanish, Greek => English

French => Italian, Portuguese, Dutch, Spanish, and Greek

Greek=>English, French

Function: web page, file (uploaded .txt, .rtf, or .html documents) and plain text translation

Other features: Dictionary of personal choice, my terms (choose how to translate, modify or add the terms of your choice)

Aiciba by kingsoft

URL: <http://www.iciba.com>

每日一练: [I wish I could 但愿我能](#). [4 28] [更多»](#)



词典

短句

翻译

豆豆

沙龙

更多

搜索

Languages: Chinese => English

Merit: its function mainly manifests in the phrase translation, and the short phrase translation, when filling the translation sentences or phase into the frame, translation will be completed. Compares to other translation, it has the superiority at speed and Chinese phrase' assurance. It is suitable to domestic user. Moreover also has the characteristic function of "every day sentence", regarding to some abc learners.

Function: phase translation



爱词霸

· 短句

[词典](#) [短句](#) [翻译](#) [豆豆](#) [沙龙](#) [更多»](#)

语法分析

搜索

语法分析

grammatical analysis

语法分析程序

syntactic analyser

不易进行语法分析的句子

sentences that do not parse easily.

矢量分析,向量解析

vector analysis

通过主观考试对他的才能进行分析。

used projective tests to evaluate her aptitudes.

尿分析

urinalysis.

无法分析的数据

unanalyzable data.

【化】元素分析,最终分析

ultimate analysis

转换语法

HTF by Shanghai Huitian Fu.com:

URL: <http://www.zhongguofanyi.com/about-us.asp>Languages:

English => Spanish, French, German, Norwegian, and Portuguese

Spanish, French, German, Portuguese => English

Function: web page and text translation

Other features: linked with web dictionaries



中国专家翻译网 (<http://www.chinatranslation.net/>)

Languages:

English, French, German,, Italian, Japanese, Portuguese, Russian, Slovenian, Spanish, Chinese
simplified ,Chinese traditional ,Korea, Dutch

Function: text translation

Other features: Provides stipendiary services



中国专家翻译网

华译网翻译公司

架起中外沟通的桥梁

[首页](#)
[公司简介](#)
[业务范围](#)
[收费标准](#)
[质量控制](#)
[公司业绩](#)
[诚聘英才](#)
[本地化](#)
[翻译论坛](#)
[人才培训](#)
[联系方式](#)

[简体中文](#)
[English](#)
[日文](#)
[德文](#)
[Arabic](#)
[法文](#)
[韩文](#)
[俄文](#)
[Search](#)
[繁体中文](#)

北京华译网翻译公司

最新动态



清华大学 著名经济学家 魏杰教授 为我司译著《开普斯链商务百科》作序

华译网翻译引进巨著《开普斯链商务百科》

国家工商总局领导为我司译著题词

社科院工经所周叔莲所长高度评价

我司新近引进哈佛《创业管理》等

国家人事部发布《关于2005年度二

全国翻译专业资格(水平)考试和

全国翻译专业资格考试四个专家委

全国日语翻译证书考试即将推出

从董事长到穷学生 1500名“高级义

阿语翻译成为中国回族青年的热门

更多-->>

国内、外主要客户

客户评价

2001年部分翻译项目

2002年部分翻译项目

2003年部分翻译项目

最近翻译的部分图书

网站本地化样稿

翻译格式要求

翻译质量要求和翻译要领

多专业翻译

多语种翻译

网站地图

友情链接

在线翻译

输入要翻译的内容(短文翻译只限5K以内)...

原文

主题: 通用

英语

汉语(简体)

翻译

Golden Beijing Translation Port 金桥译港世界通 by Netat.net

URL:

Chinese version: <http://www.netat.net/file.htm>

English version: <http://www.netat.net/english/index.htm>

Languages:

Chinese => English, Japanese

English, Japanese => Chinese

Function: web page, text, file (txt, html and rtf files) and e-mail translation (Send email to mailce@netat.net or mailec@netat.net, the translated version will be sent back to the sender.)

Other features: Chat in your preferred language. The messages will be displayed in both languages (Chinese and English). <http://www.netat.net/esalonchat.htm>

Readworld 看世界 by Readworld.COM

URL: <http://www.readworld.com/tran/>

Languages:

English => simplified or traditional Chinese

Function: web page, text and e-mail translation (send English e-mail messages to standard@mailtran.readworld.com and the e-mail message will be translated and sent back to the user)

T-Mail by T-Mail.com

URL: <http://www.t-mail.com/>

Languages:

English => Simplified or traditional Chinese, French, German, Italian, Japanese, Norwegian, Portuguese, Russian, Spanish

French => English, German, Italian, and Russian

German => English, French, Russian

Italian => English, French, German, Russian, Spanish

Portuguese => German

Russian => English

Spanish => English, Italian

Function: e-mail translation (cc e-mail messages to the server, indicate the source and TLs, the translated version will be forwarded to the receiver)

T-Text by T-Mail.com

URL: <http://www.t-mail.com/t-text.shtml>

Languages: Same as T-Mail

Function: posted text translation

TranSea 译海 by Altlan Company

URL: <http://en.transea.com/>

Languages:

English => Chinese

Chinese => English

Function: web page, text and file (txt, html and rtf documents) translation

Transtar 译星 by transtar.com.cn

URL: <http://www.transtar.com.cn/transtar/english/edefault.asp>

Languages:

Simplified Chinese => English

English => simplified or traditional Chinese

Function: web page and text translation

*The web-based automatic translation services of the above sites are still free at the time of writing the paper.

3.5 The market of web pages translation

The potential market for network MT systems is enormous. At CompuServe alone there are more than 3,000 other on-line services where MT could be introduced; and other Internet services could easily follow their lead. It has been estimated that there are currently over 40 million electronic mail messages a month. If only a small fraction of these were candidates for translation, the demand would be enormous.

In addition to electronic messages, the amount of information available in text form on Web pages can now counted in their hundreds of millions, and they are growing exponentially at a high rate (10% between 1995 and 1996). The non-English content is estimated as 80% of the total, and there is no doubt that readers everywhere prefer to have text in their own language, no matter how flawed and error-ridden it may be, rather than to struggle to understand a foreign language text. The Japanese software companies have already recognized the huge potential market and there are a number of English-Japanese translation modules available for integration with Web software. Similar Web translation software is being developed and sold for other languages, both by existing vendors of MT systems and by new companies.

A further factor will be the growth of multilingual access to information sources. Increasingly, the expectation of users is that on-line databebases should be searchable in their own language, that the information should be translated and summarized into their own language.

The European Union is placing considerable emphasis on the development of tools for information access for all members of the community. Translation components are obviously

essential components of such tools; they will be developed not as independent stand-alone modules, but fully integrated with the access software for the specific domains of databases. The use of MT in this wider context is clearly due for rapid development in the near future.

There is no gainsaying the enormous potential for the translation of electronic messages. Only a fully automatic process capable of handling large volumes with close to real-time turnaround can provide the translation capacity required by on-line markets. In addition, the online 'culture' favors rapid and 'shallow' assimilation of information; for these reasons, MT is the obvious future. It is now evident that the true niche market for MT is in 'cyberspace'. While poor quality output is not acceptable to translators, it is acceptable to most of the rest of the population. How long it will be acceptable is an open question; inevitably there will be expectations of improvement, and a challenge for the MT community must be the development of translation systems designed specifically for the needs of the Internet.

3.6 Software translation

There is a popular test to 7 translation softwares: Kingsoft Ciba 2005, Kingsoft, Dadian 2005, Dongfang Dadian 2003, Dongfang Kuaiche2003, Babylon Pro 5.0.5, Dr.eye 2005, Wanneng Duiyi2005.

