TRANSLATING SCIENTIFIC TEXTS USING MT AND MAT SYSTEMS: PRACTICAL EXPERIENCE OF A PROFESSIONAL TRANSLATOR

Olga Bezhanova.
94 Prospekt Gagarina, ap.192, Kharkov 310140, Ukraine.
Tel.: (0572) 27-03-31.
E-mail address: blekhman@lotus.kpi.kharkov.ua; super@pcl01.ai.kharkov.ua

Abstract:

The paper describes practical experience of a professional translator. The task consisted in translating 400 pages of Russian scientific materials (covering all fundamental sciences) into English within a month. The job was fulfilled using three computer-based systems: **PARS**, a Russian-English bidirectional machine translation system by Lingvistica '93 Co., **Polyglossum,** dictionary-support software by ETS Ltd., and the Random House electronic dictionary of the English language. The paper analyzes the pluses and minuses of translating scientific texts using computer programs, and gives numerous examples of translations. The main conclusion is that machine translation has no reasonable alternative when a large volume of scientific texts is to be translated professionally within a short period of time.

1. Introduction

The present paper analyzes the work accomplished on the order of The Russian Foundation for Fundamental Research. The work gave valuable material for the analysis of the modem translation facilities.

It became evident for me long ago that for translating large volumes of texts abundant in special terminology, the professional translator has to use both traditional «paper» dictionaries and something less habitual - machine translation systems and electronic dictionaries. The work that made up the subject of the present investigation can serve an example of application of such facilities for making professional translations, from the point of view of its volume, the complexity of the task, and abundance of special terminology belonging to various subject areas.

2. Task description

First, some words about the task I had to solve to meet the requirements of The Russian Foundation for Fundamental Research (RFFR). As is well known, enormous amount of research into all areas of science is carried out annually in Russia. Some of these investigations are conducted under grants from various Western foundations interested in the development of science in the countries of the former USSR. Projects that have this kind of financial support are included in the «RFFR Annual Bulletin». The 1996 directory comprises about 400 pages. It embraces titles and bibliographic information for several thousand research projects in the following areas:

- mathematics and information science;
- physics and astronomy;
- chemistry;
- biology and medicine;
- geosciences;
- liberal arts;
- databases and books issued in Russia in 1995.

The structure of the Directory was somewhat unusual for me as translator: it consisted of a brief introduction followed by approximately 5.500 titles of projects including the author's surname, the title of the project, the identification number, the name of the institution (University, research institute, etc.) where the research was carried out, and the city/area of residence of this institution; the list of the abbreviations of the titles of scientific institutions, as well as their addresses, were given in the end of the Directory.

It is evident that, except for the brief introduction, the task generally consisted in translating not complete texts but the titles of research projects, each title consisting of two to forty words.

The initial text was a Microsoft Word file of 1.2 MB. The customers required a similar English text preserving the source text styles and formatting. The customers also stipulated that the surnames were to be transliterated according to the rules of the English language, and the titles of institutions were to be translated.

Translating from Russian into English is considerably more difficult for a Russian-speaking person than translating from English into Russian because of the absence of Russian-English terminology dictionaries for quite a number of subject areas. In the ex-Union, many excellent English-Russian dictionaries of mathematics, astronomy, chemistry, biology, etc. were published, but it is absolutely impossible to find corresponding Russian-English dictionaries.

The work was supposed to be done within approximately a month. Taking into account the days-off, the translation was made for 34 days of intensive work, 5-7 hours a day, consisting in post-editing the texts translated by the Russian-English and English-Russian machine translation system PARS by Lingvistica '93 Co.

A great number of scientific terms in the source text relating to numerous subject areas required using quite a number of various dictionaries. I am sure that translating texts of such volume by one person for such a short period of time without using machine translation software is impossible.

