Through the Looking Glass to the PARS Land

Michael Blekhman

Lingvistica '93 Co.
94a Prospekt Gagarina, appt.111
Kharkov, 310140
Ukraine

1.   Background

Translating texts from one language into another is probably the most difficult task one may think of. My younger daughter, with whom I have been communicating in English since her very birthday, once pleaded, "Oh, Daddy, I will better tell you the whole story in the original than translate it into Russian! You see, I understand everything, but I can't translate!"

When I was younger (but hardly less ambitious) I translated "Alice's Adventures" into Russian. It was a challenge! The puns and the rhymes that make any English kid laugh would hardly mean anything to a Russian speaking child. So, I decided to render them in such a way that the impression they make would be equal, though the linguistic meaning might be quite different.

Here is an example. This rhyme, formally, does not sound like the famous "Father William", though it resembles Vladimir Mayakovsky's "What Is Good and What Is Bad", which is as moralizing as its English "prototype". I hope that this example will be interesting for those speaking or learning Russian.

Крошка-сын к отцу пришел,
И спросила кроха:
"Что такое 'хорошо'
И что такое 'плохо'?"

- Если мальчик стекла бьет
И баклуши тоже -
Слава о таком идет:
Очень он хороший.

- Если маме нагрубил,
Бабушке и деду,
Мне такой мальчишка мил,
Дам ему конфету.

- Если ж он цветы полил
И сварил картошку,
Я б такого отлупил
И подставил ножку.

Мудрый папа спать пошел,
И сказала кроха:
"Плохо делать хорошо,
Лучше делать плохо!"

Take also the Cheshire Cat: "Cheshire" means absolutely nothing to Russian and Ukrainian girls and boys. On the other hand, you can't do without a cat in this story. My decision was to find out another "cat", which would be familiar to Russian speaking children: Cat in Boots! And after the cat in question was found, the rest was easy: I turned him into The Cat Without Boots; having no shoes, he had a lot of humor instead, that's why he smiled so often.

It turned out, however, that translating was even an easier task than teaching a computer to translate. In my first PARS 1.00 MT system, which was marketed in 1988-1989, I tried to make the IBM-compatible mainframe translate English scientific and technical texts in such a way that the end users could extract useful information from the machine-made Russian texts.

Here are sample English-Russian translations of INPADOC data base documents (patent titles):

Source text	Target text
Functional human urokinase proteins.	Функциональная человеческая урокиназа белки.
Drainage bag and non-return valve assembly.	Дренажный мешок и не-return комплект клапана.
6-hydrogen purine derivatives.	6-водород пуриновые дериваты.
Insulin derivatives modified in the B 30 positions.	Дериваты инсулина модифицированные в 30 позиций.
Aminopolycarboxilic acids and derivatives thereof.	Aminopolycarboxilic кислоты и дериваты вышеуказанного (из указанного предмета).

Since then, up to the newest PARS-3 English-Russian 2-way system, I have never been fully satisfied. Sometimes, the electronic Eliza Doolittle merely drives me crazy. And the "cats" never stop haunting me. Once, translating a software manual into Russian, PARS came out with something like "Защитите Кота", which means "Protect Cat"; I looked into the original and was terrified to see that the system did not cope with the file name: Protect.cat! Another example of PARS's brilliance was translating Mr Green as "Господин Зеленый", that is "Mr Green in Colour"!

Back in 1979, Professor Hays, an authority in machine translation, answered my question why he considered deep structure analysis necessary in MT. He said that otherwise the user "might be deceived" by the seeming correctness of the machine output. I was only ("already", as I thought) 28, and I disagreed. Now that I am 15 years older, I have not become a champion of deep structures, but I can see that the user is really deceived, from time to time.

My elder daughter once wrote a letter to her pen friend to USA, and I had it machine translated into Russian. The translation was not bad, in general, but one sentence, though quite correct syntactically, had an absolutely different semantic meaning. The source was "I have passed all my exams", and the output "Я пропустила все мои экзамены", which means "I have missed all my exams". The reason was the ambiguity of the English verb "to pass".

2.   The PARS project

So, what is PARS? Generally speaking, it is a 2-way English-Russian-English MT system running on IBM PCs. Within 1991-1994, PARS-2 versions were marketed in Ukraine as well as in ex-Union countries. This is how it is used.

Its geography is rather wide, from Lvov in West Ukraine to Sakhalin in the Far East, in Russia. PARS has not only "collective", but also quite a number of individual end users. Among the latter, I can see two subgroups: professional translators and Russian speaking scientists and engineers. A very interesting tendency is partial reorientation of the users from English-Russian to Russian English translation, which is caused by Russian and Ukrainian penetrating Western markets. What is very pleasant, some users need both English-Russian and Russian-English translation, which proves the initial "two-way" philosophy was correct.