First, introduction of the test

1) Members: 7

- ① Jinshan Ciba 2005 (type: commercial software, price: 50 元)
- ② Jinshan Dadian 2005 (type: commercial software, price: 50 元)
- ③ Dongfang Dadian 2003 (type: commercial software)
- ④ Dongfang Kuaiche2003 (type: commercial software)
- ⑤ Babylon Pro 5.0.5 (type: free software)
- ⑥ Dr.eye 2005 (type: software)
- ⑦ Wanneng Duiyi2005 (type: software)

2)Test platform

CPU: P4 2.4G

Memory: KingMax 512MB

Hard disk: Maitui Jin zuan 9 80G

Operating system: Windows XP Professional 中文版+Service Pack 2

Second, test items

I comparison of Basic performance (20%)

1) system resources taking rate

系统要求	最低机配置	占用硬盘空间	占用系统资源
金山词霸 2005	Pentium 200、32M 内存	300MB (典型安装)	12%
金山快译 2005	Pentium 200、32M 内存	130MB (典型安装)	10%
东方大典 2003	Pentium 200、32M 内存	260MB (典型安装)	10%
东方快车 2000	Pentium 200、32M 内存	40MB (最小化安装)	14%
Babylon Pro 5.0.5	Pentium 166、64M 内存	200MB (完全安装)	7%
Dr.eye 译典通 2005	PIII 以上 CPU、128MB 内存	64MB 以上	7%
万能对译 2005	Pentium 200、32M 内存	170MB 以上	6%

After reading this form, we knew they do not request a high computer disposition, highest III and 128MB today weakest machine can use it. But the hard disk space using is different from 10MB to 300MB, which depend on the type of installment you choose. But in the system resources aspect, kingsoft and the Dongfang Kuaiche all need 10% .If your machine is not above Pentium III, the memory does not have 256MB, the speed will come across the certain influence. Dr.Eye 2005, which has a good performance in various aspects , should be an ideal choice.

2) Word stock capacity

软件名称	词库容量	能否扩充
金山词霸 2005	400 万词库, 包括 200 本词典辞书、80 个专业词库	在线更新
金山快译 2005	400 万词库, 包括 80 个专业词库	在线更新
东方大典 2003	420 万条词库, 包括 75 个专业词库	在线更新
东方快车 2003	420 万条词库, 包括 75 个专业词库	在线更新
Babylon Pro 5.0.5	60 余万条词库, 可额外安装词库包增加词库	安装词库包
Dr.eye 译典通 2005	300 万以上词库	在线更新
万能对译 2005	30 万条词库	在线更新

An important standard of translation softwares is its capacity size of words. If the word's capacity is too small, software's can not make the accurate translation. Therefore, usually speaking, bigger capacity is more greatly better. May see from table 2, Kingsoft and Dongfang Dadian both have a capacity of 4 million words, which is in the top class in software's, moreover they also could carry on the expansion through the on-line promotion.

3) Multilanguage support

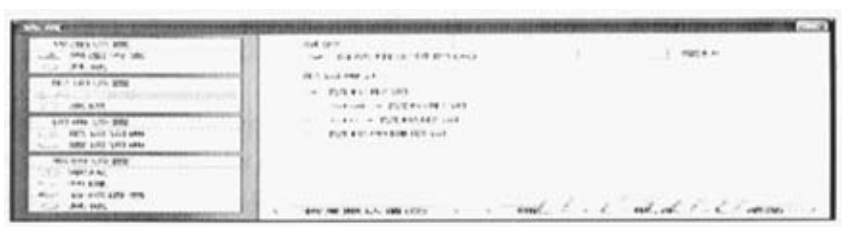
软件名称	语种支持
金山词霸 2005	英中/中英/日中
金山快译 2005	英中/中英/日中
东方大典 2003	英中/中英/日中
东方快车 2003	英中/中英/日中
Babylon Pro 5.0.5	根据安装的词库包可以实现英中/中英/中日/日中/中韩/韩中等多类型翻译
Dr.eye 译典通 2005	英中/日中
万能对译 2005	英中/中英/日中/中日/中韩/韩中/中俄

According to the table, nearly all translation software's have provided three kinds translation services of English-Chinese, Chinese-English and Japanese-Chinese. These modest functions are

enough to use.

II Word-took mode (10%)

1) word-taking to translation



Jinshan Kuaiyi and Dongfang Kuaiche can translate whole text , others provided the word-taking to translation by mouse on the screen. The way of translation facilitates clients greatly.

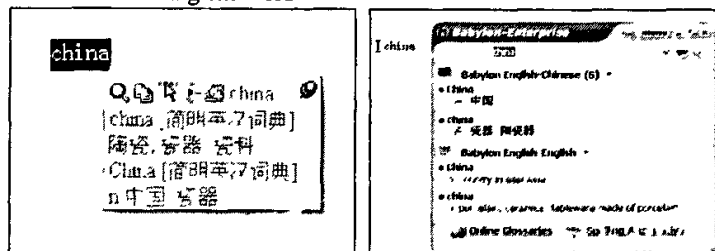
2) supporting by other kinds of documents

软件名称	文档格式支持
金山词霸 2005	HTML、PDF、DOC、TXT
金山快译 2005	——
东方大典 2003	HTML、PDF、DOC、TXT
东方快车 2003	——
Babylon Pro 5 0.5	HTML、PDF、DOC、TXT
Dr.eye 译典通 2005	HTML、PDF、DOC、TXT
万能对译 2005	只支持 TXT，其它程序可以对菜单部分进行翻译

Kingsoft Ciba,Dongdang Dadiandian,babylon,Dr.eye provided good supports to our commonly used documents ,including HTML , PDF ,DOC and TXT.

III quality for word translation (25%)

1.Test of translating one word



Kingsoft Ciba and Babylon can provide 2 accurate explanations of "China", but others do not

have the performance .Kingsoft provide the classified table to inquiry It also has divided into doze kinds of explanation of English word ,in which has the more exhaustive minor division.

2.Expansion of explanation

The test of explanation's expansion is mainly about manual-input word translating in the software, and the exam explanation in them.



Only Babylon 's content between explanation's expansion and screen-take translations the same. Other four translation softwares provide more information. particular, kingsoft Ciba, Dongfang Dadian and Dr.Eye have not only every word's explanation and the examples, but also kinds of applications. The way will be helpful to learners.

IV whole text translation quality (15%)

Original text:

Our term china for porcelain or ceramic ware is a shortening of chinaware and probably china dishes. Although the word china is identical in spelling to the name of the country, there are 16th- and 17th-century spellings like chiney, cheney, and cheney that reflect the borrowing into English of the Persian term for this porcelain, chn. The Persian word and the Sanskrit word cn, Chinese people, which gave us the English name for the country, go back to the Chinese word Qin, the name of the dynasty that ruled China from 221 to 206 BC



Because this section is quite specialized, strong, the difference among three translation softwares it not obvious. But after carefully analyzing, it is not difficult to find out kingsoft' superiority to translate in detail at some aspects. But it needs to point out, Kingsoft has special resources. If browsing through articles before translating, choosing the certain field of the content, the accurate rate will be improved.

V characteristic function (30%)

1、Kingsoft

a : A function of mini word-reciting . When running the software, prepared words will appear on the top of the screen.

b : Keep difficult words and make cards.

c: Rich in appendixes

d: Cards in offices



2、Kingsoft Kuaiyi

a: Web translation and localization of the other softwares

b: When translating, speed or quality can be chosen according to the need

c: Connect with kingsoft Ciba

3、Dongfang Dadian

a: Mouse-drew to translation windows

b: Search relevant phrases

c: Inlaid offices

4、Dongfang Kuaiche

A: Permanent localization

b: Connect with IE

C: Multi- languages search with google

d: Encoding conversion to support GB、BIG5、HZ、GBK simple、KSC、GBK traditional、Euc_JIS、New_JIS、Unicode

5、Babylon

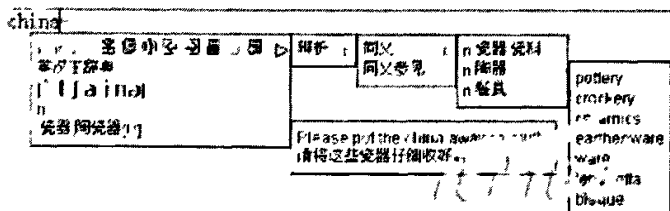
a: Monetary unit transform and update



b: Multi- languages transform

6、Dr.eye

a: Screen -taking word and button of “More Results “ links more usages.