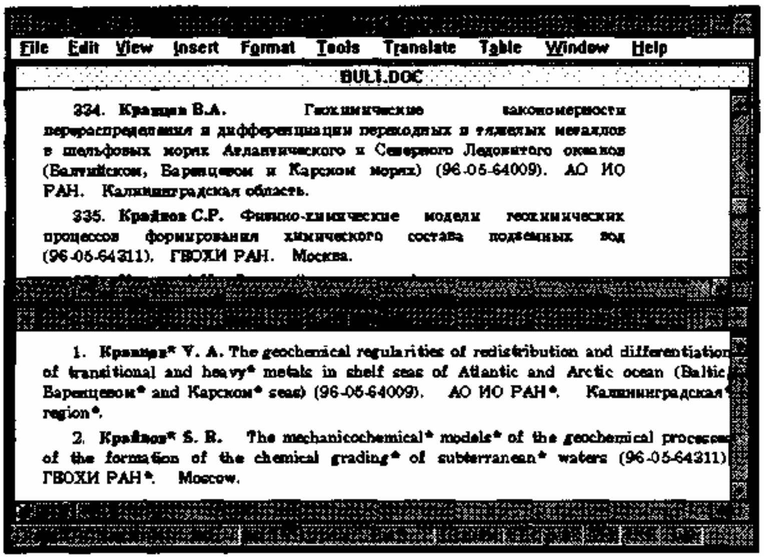
It is also necessary to say that, due to the large number of misprints in the source text (the better half of which composed Latin letters instead of Cyrillic ones in Russian words), the machine translation quality in the first instance turned out to be lower than it could be if the text had no misprints. All Russian words having Latin letters were left untranslated by the system, which considerably complicated post-editing. That is why machine translation was preceded by context substitution of Russian letters with their Latin «analogs», which slowed down the whole process.

As it was already mentioned, the initial, draft translation was performed by the PARS machine translation system. PARS for Windows is linked with WinWord 6.0 and

WinWord 7.0. The system has the following characteristics that considerably simplify translation:

a) the user may set the dictionaries to be used according to the subject area, as well as their priorities;
b) it is possible to modify the dictionaries: translating the source text by portions, the user clears up the translations of the words missing in the system dictionaries and enters these words into the corresponding dictionaries. Further on, the system will use these words, which considerably raises translation quality. It is necessary to note that entering words and phrases into PARS dictionaries is extremely simple compared with other MT systems I am aware of, and this option is very easy to master for the user who wants to work with this system professionally;
c) PARS features automatic transliteration of proper names, which was very useful for translating the Directory;
d) PARS translates directly in WinWord, preserving the source text format in the target text.

Before the translation session, the Polyglossum system of dictionaries by ETS Ltd. was activated on CD-ROM, which made it possible to access any of the dictionaries by pressing Alt+Tab without exiting from WinWord.


File Edit View Insert Format Tools Translate Table Window Help
BUL1.DOC

334. Кравцов В.А. Геохимические закономерности перераспределения и дифференциации переходных и тяжелых металлов в шельфовых морях Атлантического и Северного Ледовитого океанов (Балтийском, Баренцевом и Карском морях) (96-05-64009). АО ИО РАН. Калининградская область.

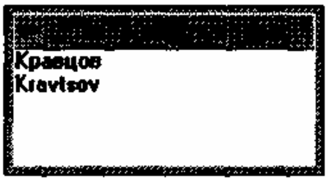
335. Крайнов С.Р. Физико-химические модели геохимических процессов формирования химического состава подземных вод (96-05-64311). ГЕОХИ РАН. Москва.

1. Кравцов* V. A. The geochemical regularities of redistribution and differentiation of transitional and heavy* metals in shelf seas of Atlantic and Arctic ocean (Baltic, Баренцевом* and Карском* seas) (96-05-64009). АО ИО РАН*. Калининградская* region*.

2. Крайнов* S. R. The mechanicochemical* models* of the geochemical processes of the formation of the chemical grading* of subterranean* waters (96-05-64311) ГЕОХИ РАН*. Moscow.

Fig. 1.

As a result of machine translation, the screen is split in two WinWord windows: the source text is in the top window, and the bottom one contains the unedited, draft translation produced by PARS (Fig. 1). This makes it possible to post-edit the text comparing it with the original. The polysemantic words as well as words beginning with capital letters, i.e. potential proper names, are marked in the target text with asterisks. The translator can choose a more suitable translation option of a polysemantic word (phrase), and the transliteration of the proper name (Fig. 2).



Кравцов
Kravtsov

Fig. 2.

Words not found by PARS were looked up in Polyglossum dictionaries.

3. Translating subject area oriented chapters

Mathematics and information science

This chapter of the Directory comprised 800 titles of research projects in the field of mathematics and information science. For the translation of this chapter, the following dictionaries were set up in PARS (in a descending order of priorities):

1) programming dictionary (25,000 terms in each part, Russian-English and English-Russian);
2) technical (76,000 terms);
3) general (35,000 words and phrases).