There are people who have negative attitude to PARS (or maybe to MT in general), saying that it is easier to translate the whole text manually than to post-edit the machine product. The most unexpected thing, however, is that the higher professional skills, the better attitude to MT. I discussed the problem with a lady translator who worked at the Izvestia Concern, Moscow, where PARS is implemented. Izvestia is Russia's most influential daily, and PARS 2.04 had been selected among the MT systems marketed in Russia for translating the VWD Information Agency data base. I was sitting there for about an hour and a half, working with PARS, while she was manually translating VWD information, using a standard 2-window text editor and a pile of "paper" dictionaries. I noticed that she only managed to translate 2 documents within that time, though, certainly, she was supposed to give "ideal" translation, which is unattainable for any MT software. It only took PARS 2.04 a minute to rough translate one document on an IBM/286 PC, and the translation quality seemed not bad to me, the contents being quite clear. However, when asked if PARS could help her, she said, "You see, it seems useless since the machine product needs too much post-editing, which is a terribly boring task for me".

On the contrary, one of my clients, Mr Vladimir Kolykhmatov, who is a very experienced translator working for the Moscow Agency of the Dupont Company, translating 80-90% from Russian into English and 10-20% from English into Russian, says he cannot do without PARS, adding "I am too lazy to do the job all on my own; I prefer to post-edit the raw translation instead of typing everything in with my fingers".

I am very glad PARS has also been appreciated by non-translator end-users. In particular, Dr. Boris Piskunov of the South Sakhalin Institute of Geology sent me a PARS-made English translation of his scientific report he was supposed to submit to his South Korean partners, with the commentary that he was quite satisfied as the volume of post-editing needed was not large. I have also received similar information from Dr. Vladimir Petushkov of the Ukrainian Welding Institute. Both of them mastered entering new words into PARS dictionaries, saying, however, that the dictionary extending routine needed improving since it was not easy to encode Russian words, especially the verbs. This is really so, which is why I decided to work out a much more flexible routine, with automatic encoding, which was implemented in the PARS 3.0 system.

3.   PARS-3

In July, 1994, we developed the first version in the PARS-3 series (see Fig.1). It runs under MS DOS and under Windows 3.1 and it can operate in network configurations under Novell.
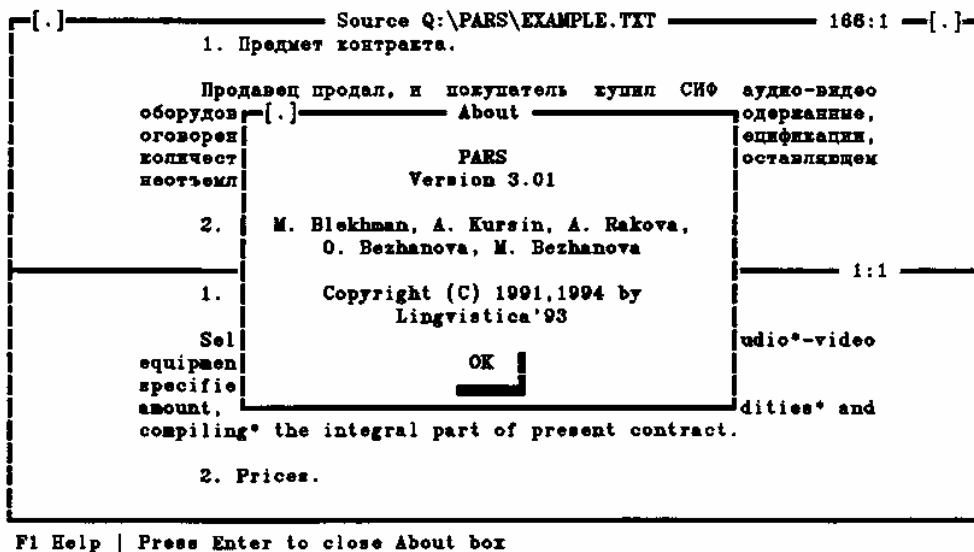

Fig.1. 'About' PARS 3.01

Developing PARS-3, we tried to take into account the main advantages and drawbacks our end users had pointed at. It was clear that, being as imperfect in the output quality as any "average", that is marketable, MT system, PARS-3 was to compensate for the translation errors in some way. Not that we don't pay attention to improving translation quality (on the contrary!), but the "cats" do come out of the shadows, from time to time. So, we decided to make PARS as user friendly and "recognizable" as possible.