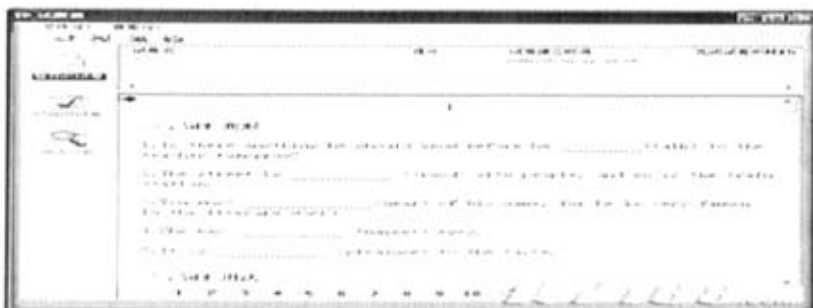


b: Automatic diagnoses of Multi- languages

c: Japanese input without Japanese operating system and Word processor

d: Promoting every week

e: Testing



7、Wanneng Duiyi

a: Personal corpus

(6) finally evaluates

Every software has advantages and shortcomings .In order to get a direct-viewing result, we give graded comparison showed as followed:

软件名称	金山 词霸 2005	金山 快译 2005	东方 大典 2003	东方 快车 2003	Babylon Pro 5.0.5	Dr.eye.译 典通 2005	万能 对译 2005
基本性能比较 (20%)	17	17	16	16	16	17	17
翻译方式(15%)	15	----	15	----	13	14	10
单词翻译质量 (20%)	20.	----	20	----	15	17.	16.
文本翻译质量 (15%)	----	14	----	14.	----	13.	----
特色功能(30%)	28	27.	26	27	15	27	10
合计	80	58	77	57	59	88	53

Generally speaking, Dr.Eye has obtained a good result through the word translation and the full text translation But for a single function, domestically produced software got the advantages,

How to choose suitable softwares? There are some recommendations:

- 1) To users who focus on the word translation : Kingsoft Ciba2005;
- 2) To users who focus on the whole text translation : Kingsodt Ciba,Kingsofe Kuaiyi;
- 3) To users who focus on the word translation and whole text translation : Kingsoft Ciba2005+Dr.eye

3.7 Summary

Translation is a science and an art. It is also one kind of a creative job. Because of its precise, complex, tedious and arid, people proposed the MT. Although many people pay the massive care and research, there is not a practical system until now. This has further proven the language's complexity. Main reason why web translation and software translation can't receive a satisfied quality is the limitation of the artificial intelligence theory and the algorithm .It is not realistic to complete the creative job only by web translation and software translation.

Chapter 4 CAT (computer assistance translation)

Dongfang Kuaiche and Kingsoft Kuaiyi can translate the whole text, but both of their translations are lack of readability and accuracy. The principle of the software translation is to regularize grammars and applies the massive types of sentences mechanically. The way of translations just piled on words and expressions. Although this kind of software translation may help the non-specialized people to glance general idea, it is difficult to satisfy the specialized translators, as well as enterprises, the scientific researches and the governments.

The development and popularization of computer assistance translation (CAT) become a trend. CAT does not rely on the automatic MT. It makes translators engage in and complete the translations. With TM (Translation Memory), the user's information corpus can be established. The corpus does not need to duplicate the work of the translations. CAT enhances the translation efficiency, the accuracy and the consistence of terminology. In this chapter, author will introduce famous CAT instruments, Germany Trados and Chinese "Yaxin", and their works principles and applies. Definitely, the more efficiency and accurate translations of Sci-tech document may be carried on.

4.1 Characters of TM technology

The Sci-tech document translation faces specialized fields, covers the narrow scopes. Therefore, the translation meets the repetition at the varying degree. How to use completed parts of translation, enhances the working efficiency and save cost are urgent problems to solve. TM may eliminate the repetition and guarantee translations' quality.

TM technology is a kind of translation assistant tool, which can save all the user's translations materials in translations memory corpus. In the processing of translation, the system splits out the same or the similar translations resources from the translation corpus, and produces the reference translations automatically. What should translators do is only concentrating the new content. This is one characteristic.

The other memory characteristic of TM technology is that memory corpus can edit, study and automatically save new translation parts, accumulation and consummation of the corpus are finished automatically when the translator is translating or in the revision.

4.1.1 The advent of TM system

Translation technologies flourish in the 1980's when computational equipments and resources became more widely available. At that time, large translation agencies began using specialized, automated tools to help their translators work more efficiently and consistently. These tools assisted translators on storing previously-translated materials, provided database of specialized terminology, and allowed translators to access the information to subsequent jobs. These specialized tools marked a different direction in the development of translation technologies: far from automating the translation process, these technologies were intended to assist human translators, allowed translators to work more efficiently, with less repetition and more consistency, translation memory systems became popular.

Translation memory system is usually integrated into a word processing program .translators translate, reuse former information and choose suggested segments conveniently.

4.1.2 What's significant of MT

What's significant of MT? Translation memory systems can be particularly suitable for translators to apply in highly-technical documentation, large documentation sets, multiple related documents, and documents with specialized vocabularies. Using this technology, translators can offer highly consistent, efficient, and cost-effective translations:

First, translation memory systems can ensure that the translated documents are consistent, including common definitions, phrasings, terminology, and so on. This is particularly valuable when different translators are working on a single project and when translation jobs span many different documents. The database of translations and terminology helps promote consistency within a documentation set, as well as consistency in subsequent translation projects.

Second, translation memory systems can help speed up the overall translation process. Because these systems remember previously-translated material, they eliminate the need for re-translating boilerplate text, instructions that are identical across different documentation series, glossaries of terms common to an industry or documentation set, or warnings and cautions present throughout documentation. The translator can translate the material one time, then reuse or tweak it at subsequent instances.

Third, translation memory systems not only make the translation process consistent and efficient, but also reduce cost for needs of long-term translation. For example, a documentation , a translator need translate 100% for the first time, but only 80% of the next time, 75% of the next, and so on. Likewise, a warning message could offer one time and reuse throughout an entire

documentation or series of documents. Additionally, building a database of terminology and phrasing specific for products, companies, or an industry, the translation memory system makes subsequent translation jobs even more economic because a portion of the material is already translated.

4.1.3 Function of TM

<http://www.trados.com/> TRADOS, Translators Workbench
<http://www.star-ag.ch/> STAR, Transit and Term-Star
<http://www.corel.ie/catalyst/index.htm> Corel, Catalyst
<http://www.sdintl.com/> SDL, Sdlx
<http://www.qsoft.de/ibmtrans.htm> IBM, Translation Manager
<http://www.mendez.com/english/default.asp/> Lernout & Hauspie MT Systems
<http://www.globalink.com/> Globalink
<http://www.qsoft.de/ibmtrans.htm> IBM Translation Manager
<http://www.systransoft.com/> Systran
<http://www.logos-ca.com/> Logos MT Demos
<http://babelfish.altavista.com/cgi-bin/translate> or <http://www.systransoft.com/>

There are numerous TM tools available on the international web. All of them occupy two main functions: preservation segmentation of source text, preservation of the original typesetting and layout.

Preservation segmentation means that a translator does not need to strain his/her eyes and brains looking for the source sentence translated in the beginning of the text. The software offers you this segmentation automatically using a table or a highlighted colored background (as Trados does). Thanks to this simple technology, you will never skip a sentence or even a paragraph in translating hard copies or overwriting electronic documents.

Translation memory means that you never have to translate the same sentence, phrase or word twice. You do it only once, and when you come across the same sentence, phrase or word next time, a program suggests your previous translation, which can be edited according to the new context. This is of importance primarily, but not exclusively for technical translators, because Sci-tech documents tend to have plenty of repetitions, for example, when working with Volvo Cars International, about 60% of the sentences in the brochure on the Volvo S60 were sentences had already translated working on the Volvo S/V70, S80, S/V40 brochures. In such cases, which are not at all rare, a Translation Memory saves us a lot of time and ensures perfect consistency of terminology, which is appreciated by the clients.