It is to be noted that these dictionaries turned out to be not enough for the complete translation of the chapter, so I also used the Polyglossum system of electronic dictionaries to post-edit this portion, namely its mathematics and polytechnical dictionaries, as well as the largest paper Russian-English dictionary, by Prof. A.I. Smirnitski (55,000 words and phrases).

On the whole, the chapter «Mathematics and information science» was translated by PARS fairly well, especially the titles of research in the field of information science due to the programming dictionary by Dr. M.S. Blekhman. On the other hand, some purely mathematical terms (for example, “tetrahedron”) were unfamiliar to PARS, and I had to look them up in the Polyglossum mathematical dictionary.

The main difficulty when working with this chapter consisted in translating phrases comprising surnames of «foreign» mathematicians, as, for example, Langevin equation. Because many similar phrases were absent both in PARS and in Polyglossum, I had to look them up in The Great Soviet Encyclopaedia, which presents names of well-known scientists in their native languages.

It is also necessary to note that the first chapter of the Directory, «Mathematics and Information Science», was automatically translated as a whole, which slowed down post-editing because WinWord works much slower with long text portions. The rest of the text was translated by portions comprising 300-400 titles each.

Physics, astronomy

The following dictionaries were set up in PARS for the translation of this chapter consisting of 1290 titles:

1) technical;
2) radioelectronics (50,000 terms);
3) microelectronics (20,000);
4) general dictionary.

Due to the absence of a special dictionary of physics and astronomy in PARS, post-editing of this chapter was more difficult than of the previous portion. The Polyglossum dictionaries, comprising about 1,500,000 terms of various subject areas, were of great help. When post-editing the chapter «Physics, astronomy», I made use of the Polyglossum technical dictionary.

The main problems arose in rendering the names of the planets and their satellites, which I managed to find in the English-Russian astronomy paper dictionary.

Chemistry

For the translation of this chapter (659 titles), the technical and general dictionaries were set up in PARS.

This chapter was the most difficult to translate since it comprised quite a lot of specific chemical terms, such as фталоцианин (phthalocyanine), редокс (oxidation-reduction), рацемат (racemoid), аценафтен (acenaphthene), гваяцил (guaiacyl), etc.

I had to look up the words not found either by PARS or by Polyglossum in the Russian-English Dictionary of Chemical Reactions and in the English-Russian Dictionary of Petroleum Chemistry and Processing because, generally, the difficulty consisted in spelling of the unknown chemical terms.

For example, it was clear that the English translation of the term «стирил» could not differ seriously from the Russian variant, but I was not sure whether it was «styril» or «stiril». I found the word «styryl» in one of the paper dictionaries, which put an end to my doubts.

It was very difficult to translate complex terms consisting of several components, for example, «винилхалькогенополигалогенбензол». PARS failed to translate such words, that is why it took me 6 days to post-edit this comparatively short chapter.

Coming across a word consisting of several components, I usually broke it in sense-bearing parts and translated them in turns. Thus, the term «винилхалькогенополигалогенбензол» was broken into «винил», «халькоген», «полигалоген», and «бензол». The resulting «simple» words were translated and united in one. It's only natural that such work was very labor-intensive and occupied much time.

Biology, medicine

This chapter (908 titles) was the second most difficult to post-edit. The following PARS dictionaries were used for translation:

1) medicine (20,000 terms);
2) aviation medicine (24,000 terms);
3) technical;
4) general.

The main difficulty consisted in translating the names and genders of animals and insects. I could not do without the Russian-English paper dictionary by A.I. Smirnitski, in which I found such terms as «иглокожие», «ракообразные», - «echinodermata», «Crustacea», etc. The dictionary by A.I. Smirnitski comprises quite a number of biological terms, and I made the greatest use of this dictionary for post-editing the chapter «Biology, medicine».

Also, I used The Great Soviet Encyclopaedia, in which I found such rare terms as: ветвистоусые (Cladocera), булавоусые (Rhopalocera), сивуч (Eumetopias jubatus).

Besides, I used intensively The Random House Unabridged Dictionary, including its electronic variant. It let me clear up the spelling of such words as Leishmania, Pteropoda, etc.