The first thing to mention is how PARS-3 looks: its interface is quite familiar to those who has experience of working with Borland's products, being in accordance with the CUA standard; the numerous HELPs and menus are either in Russian or in English, so that the user could choose. All the standard functions are supported, and quite a number of specific ones added.

PARS-3 has a flexible built-in text editor. It will be made compatible with WordPerfect and Word for Windows, so the user will be able to enter PARS and have any Word or WordPrfect file translated into English or into Russian.

Using the built-in editor, you may also select any of the synonyms of the ambiguous words and have it pasted into the target text with a single keystroke. A list of "new" words is created automatically in every translation session, and you may have a word or idiom entered into the dictionary directly from the target text.

Point No 1 in PARS-3 is the new approach to dictionary management.

PARS supports convertible English-Russian-English dictionaries. Convertibility permits you to type-in, for example, a Russian word with its translation, and the system will immediately automatically set the English-Russian correspondence. In other words, one and the same dictionary is used both for English-Russian and for Russian-English translation requirements.

The user may browse the dictionary, arranged either by the Russian or by the English parts, edit it, pass on to any word by pointing it fully or by initial letters.

The user also has a very flexible routine of adding new words to the dictionaries. The program features an automatic encoding routine, which lets the user enter a Russian word in its initial full form and have it automatically encoded; in this case, the program recognizes the grammatical characteristics of the word: part of speech; aspect (perfective or imperfective for verbs); gender (masculine, faminine, neutral, plural for nouns), etc.; type of the morphological paradigm: singular and plural noun forms as well as the whole conjugation paradigm of the verbs, including personal, adverbial participle, short participle, and imperative forms. The program splits the word into the stem and the ending, and separates them with a vertical line for clearness.

Here is an example of the conjugation of the Russian verb "делать": дела¦ть, дела¦ют, дела¦ет, дела¦л, дела¦ли, дела¦ла, дела¦ло, дела¦ю, дела¦ем, дела¦ешь, дела¦ете; in the perfective form, future forms are used instead of present: сдела¦ют, сдела¦ет, etc.

The form of adverbial participle (дела¦я, сдела¦в), short-form participle of the verb in perfective aspect (сдела¦н), and imperative mood form (дела¦й) are also determined automatically. Fig.2 shows how PARS automatically encodes a verb, displaying the list of verb forms (the Russian verb "чувствоваться" has been encoded).
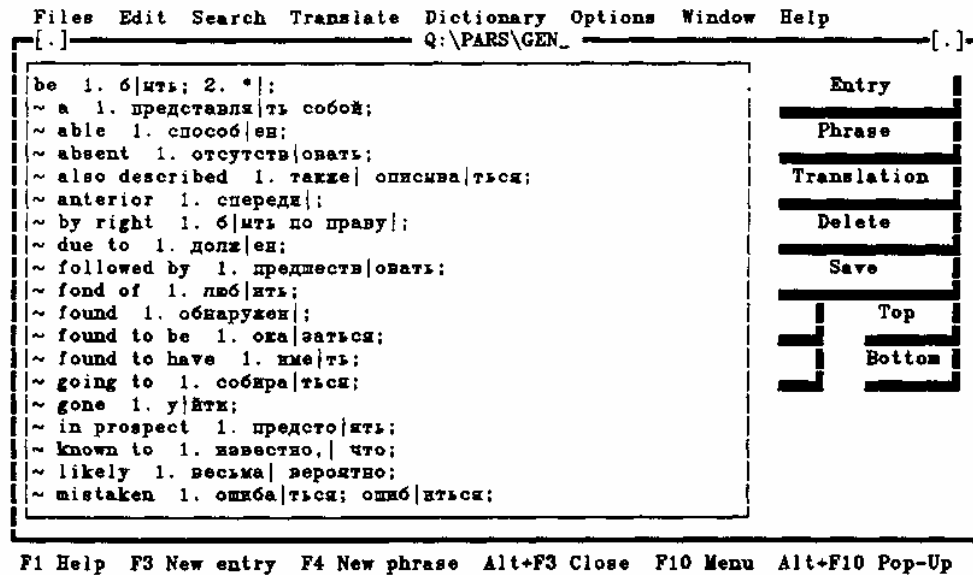
```
 Files  Edit  Search  Translate  Dictionary  Options  Window  Help
┌──┌─[.]──────────────────── Q:\PARS\GEN_ ────────────────[.]┐─────────┐
│  │      ┌─────────────────[.]──────────── Russian Phrase Edit ──────────────┐  │
│  │чувств│оваться │ чувствоваться                                           │  │
│  │      ┌─[.]─────────────────────────────────────────────┐                │  │
│  │      │  Types of conjugation                           │                │  │
│  │      │                                                 │                │  │
│  │      │  пита│ться           .   чувств│оваться   ▓     │ ▄▄             │  │
│  │      │ >чувств│оваться <     ▓   чувств│уртся            │ ▄▄             │  │
│  │      │  рос│                 .   чувств│уется    ▓      │ ▄▄             │  │
│  │      │  оторв│аться          ▓   чувств│овался           │ ▄▄             │  │
│  │      │  хач│аться            ▓   чувств│овалксь          │                │  │
│  │      │  воздерж│аться        ▓   чувств│овалась          │                │  │
│  │      │  колеб│аться          ▓   чувств│овалось          │                │  │
│  │      │  боро│ться            ▓   чувств│урсь             │                │  │
│  │      │  расста│ваться        ▓   чувств│уехся     ▄▄     │  │
│  │      │  мел│                 ▓   чувств│уеньсь           │ ▄▄             │  │
│  │      │  осипа│ться           ▓   чувств│уетесь           │ ▄▄             │  │
│  │      │  сда│ться             ▓                           │                │  │
│  │      │                                                 │                │  │
│  │      │          OK ▌       Cancel ▌      Help ▌         │                │  │
│  │      └─────────────────────────────────────────────────┘                │  │
│  └──────────────────────────────────────────────────────────────────────────┘  │
└─────────────────────────────────────────────────────────────────────────────────┘

 F1 Help
```