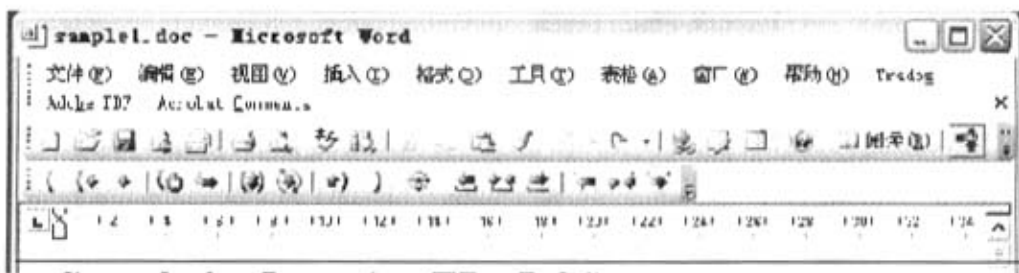
Preservation of the original typesetting and layout means that you no longer need to worry about preserving the original typesetting and layout.

4.2 Introduction of Trado

Trados is software, which provides solutions to translation companies, enterprises and organizations with massive translations duties. It based on UNICODE (unified character code), supports 55 languages. Because of its working platform integrates Microsoft Word perfectly, the user carry on the interactive translation in the Word environment, and may effectively use the Word's processing function of characters.

4.2.1. Advanced TM —— Translator's Work Bench

Translator's Work Bench is the core of TRADOS software ,which can the automatic deal with the same parts of the translation ,keep results in backstage. After completing the installment of the system , the TW4Win (8 function strips) can automatically increase to the Word. The users may carry on the related operations directly in Word. The working platform of Trados is WorkBench+Word (As chart 4.2.1 shows).



Picture 4.2.1

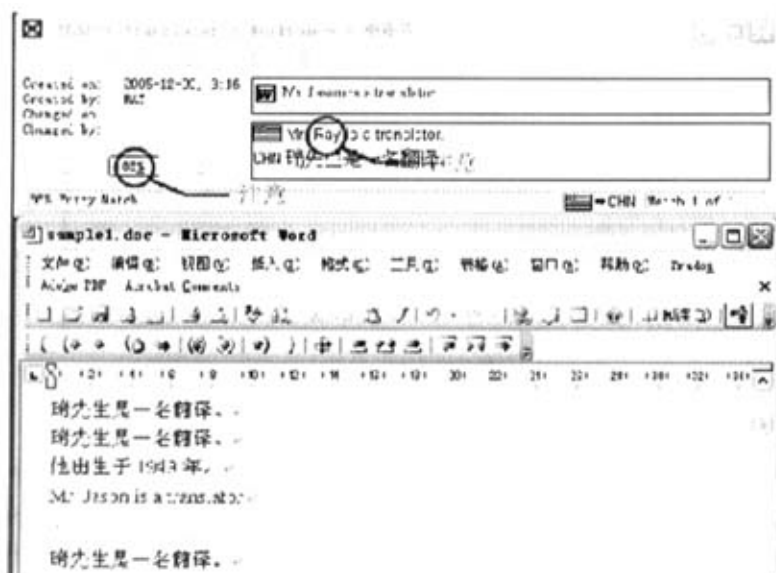
Work routine of Trados is:

(Analyze). Work Bench may carry on the analysis to the source material, search Same/similar resources from translation memory corpus to current text .The software automatically analyzes material's degree of similarity, terminology category and document format. Users estimate the workload, the difficulties and the work load what TRADO S undertakes.

(Translate). Work Bench may choose one or many documents to translate. In translation process, Work Bench may complete the correct translation (match rate 100) (As picture 4.2.2 shows).To the similar sentences and clauses units, TM may establish suitably match parameter. (As chart 4.2.2 shows). System translates sentences and clauses above this match parameter, and marks parameter which does not match with the different color in order to revise in the Word environment. When the users revise the words which do not match, Work Bench can automatically demonstrate the match material from the translation memory corpus. Translators can choose and take the model.



Picture 4.2.1



Picture 4.2.2

Work Bench has the function of comprehensive corpus management. The users may build many different resources' corpus. When translating, users find out the appropriate memories to the materials. We can encounter the situations in the translation process, that is, a sentence has one meaning in this time, at that time another .Two translations are correct, different in environments .

Considering this kind of situation, Trados can solve this problem. The index of resources of Work Bench system is flexible, users retrieve in bilingual corpus with steps of precise inquiry and fuzzy inquiry, also retrieve date and people.

(Clear) after comparing translations, system can automatically save the original texts into the corpus. Users eliminate the original text, and save the current document in Word environment.

(rearrangement) After translating, rearrangement is a barrier for electronic interpreters. It wastes massive time of translators to match the original documents to the TXT documents and the HTML documents with graphs. Testing engineers use RTF documents to keep original format strictly. TRADOS can maintain paragraphs, the frame, bottom grain, the subscript, the typeface, the mark from the original documents, and retain the Internet hypertext links in the documents. These technical characteristics are suitable to the Web documents, especially documents of current mainstream on line, such as, RTF, DOC, HTML, SCML, XML, FM, MIF and so on.

Supporting TMX (Translation Memory Xchange) is a big merit of Trados productions. Trados Conforms to the TMX standard ,which can deal with the database exchanges among the different productions and database.

4.2.2. Standardized tool of technical terminology ____MultiTem

For the localization of international companies' products and the promotional operation systems, Sci-tech document translation has extremely strict request of accuracy and the unification. TRADOS's MultiTem solves the problem of standardization of technical terminology. The user can build the special-purpose terminology corpus according to need independently to guarantee terminologies' consistent and accuracy, especially for the young work team and the company. At present, Microsoft, Oracle and other companies have their own terminology corpus. MultiTem mainly applies in the promotion and the localization of the products.

MultiTem has similarly place with the electronic dictionary: they all are a glossary corpus. But, MultiTem is quite superior in the composition: sharing and index. MultiTem provides the simple coordinated of words, and supports the graph, complex translations, for examples, context comparison clauses.

The user can input the terminology directly in the Interface and Excel. The system will automatically transfer them into MultiTem corpus.

4.2.3. Interactive memory corpus ____WinAlign

WinAlign uses the translation resources fully to establish the documents based on Work Bench,

which can deal with valuable translation materials used in the former projects and reuse corpus' resources. WinAlign maximize the resource and efficiency

WinAlign takes the paragraph as a unit and makes the source text and target text match through the structure recognition and inspection of words, paragraph marks, sentences, clauses and sentences. After matching, the system will reveal the match results automatically. It may shut off the segments, which do not match directly, and connect correct matches. If translations of Trados is difficult to understand, users can help Trados match through the definition mark, the numeral, the form and other some elements (various names of products.)