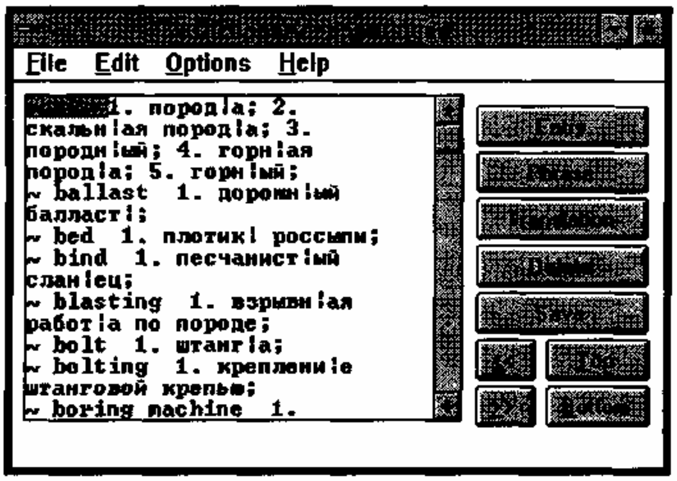
<u>Geosciences</u>

This chapter comprised 752 projects in such subject areas as geology, paleontology, archaeology, ecology, etc. PARS translated this chapter very well, owing to the presence of geological and ecological dictionaries; this raised the translation quality several times as compared with translating such chapters of the Directory as «Chemistry» and «Biology and Medicine».

The chapter was translated in two stages. It was split into two portions of nearly equal sizes that were translated by PARS with the following dictionaries:

The first portion:

1) geological dictionary (11,000 terms) (Fig. 3);



```
1. пород!а; 2.
скальн!ая пород!а; 3.
породн!ый; 4. горн!ая
пород!а; 5. горн!ый;
~ ballast  1. породн!ый
балласт!;
~ bed  1. плотик! россыпи;
~ bind  1. песчанист!ый
слан!ец;
~ blasting  1. взрывн!ая
работ!а по породе;
~ bolt  1. штанг!а;
~ bolting  1. креплени!е
штанговой крепью;
~ boring machine  1.
```

Fig. 3.

2) technical dictionary;
3) general dictionary.

Embarking on the translation of this chapter, I didn't yet know that it comprised many documents on ecology, that is why the ecological dictionary was not chosen for translating the first portion. When post-editing it, I saw that the better half of the words not found in the system dictionaries related to ecology, and I also set up the PARS ecological dictionary (18,000 terms) for translating the second portion.

I also made some conclusions (these are discussed below) connected with setting up the priorities of the dictionaries, which made me indicate the dictionary of technical terms as the prioritized dictionary for translating the second portion. Thus, the list of dictionaries looked like this:

The second portion:

1) technical dictionary;
2) geological dictionary;

3) ecological dictionary;
4) general dictionary.

When post-editing this chapter, I actively used The Random House Unabridged Dictionary to clear up the spelling of such words as Тетис (Tethys), пегматит (pegmatite), and many others. In particular, I made use of the table of geological periods given in this dictionary. It is to be noted that this was the only source where I managed to find translations of a large number of geological terms. It only took me 4 days to translate this chapter, which would have been impossible without using the Random House dictionary.

Humanities and social sciences

This chapter (214 titles) was the easiest to translate. I set up the following PARS dictionaries:

1) general;
2) economic dictionary (55,000 terms).

Despite the fact that this chapter occasionally comprised separate terms of biology, geology and ecology, the number of words not found by PARS was very small.

«Databases of the 95-96ies»

This chapter was translated by PARS very well using the general and programming dictionaries. Post-editing consisted in making minor corrections to the machine translation.

4. Translating and post-editing

This part discusses the process of translation itself, including the difficulties encountered, as well as the ways of overcoming them.

The draft translations performed by PARS were of different quality, depending on the subject area of the source text and, accordingly, on the presence of terminological dictionaries.

A few translations required no post-editing at all or «cosmetic» post-editing. For example:

> Озернюк Н.Д. Механизмы метаболической стабильности процессов развития.

Machine translation:

> 133. Ozernyuk N.D. Mechanisms of the metabolic stability of the development processes.

Or:

> 151. Офицеров В.И. Исследование структурной организации конформационных антигенных детерминант на примере белков оболочки вируса гепатита А.

Machine translation:

151. Ofitserov V.I. Research of the structural organization of conformational antigenic determination on the example of the proteins of the capsule of hepatitis virus A.