Fig.2. Automatic determining verb forms

The dictionary entry in the PARS-3 system resembles that in traditional "paper" dictionaries: all the idioms beginning with the head word are displayed in the same panel, the head word being substituted with the tilde (~) sign. Figs. 3 and 4 show Russian-English and English-Russian dictionary entries, respectively.

```
 Files  Edit  Search  Translate  Dictionary  Options  Window  Help
┌─────────────────────── Source Untitled ──────────────────────────┐
│                                                                   │
│  ┌─[.]─────────────── Q:\PARS\GEN_ ──────────────[.]┐             │
│  │  ┌────────────────────────────────────┐ ┌───────┐ │           │
│  │  │автоматизированн│ый  1. automated; 2. . │ Entry ▌ │           │
│  │  │machine-aided; 3. automatic;        │ ▀▀▀▀▀▀▀   │           │
│  │  │~ перевод│  1. machine-aided         │ Phrase ▌ │           │
│  │  │translation;                        │           │           │
│  │  │                                    │ Translation▌│         │
│  │  │                                    │           │           │
│  │  │                                    │ Delete ▌ │           │
│  │  │                                    │           │           │
│  │  │                                    │ Save ▌   │           │
│  │  │                                    │      ▀▀   │           │
│  │  │                                    │   Top ▌  │           │
│  │  │                                    │ ▀▀        │           │
│  │  │                                    │ Bottom ▌ │           │
│  │  └────────────────────────────────────┘ └───────┘ │           │
│  └────────────────────────────────────────────────────┘           │
│                                                                   │
└───────────────────────────────────────────────────────────────────┘

 F1 Help   F3 New entry   F4 New phrase   Alt+F3 Close   F10 Menu   Alt+F10 Pop-Up
```

Fig.3. A Russian-English word entry

We have developed a Russian grammatical polytechnical dictionary of 25,000 separate words, with full morphological information, and it is constantly extended. The dictionary will be part of all our Russian-based MT systems. Similar grammatical dictionaries will be developed for all the languages included into the PARS-3 and PARS-4 projects, such as English, Ukrainian, German, Italian, Spanish, French, Dutch, Portuguese.

```
 Files  Edit  Search  Translate  Dictionary  Options  Window  Help
 ┌[.]──────────────────────── Q:\PARS\GEN_ ─────────────────────[.]┐
 ┌──────────────────────────────────────────────┐
 │ be  1. 6|ить; 2. *|;                          ·│  ▐ Entry      ▌
 │ ~ a  1. представля|ть собой;                   │
 │ ~ able  1. способ|ен;                          │  ▐ Phrase     ▌
 │ ~ absent  1. отсутств|овать;                   │  ───────────────
 │ ~ also described  1. также| описыва|ться;      │  ▐ Translation ▌
 │ ~ anterior  1. спереди|;                       │  ───────────────
 │ ~ by right  1. 6|ить по праву|;                │  ▐ Delete     ▌
 │ ~ due to  1. долж|ен;                          │
 │ ~ followed by  1. предшеств|овать;             │  ▐ Save       ▌
 │ ~ fond of  1. люб|ить;                         │  ───────────────
 │ ~ found  1. обнаружен|;                        │  ▐▌    Top    ▌
 │ ~ found to be  1. ока|заться;                  │  ───
 │ ~ found to have  1. име|ть;                    │  ▐▌  Bottom  ▌
 │ ~ going to  1. собира|ться;                    │  ▐▌          ▌
 │ ~ gone  1. у|йти;                              │
 │ ~ in prospect  1. предсто|ить;                 │
 │ ~ known to  1. известно,| что;                 │
 │ ~ likely  1. весьма| вероятно;                 │
 │ ~ mistaken  1. ошиба|ться; ошиб|иться;         │
 └────────────────────────────────────────────┘
 └─────────────────────────────────────────────────────────────────┘

 F1 Help  F3 New entry  F4 New phrase  Alt+F3 Close  F10 Menu  Alt+F10 Pop-Up
```