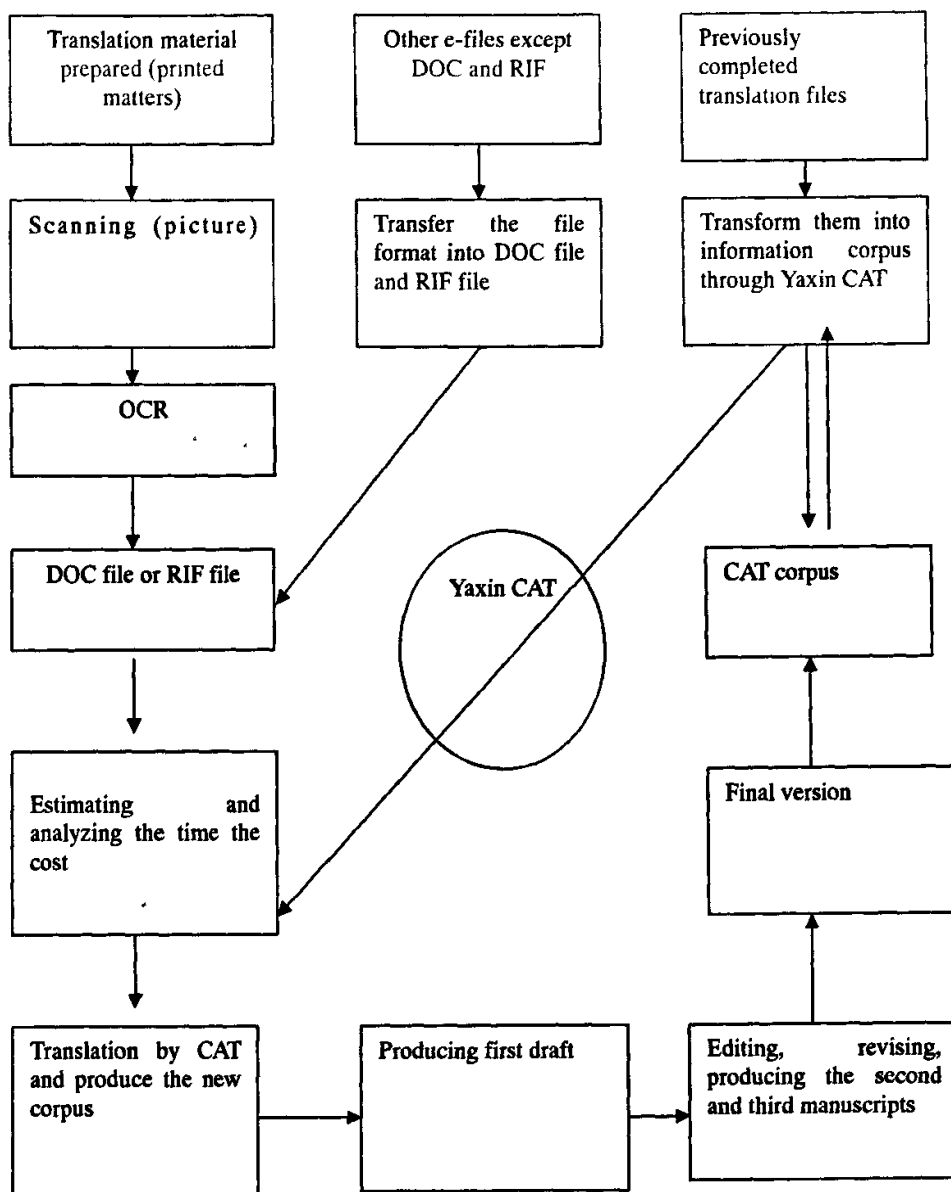
4.3 Introduction to Yaxi CAT

As the large and middle scale enterprises, the scientific research units, the translation companies, the localized companies, foreign affairs departments, translations are massive, especially artificial written translation, which does not have the unification work flow, high efficiency, moreover, low cost.

According to this situation, Beijing Yaxin profited from the domestic and foreign advanced translations experience, develops "Yaxin CAT (Computer Aided Translation)—the specialized translation platform ", and provides sets of translation solutions for the enterprise users like CAD.

Along with the use and study, more and more new word, grammars, the sentence patterns store in the corpus. CAT can save more time for the user .Because CAT does no depend on machine's automatic translation, and let people participate in and complete the translation.

The uses can establish the unified work flow to enhance the efficiency and quality greatly, as the picture shows.



4.4 . Comparison between TRODAS and Yaxin CAT

Similarities:

CAT and TRADOS use the advanced TM technology which is the core of translation technology. TM can accumulate translators' wisdom and the experience.

Differences:

The TRADOS integrates Word directly, avoids typesetting after finishing the translation, and translates graphs and formula in the original text. Its operation is simple, and the design is advanced. Trados supports 55 languages. But, TRADOS does not have the function of electronic dictionaries.

By the automatic memory, intelligence analysis and interactive technology, Yaxin CAT is the domestic origination and in leading level of the world. Yaxin has three superiorities: Memory. Yaxin automatically produces the correct translation; Analysis. With the open grammar system and interactive operative system, convenient man-machine interactive system, Yaxin carries on translations and revisions. Because of Chinese software, it is advantageous for application. The most prominent characteristic is 23 specialized corpuses (in the electricity dictionary) and exchanges among them willfully. It was said that after being familiar with this software operation and application, a graduate of senior school can carry on the special translation.

Chapter 5 Project management of Sci-tech document translation

5.1 MT and human translation

This section introduces the processes of traditional translation and MT

5.1.1 The traditional translation's process:

- 1) Translators read the whole text cursorily; and grasp the general idea of the content. When start translating Sci-tech documents, they meet many new materials and the terminology. Translators should have an intimate knowledge of related specialized field, understand entirely the related society background. It is extremely advantageous before the formal translation
- 2) What translators should do includes: read intensively chapters by chapters, sections by sections, sentences by sentences, analyze the full text, make each word thoroughly with the concrete sentence meaning, and clarify the paternal construction and the grammar relations. To some new words, translators analyze their implications according to the context (commonly used languages or specialized category), look into the corresponding dictionaries carefully, and get the right explanation. You know, hasty makes mistake.
- 3) In the process of translating, translators should not translate one sentence you just read. The right way of translation is to translate paragraph by paragraph. It is advantageous to get distinguished meanings according to the context. It is convenient to pay attention to the logic relations and link up sentence by sentence, paragraph by paragraph and make translation smooth in the end.

5.1.2 Cognitive process

It is necessary to understand the functioning of the human brain to understand the essential principles underlying MT.

The first stage in human translation is complete comprehension of the SL text. This comprehension operates on several levels:

Semantic level: understanding words out of context, as in a dictionary.

Syntactic level: understanding words in a sentence.

Pragmatic level: understanding words in situations and context.

Furthermore, there are at least five types of knowledge used in the translation process:

Knowledge of the SL. It is the foundation to allow us to understand the original text.

Knowledge of the TL. With it, translators can produce a coherent text in that language.

Knowledge of equivalents between the source and TLs.

Knowledge of the subject field as well as general knowledge. Both of them aid translators to understand the text of SL.

Knowledge of socio-cultural aspects, that is, translators have the knowledge of customs and conventions between the source and target cultures.

Given the complexity of the phenomena that underlie the work of a human translator, it would be absurd to claim that a machine could produce a target text with the same quality as done by a human being. However, it is clear that even a human translator is seldom capable of producing a polished translation at first attempt. In fact, the translation process comprises two stages: first, the production of a rough text or preliminary version in the TL, in which most of the translation problems are solved, but the translation is far from being perfect; and second, the revision stage, varying from merely re-reading the text, translators make adjustments of the minor mistakes and the implementation of radical changes. It could be said that MT aims at performing the first stage of this process in an automatic way so that the translators can proceed directly to the second, and carry out the meticulous and demanding task of revision. The problem is that the translator now faces a text which has not been translated by a human brain but by a machine. The change requires the new approach because the errors are different. It becomes necessary to harmonize the machine version with human's process, judgments and experiences. MT is thus both an aid and a trap for translators: an aid because it completes the first stage of translation; a trap because it is not always easy for the translator to keep the necessary critical. Although there is a distance from a version translated by human being, at least in a rudimentary way, machine has already done its work and decreased the burden of translators.

5.1.3 The MT's process

The MT started from the end of 1940s last century, and has already developed many different methods until now. Regards as rule based MT; translation process is under the guidance of linguistics and a reversal process among symbols. This method requests the operational SL, that is, the computer ("understands") the form of SL. Followings are examples to explain translation process of Chinese-English MT.

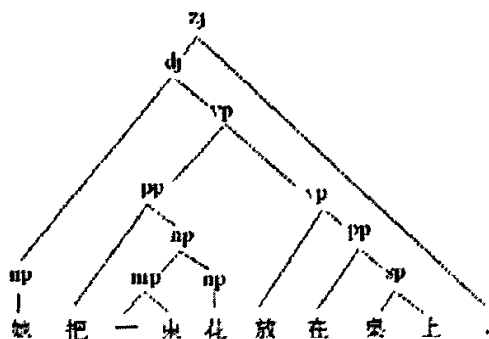
1) The lexical analysis of the SL. In this stage, Machine uses stratification knowledge of SL, distinguishes string of character in the SL text, obtains the syntax and semantics knowledge of each word from the dictionary and corpus, and prepares usages in following processing.