Such cases were rare, but they did occur.

On the other hand, in some cases the translation offered by PARS was to be changed completely to obtain the correct text. The fact is that if the system dictionary doesn't have a set expression, PARS translates it word by word, sometimes making it hard to understand.

For example, the phrase «принимающий решения» was translated «receiving decisions», on analogy with «receiving letters», the phrase «образ жизни» was rendered as «image of life» instead of «way of life», etc.

Some translation problems were caused by the absence of some geographic names in the PARS general dictionary as, for example, Yekaterinburg, Petropavlovsk-Kamchatka, Kabardino-Balkariya, etc. At the same time, in the very beginning of my work, I entered these words into the dictionary, which simplified post-editing.

One of the merits of the PARS system is the selecting translation variants option. Here is an example .A title that relates to biology or medicine is to be translated:

580. Носиков В.В. Поиск и изучение антигенных детерминант, связанных с аутоиммунной деструкцией островковых бета-клеток при инсулинозависимом сахарном диабете, с сипользованием библиотек бактериофагов, зкспрессирующих широкий спектр разнообразных пептидных зпитопов.

Machine translation:

580. Nosikov V.V. Search and the studies* of antigenic determinants bound with the autoimmunity disruption of islet beta-cages* at инсулинозависимом sugar diabetes, using the libraries of bacteriophages expressing the broad* spectrum of diversified* peptide epitopes.

A double click on the asterisk will display the list of translation options for this word/phrase. Having chosen one of the variants and pressed the button, the post-editor inserts it into the text instead of the initial one.

In the above text, the following translation variants were offered: studies (research, analysis), cages (cells), broad (wide, capacious, extensive, large-scale), diversified (miscellaneous, diverse).

One of the most important features of the PARS system is that the translator may set up dictionary priorities. When choosing the dictionaries to be used in the translation session, it is recommended to set them up in the optimum order, placing on the top of the list the dictionary which is going to be the most frequently used one in this session. It is common knowledge that a word can have quite a lot of translations out of context, and the right translation depends on the subject area. A good example was given above: the medical term «бета-клетки» was translated «beta-cages». In this case, the system treated the polysemantic word «клетка» as a general-usage one, not as a medical term.

Another example shows a different situation: the system understood the Russian word «опыт» as «experiment», so the machine translation of the sentence «in the context of the European experience» was «in the context of European test*». The translation option for the word «опыт» was «experience», which was quite all right in this situation.

When translating the chapter «Geosciences», I understood very well the importance of setting up dictionary priorities in PARS. As was mentioned above, the first part of the chapter was machine translated with the geological dictionary having the highest priority. The fact is, I believed that geology would be the main topic in this chapter. However, I did not take into account that those texts comprised many common words, which would be translated as their «geologic equivalents».

For example, «разработка нового метода» was translated «the mining of new method», while the word «development» was to be used.

Without doubt, when assigning priorities to the dictionaries, one should remember that it's impossible to foresee all the cases of the usage of a word in any context. The system does not «feel» differences between shades of meaning.

For example, when translating the chapter «Physics, astronomy» the phrase «осажденный магнетик» was translated «besieged magnet». In this case, the system failed to distinguish between the physics term «осаждение» (reduction) and the general-usage «осаждение» - (to besiege a city or fortress). The reason is quite simple: in the dictionary, the word «besiege» is the first translation, and the system gives preference to it. I want to note another interesting feature of the PARS system - the possibility of transposing the translations in the dictionary.

I'd also like to add that the draft machine translations produced by PARS sometimes had funny peculiarities. Thus, sometimes Russian surnames were translated into English as general-usage words: «Бобров - Beavers, Зубов - Teeth», etc. In such cases, I simply used the transliteration facility.

5. Conclusion

The translation of the RFFR Directory was a very interesting and useful work for me. It helped me make up the technology of working with similar texts, clear up, in what dictionaries I can find terms relating to definite areas of knowledge. Experiments with the choice of dictionaries for each of the chapters helped me clear up what dictionaries would be better to use for translating texts of definite subject areas.

For the PARS system, this work turned out to be even a more useful stage of its development. In the process of post-editing, new words and phrases were entered into the dictionaries. New dictionary projects were launched. Presently, PARS comprises dictionaries of mathematics (80,000 terms), chemistry (50,000 terms), a much larger dictionary of geology (27,000 terms), etc.