Fig.4. An English-Russian word entry

The English-Russian-English dictionaries we have developed contain about 200,000 word-entries altogether. They cover general usage words as well as such topics as computers, microelectronics, economy, engineering, some branches of medicine. New dictionaries are being developed, and the existing ones extended.

The system features several translation modes, set up by the user. A text can be translated from English into Russian or from Russian into English; synonyms may be displayed. The whole text may be translated or just its fragment.

Up to 4 dictionaries may be used in each translation session, their priorities set up by the user.

PARS-3 makes extensive use of extended memory (if the computer has no extended memory, the system will work, though slower).

Average translation speed, when 2 dictionaries are used, varies from 10 sec/page on an AT/386 SX to 2-3 sec/page on an AT/486 DX.

PARS-3 is a combination of direct and transfer-based translation philosophies; some semantic categories are used, such as "time", "number", "geographical notion". We are planning to develop a much more powerful transfer grammar, and, which is even more important, an example-based semantic disambiguation routine to avoid some "cats" in the output texts. However, I know but too well that too much transfer is in no way better than no transfer at all; sometimes you have the sentence "transferred" in such a way that it becomes incomprehensible.

As to the translation quality, generally, the evaluations that come from the end users vary from very flattering to rather critical. I developed a 7-point scale for evaluating the output quality, which is based on practical tasks the end user wants to solve, and on common sense. The average mark for PARS output is 5, and if 6 as the average mark is ever attained I will be glad.

In August, 1994, PARS 3.0 was exhibited at The European Conference of Artificial Intelligence (ECAI) in Amsterdam, The Netherlands. The numerous visitors did not speak Russian, so they could hardly evaluate translation quality. Mr J. V. Pinchasik, who is my colleague and representative in The Netherlands, invented what he calls "The Iterative Reverse Translation"

approach, which consists in comparing the source text with the reverse translation of the machine product. Being a bidirectional MT system, PARS can easily do both.

Here are some of the numerous examples Pinchasik provided. :

Original English Texts
Source: Financial Times,
July, 22, 1994

Reverse English Translations

1. IBM earnings rise surprises
   Wall Street.

1. Increase of earnings astonishes
   Wall street.

2. Bundesbank raises hope of fall
   in rates.

2. Bundesbank hoists hope on
   the tumble of rates.

3. Telegraph's chairman in sharp
   attack on Cazenove.

3. Chairman of Telegraph in sharp
   attack on Cazenove.

4. Italy fails to decide economic
   policy.

4. Italy fails decide economic policy.

5. Latin Americans rediscover their
   neighbours' markets.

5. Latin americans discover again
   markets of its neighbors.

6. Ecuador Parliament rejects
   Telecom sale.

6. Parliament of Ecuador rejects
   selling Telecom.

7. UN lawyers take steps to set up
   criminal court.

7. Lawyers UN taking steps in order
   to set up the court of criminals.

8. Japan puts conditions on backing
   Taiwan's bid to join GATT.

8. Japan puts conditions on
   maintaining the proposal of Taiwan
   to affiliate GATT.

Alcatel in China

Alactel in China

Alcatel Australia has signed
contracts, valued at a US$56 sum to
provide telecommunication equipment
to China, agency AP-DJ reports from
Canberra. The company will supply
and install digital switching
exchanges to the autonomous
regions of Tibet and Ningxia and
the province of Gansu.

Alcatel Australia signed contracts
on sum of 56 dollars USA for
the ensuring of China by the equipment
of telecommunication, discloses
agency AP-DJ from Canberra. Company
will supply and establish the digital
switching of exchanges for
the autonomous districts of Tibet and
Ningxia and for province Gansu.

ACKNOWLEDGEMENTS