For example: Chinese sentence "她把一束花放在桌上". Finishing the process of lexical analysis by machine, we get the result as showed below:

她/r 把/p 一/m 束/q 花/n 放/v 在/p 桌/ng 上/f 。/w

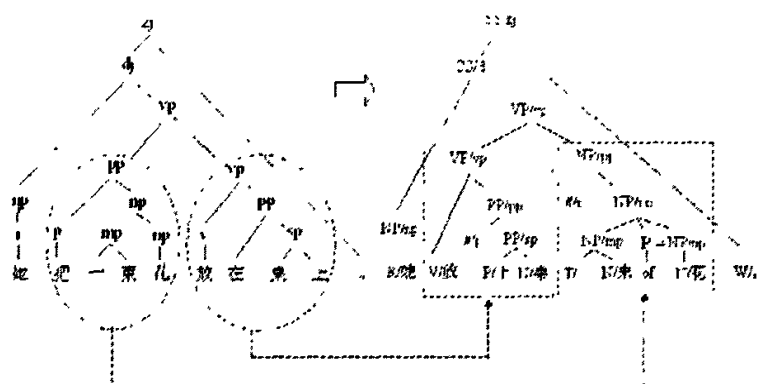
Characters are the lexical category marks, *r* expresses the pronoun, *p* preposition, *m* numeral, *q* classifier, *n* noun, *v* verb, *f* locative, *w* punctuation mark (the following marks use the same meaning with these).

2) The syntax analysis of the string of SL. In this stage, machine analyzes the syntax structure as well as features of structures. Machine uses the syntax stratification, and the relevant rules. The process is composed by two parts, one part is the rule (Context Free Rule), which indicates the phrase relations, for example: A noun phrase can describe as followed: $Np = mp + np$. Moreover, a phrase is a group (unification formula). The new language unit attributes information reuse. The result of syntax analysis can express in a syntax tree. The syntax analysis of the Chinese sentence showed above will be able to obtain the syntax tree, which is shown in the picture below. (The picture leaves out the information of lexicon).



Picture 5.1.3.1

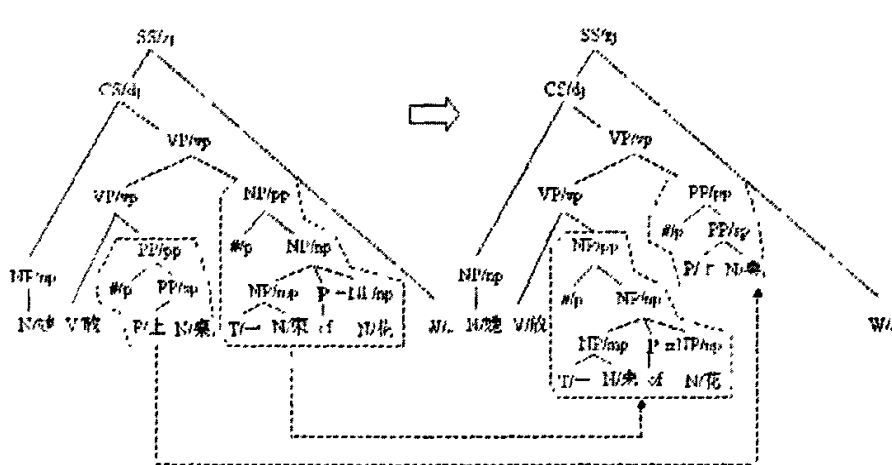
3) Transforms the syntax structure of SL into the TL, the structure transforms the mainly use corresponding relations between SL structure and the TL structure. And through a group of transforms rule transforms the syntax tree of the SL syntax tree to the TL. The transformation rule has listed the SL syntax structure as well as the corresponding TL structure, and described this kind of transformation condition, regarding the above theory, the chart -2 described this kind of structure t transform, right is after the transformation and obtains English structure tree, each tree point sign two oblique lines__ separation category mark, left side of the category is results after the right side transformation.



Picture 5.1.3.2

In the tree derivation (including the under chart 2, 3), Chinese phrase mark indicated by the lowercase letter, English phrase mark by capital letter expression. Zj expressed the entire sentence, dj expresses the minor sentence, np expressed noun phrase, vp expressed the verb phrase, pp expressed the preposition phrase, mp expressed the number lexical category phrase, sp expressed quarter lexical category phrase, SS expressed the English entire sentence, CS expresses the minor sentence, W is punctuation.

4) Adjustment of the TL structure. Usually, after finishing the structure transformation, the syntax structure also retains many traces of SL's structure. Machine needs the adjustment according to the syntax knowledge of TL, picture5-1-3-3 shows the translation structure after adjustment



Picture 5.1.3.3

5) The word choices of the TL text .The choices mainly need corresponding knowledge between two languages. This work is not easy to complete. At usual ,the word of SL often corresponds many words in the TL ,for instance ,in Chinese phrases “开飞机和开门” ,the word “开” is translated into English separately, that is, "fly" and "open". How to choose correct translation depends not only the bilingual corpus, but also considerations to the contexts carefully. The words of the Chinese sentence can get the correspondences showed below.

她↔she 放↔put 一 ↔a 束↔bunch
花↔flower 上↔on 桌↔table
↔

6) Achievement of translation. After the process of transformation as well as choice of the words, strings should be moved away, for example, the leaf from the structure tree's which shows in Pic.5-1-3-3 and arranged words in order. We must pay attention to the difference between SL and TL, adding or omit something, and we will obtain the appropriate form showed below. For instance, in this example "put" becomes the third person form, that is "puts"; "flower" must become the complex form, "flowers". Regarding the above example, finally computer makes the following English translation (# is to express spatial form, and the translation "of" is depends on "blank").

她 放 一 束 P/# 花 上 桌
She puts a bunch of flowers on table.

From the main introduction showed above, the MT involves the SL analysis, the transformation of SL to the TL, the production of TL and so on . Each process all needs the language knowledge to tell the computer to make the correct choice.

5.1.4 Analysis of Some Errors in Machine-translated Texts

The purpose of analyzing errors in machine-translated texts is to show the translation done by a famous net system as an example of MT, give suggestions to use MT and improve the research of MT. An article is from a Sci-tech document with a topic of “Combined physical, chemical and biological treatments of wastewater containing organics from a semiconductor plan”. In this case we will represent the English-Chinese translation produced by www.zhongguofanyi.com, a simple revision by a human and a human translator.

SOURCE TEXT: Wastewater containing organics from a semiconductor plant was experimentally investigated in this study. The wastewater is characterized by strong color, high chemical oxygen demand (COD), a large amount of refractory volatile organic compounds and low biodegradability. Because of these characteristics, treatment of this wastewater by traditional activated sludge method is essentially impossible. (52 words)

www.zhonguofanyi.com	Revision	Human translation
污水包含有机物从半导体植物实验性地被调查了，在这项研究中。污水为强的颜色描绘，高化学氧需求(鳕鱼)，很多加工困难的挥发性有机化合物和低 biodegradability 。由于这些特征，这污水的治疗由传统被激活的烂泥方法根本上是不可能的。.(101 words)	对有机污水的半导体进行了实验性的研究。污水的特点是亮度高，COD 高，不容易挥发的有机化合物多，生化低。由于这些特征，传统的有效治理污水的方法从根本上是行不通的。(78 words)	本文对半导体有机废水的处理进行了试验性的研究。半导体废水的特点是：色度高，COD 含量高，含大量难降解挥发性有机物，可生化性差。因此传统活性污泥法不能根本有效地处理半导体有机废水。(87 words)

The first point to be made is that MT focuses on the SL, while human translation aims at comprehension of the TL. MTs are always inaccurate because they take the words from dictionaries and based on the situational limitations set by the program designer. Various types of errors can be seen in the above translations.

1. Errors that change the meaning of the lexeme

Words or phrases are apparently correct but the translations are not suitable to the meaning in the concrete context:

Original:	strong color——treatment ——activated
www.zhongguofanyi.com:	强的颜色——治疗——被激活的
Revision:	亮度高——处理方法——有效的
HT:	色度高——处理方法——活性的

2. Words without meaning:

Original:	biodegradability——COD
-----------	-----------------------

www.zhongguofanyi.com: biodegradability——鲟鱼

Revision: 生化——COD

HT: 可生化——COD

3. Errors in usage

The translation is understandable in that the MT produces the meaning but does not respect usage:

Original: was experimentally investigated in this study

www.zhongguofanyi.com: 实验性地被调查了

Revision: 进行了实验性的调查

HT: 进行了试验性的研究

Original: this wastewater by traditional activated sludge method is essentially impossible.

www.zhongguofanyi.com: 这污水的治疗由传统被激活的烂泥方法根本上是不可能的。

Revision: 传统的有效治理污水的方法从根本上是行不通的

HT: 传统活性污泥法不能根本有效地处理半导体有机废水。

As mentioned above, human translation concentrates on the TL. Translators always prefer reproducing meaning departing from the literal meaning of SL, if necessary, in order to let people of native language understand. For example, human translator clearly chose “色度高, COD 含量高, 含大量难降解挥发性有机物” as a better contextual translation than “亮度高, COD 高, 不容易挥发的有机化合物多, 生化低。” ,and than the literal versions seen in the MTs.

MT aims primarily at general idea, but not at the production of a perfect target text. It is important to follow two basic rules in order to make the best use of MT. First, we need to recognize the types of texts, for example, poetry is not suitable for MT. Second, it is essential to focus on the source text, as even one letter can radically change the meaning of the whole sentence. For example, the omission of an “s” in the source text will make the MT incomprehensible. So, it is important to analyze SL before the translation.

5.2 Project Management in Sci-tech document translation

The aim of project management is to avoid the mistakes mentioned above and improve the quality and speed of the Sci-tech document translation.

5.2.1 Basic information of Project Management

According to the definition of American Project Management institute—PMT, project is a service which creates the specific production ,has the time sliking. In short, project is: performed

by people; constrained by limited resources; planned, executed, and controlled. The Sci-tech document translation belongs to Project Management

The translation of Sci-tech document has characteristics of Project Management:

- 1) sense of purpose, namely the distinctive translation achievement;
- 2) the process of translation is a new creation again, not the simple pure duplication. Each translation issue has the explicit time limit, not the unceasing repetition again and again
- 3) the conditionality, the translation activity receives the detachment condition restriction in the certain degree, in particular time resources, human resources, and information limit. Any issue all having the time. Therefore the time resources are especially important to the translation activity. When has the translation demand, it is not far away from submitting the achievement.
- 4) human formerly resources of technical content can not be utilized when organizes people to be engaged in the specialized translation in the short time because it is a truly complex and arduous work. This needs to introduce the idea and the method of "Project management" which makes the translation systematical and effective.

The project management takes the Sci-tech documents as the object system. Through a temporary special flexible organization, the leader of project management achieves the plan efficiently and realizes dynamic management. The principle of the project management is "the whole—decomposition—synthesis". Decomposition means the project is divided into different units. Everyone is responsible for one part. According to the principle, project team should build effectively first.

The Project management cares about the task's schedule. That is, translator should complete the target according to specifications and requests, and submits the achievement (deliverable) on time (within budget). When we take over some translation projects, our customer offers the demand and the expectation. How to meet and realizes these, the urgent matter is to set up the system which MT aided. When there is a large-scale translation activity, there will be an effective management. We can employ an manager, too. He decides the time progress, the quality requirement, Meanwhile, the explicit division of labor and the cooperate partners. Terminology unification becomes easy.

5.2.2 General idea

- 1) Aid tool of human translation. MT is an aid tool, we emphasizes it is advantageous to use corresponding functions of the computer to adapt and optimize current translation pattern.
- 2) Aim at automation. Project management offers full scopes to computer's superiorities. (Technology of natural language processing, technology of Database, technology of retrieval and the computational method), provides each kind of tool to assist the translation effectively and Let

the computers undertake more work.

3) Center on database. The management gives emphasis at classification of the database, which centralizes and preserves in the computers. One part of database management is completed by human, and the other is by computer automatically.

4) Method of Management. Project management takes a Sci-tech document as a "project". After reaccepting a translation task, the translator need set up a "project" firstly. Management mechanism starts to track the duty advancement and the complete database, establishes relation automatically between task and database, and Coordinates database transforming, renewal and preservation.

5) Arrangement of work flow, as the picture showed in introduction. The work flow of the project management contains 5 parts.

①Collecting the information

As shows in picture 5.2.2.1

Required information	Information's source	Gain way
Specialized knowledge	Original texts	writer
Definition	Related Sci-tech documents	web search engines
Knowledge spots	Related web pages	Libraries
Common Knowledge	Specialized dictionaries	Related organizations and experts

Picture 5.2.2.1

By computational tool, people store information according to classifications

A, Non-electric information is inputted into computers. So, system can control exterior material

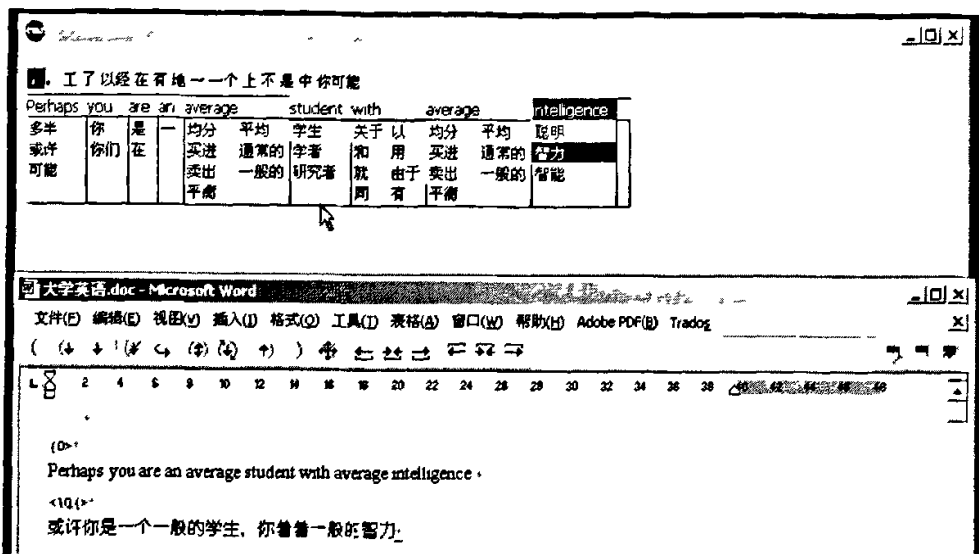
B, Dictionaries and terminologies are inputted into corpus, according to the form required

C, label the words and sentences with computational marks.

In information processing, users formalize the primary database and code the languages in order to help computer search knowledge saved in the corpus. The related information is stored in computes as knowledge library for system.

②Analyzing original text

Translators analysis original texts based on tools of exchangeable information, make e-notes, so as to focus on the problems.



③Translating original text

People use Translation Memory, terminology management mechanism, corpus to help translation. People can ask retrieve. Computer can keep new TL text and SL text in corpus.

④Proofreading the translation

The translator may use uneven way to inspect situation between the original text and the translation, inspect the language quality, chapter structure and logic of translation itself with the full text way.

⑤recycling Database

Before project is ending, management mechanism reorganizes the complete database, derives it to uses in the future, and clears other database according to translators' request.

5.2.3 Method of the project of Sci-tech document translation

5.2.3.1 Management preparation

After receiving the explicit project, the leader of project should accept the principle of minimizing unused resources ,set out the plan in detailed, and arrange the progress .For example, using Work breakdown structure—WBS, management sets up responsibility matrix, makes everyone is responsible for only one part of the work.

In decomposed duty, translators can use CPM, Critical Path Method, the most important route in a long-term work to discover the critical path. Through increasing human resources and cashing, the translation task is finished in the scheduled time. In brief, work in part is for using each kind of

resources fully and effectively.

5.2.3.2 Management in the project

Management in the project of Sci-tech document may be realized through the synchronization control, that is, compare actual situation to plan deviation, translators take the effective action to prevent the major problem accumulated. In the translation activity, team members must maintain close relation and communicate with examines and editors to avoid technical mistakes, and to realize translation quality's control.

5.2.3.3 Management after the project

After finishing a translation mission, members in some translation organizations or departments thought it is enough. Many experiences have not summarized promptly result in the similar question and wrong always appears. Now, as a project, the team leader should let members understand and communicate the methods, ideas and problems in the project, analyze, summarize and correct them promptly in order to improve management and the translation quality.

5.2.4 Advantages of Project management

1)Using the method of Project management__ target management and quality control in the activities of Sci-tech document translation, translation quality can be guaranteed through closing control in various stages and promptly correcting errors. Before starting the project, through forward control, members in the team collect information and predict every kind of possibility they will meet in the project, for examples, the terminology unification, the article form and bibliographies. Something difficult is done.

2)Decomposing work, everyone in the project can tie with his own responsibility and require his own time It is a clear and effective work way. Effective team is established and efficiency is enhanced through Work breakdown structure and responsibility matrix.

5.3 Open resource

Along with the development of Internet technology, some translation organizations overseas have established the special websites on-line to introduce translation achievements, translation researches, translation terminologies and hot topics of discussion .All these webs have been very great transparency. At present, the domestic organizations have not established so many special

websites. It's a big distance far from the western nations.

In our country, translation needs to follow the steps of globalization, meet the request of knowledge-economy society. Translators are longing for an environment and can fully use the modern methods of information times. So, the author suggests that Translator's association group of China establish a website in various provinces and cities cooperates and forms_a nationwide translation network, more complete finally. The platform may set up four columns as followed:

5.3.1 Translation corpus

We can utilize academic units of large and middle scale in each place, and cooperate with the software companies to explore many corpus of technical translation to improve the capacity of resources' usage .In the future, the author believes translation's resource_shared will be realized on line.

5.3.2 Intellectuals' bank

In this corpus, we may files the different specialists and the different ranks of translators in the intellectuals' bank. These intellectuals distributed in the associations: academies, universities, and information departments of the governments and companies and so on. By the traditional way, it is very hard to organize the intellectuals, who will play the very important role in the times of knowledge economy. Therefore, it is the net that is the important way to organize intellectuals. By net, translators also can receive translation courses and translation trainings.

5.3.3 Translation cooperation

In this column, technical translators can share the latest information of translation research. Translation projects of Sci-tech Documents in a big scale can be organized and finished. Volunteers in the column can offer some references which are needed urgently. At the same time, this column can introduce some research results of Sci-tech document translation. It is advantageous for translators to understand translation materials promptly and reduce the waste of the repetition translation.

5.3.4 Translation forum

Translators can exchange the academic information and techniques and build usenets here. Translation forum can realize "diagnoses" of the difficult translation problems by translators from all over the world at the same time. Translators can also strength the communication with foreign

colleagues here. When meeting the update terminologies, translators can ask the foreign colleagues to get the accurate meanings of the original language. The nature of the translation determines that translators will understand the culture and technology first, and then introduce them to person of TL. Translators will come to realize the purpose of net and the MT. So, transforming the idea, raising the innovative ideology, fully using information in the Internet, eliminating the manual workshop of translation and organizing translators are extremely important. Translation should be industrialized to push forward the development of Sci-tech document translation, and form the Chinese characters and superiority.

Chapter 6 Conclusion and suggestion

It can be drawn from the previous discussion that MT plays a positive role in the process of the Sci-tech document translation, when translators and specialists face massive tasks of Sci-tech document translations and adopt the correct strategies, i.e., CAT, project management. MT can decrease Sci-tech translators' burden. Because of the limitation of the artificial intelligence, the realization of automatic translation still has a long way to go. We have to use computer assistant translation in practice. There are so many translation assistant tools such as web translation and software translation. The author analyzes several kinds of them, and gives some suggestions on how to choose and use the relative tools. The author thinks it is not enough to use CAT and project management, so in chapter 5 the author gives the suggestions to set up information_shared platforms of information, where people can communicate with others, share information of translation project and receive relative trainings and curricula. Communication and study in the MT platform can improve translators' habits. Through the development of the machine and the endeavor and human being, the author believes quality of Sci-tech document translations will be greatly improved; speed of Sci-tech document translations will become higher with the development of science and technology.

There are two ways to improve quality of the MT, one is to confine the MT in certain fields, such as bank service and financial service, and the other is to permit people to involve in MT. The essence of current study explores the potential combination of the ways mentioned above. The thesis is based on a limited subject for MT of Sci-tech documents, the ideas of CAT and suggestions of MT can be applied in other relative fields in the future study and promote the development of translation career.

ACKNOWLEDGEMENTS

Here I will send my sincere thanks to my tutor , Professor Du Yaowen, who gives me a lot of valuable advice on my thesis with great patience and effort. My thanks also go to my teachers in Department of Foreign Languages, particularly to Professor Zhao Anyuan, Professor Hao Mei, Professor Guo Aiping, Professor Hou Tao, Professor Feng Shimei, and others.

Still I would like to express my appreciation to my dear classmates and friends who give me a lot of help and encouragement.

Finally I should also express thanks to my dear parents for their encouragement and support all along the way.

BIBLIOGRAPHY

1. Bozena Henisz_Dostert,R.Ross Macdonald,Michael Zarechnak, "MT", mouton publishers,1979
2. Christopher S.Butler, "Computers and Written Texts", Blackwell, 1992
- 3.D. Arnold, "MT: An Introduction guide" NCC Blackwell Oxford(1994)
4. Eugene A.Nida, "Toward a Science of Translating", Shanghai Foreign Language Education Press, 2004
5. Eugene A.Nida, "Language and Culture", Shanghai Foreign Language Education Press, 2001
- 6 Jonathan Slocum, "MT systems", Cambridge university press, 1987
7. Kristin R.Woolever,Helen M.loeb, "Writing for the computer industry", prentice hall,1994
8. Madeleine bates, Ralph m.Weischedel, "Challenges in natural language processing", Cambridge univercity Press
9. Newmark, Peter. "A Textbook of Translation". New York: Prentice Hall, 1988.
10. Peter Newmark, "a Textbook of Translation", Shanghai Foreign Language Education Press, 2001
11. Walter A.Sedelow ,Jr.Sally Yeates Sedelow, "Computers in Language Research", Mouton Publishers,1979
12. Wolfram Wilss,"The Science of Translation", Shanghai Foreign Language Education Press,2001
13. W. J. Hutchens, latest Development in MT Technology: Beginning a New Era in MT Research. In: Proceedings of 14 . MT Summit-IV, Kobe, Japan, 1993
15. W.J.,Hutchins Chichester "MT: Past, Present, Future", England(1986)
16. Sergei Nirenburg, "MT", Cambridge university Press, 1987
17. 丁往道, 吴冰, 钟美荪, 《英语写作手册》, 外语教学与研究出版社, 1998
18. 桂乾元, 《翻译学导论》, 上海外语教育出版社, 2004
19. 冯志伟, 《自然语言机器翻译新论》, 语文出版社, 1995
20. 冯志伟, 《自动翻译》, 上海知识出版社, 1987 年
21. 冯志伟, 《自然语言机器翻译新论》, 语文出版社, 1994 年。
22. 冯志伟, 《自然语言的计算机处理》, 上海外语教育出版社, 1996 年
23. 冯志伟, 《机器翻译研究》, 中国对外翻译出版公司, 2004
24. 胡壮麟, 姜望琪, 《语言学高级教程》, 北京大学出版社, 2002
25. 刘涌泉, 《中国的机器翻译》, 知识出版社 , 1984
26. 刘宓庆, 《中西翻译思想比较研究》, 中国对外翻译出版公司, 2005
27. 刘重德, 《西方翻译理论研究》, 中国对外翻译出版公司, 2003

28. 刘颖,《计算语言学》,清华大学出版社,2002
29. 吕俊,侯向群,《英汉翻译教程》,上海外语教育出版社,2001
30. 秦秀白,《英语语体和问题要略》,上海外语教育出版社,2002
31. 姚天顺,王野翊,《自然语言理解》,清华大学出版社 1995
32. 易绵竹,南振兴,《计算语言学》,上海外语教育出版社,2005
33. 翁富良,《计算语言学导论》,中国社会科学出版社,2005
34. 俞士汶,《计算语言学概论》,商务印书馆,2004

攻读学位期间发表的学术论文

Paper Published During the Academic Years
for the Master Degree

发表论文

杜耀文，程霞， 马克思主义与西方社会科学在中国的传播和翻译，《山西高等学校社会科学学报》，2006,18（5）： 124—127

Du Yaowen, Cheng xia “Western Social Sciences and the Spread and Translation of Marxism in China”, Social Sciences Journal of Colleges of Shanxi, 2006, Vol.18（5）： 124—127