



In 1993, I set up my own private company, which I named *Lingvistica '93*. It resided in Kharkov, the second largest Ukrainian city, and developed machine translation '98 Inc., and it resides in Montreal, Canada, where my family and myself came in 1998.

Our next MT system, PARS/Ukrainian-English (PARS/U), was to some extent the result of my reading the paper by Boyan Onyshkevych in a book on machine translation edited by Sergei Nirenburg, my University friend and colleague. That paper described an English-Ukrainian prototype MT program, and I decided to make a system to translate between Ukrainian and English. In 1992 we were financed by the computer department of the Ukrainian Supreme Rada and came up with the world's first Ukrainian-English-Ukrainian operational MT system. Now its newest version is marketed in North America by the Montreal-based *Yevshan* company. It was presented at the AMTA conference held in Montreal in 1996, at MT Summit VI, in San Diego, in 1997 [4,6], and at the University of Toronto in 1998.

The DOS version of PARS/D, the German-Russian system, was created in 1994 on the order of the *Izvestia Concern*, Moscow, to translate VWD information messages. Unfortunately, we failed to complete the project for a number of reasons. In 1997, however, we resumed the work due to financing from the Hamburg-based *Igor Jourist Verlag*. PARS/D translates between Russian and German, both ways.

The PARS/DU Ukrainian-German-Ukrainian system, again the world's first for this language pair, was developed in 1998 in the framework of the KOPERNIK project launched by the Ukrainian Ministry of Education.

In 1999 we launched another PARS project, a German-English bidirectional system.

Being the author of 5 MT projects, I am often asked a philosophic question: «What is translators' attitude to MT in general, and to PARS, in particular?» Well, before joining the cohort of machine translators, I became a translator. My major achievement was translating *Alice's Adventures in Wonderland* from English into Russian. So my first impulse was to say that being a translator is by no means an advantage in developing an MT system because translating is art, while you can't make anyone, including a computer, an artist. In other words, MT...is not translation. I remember very well translating *Alice's Adventures*. You know what I had to do? Trying to make the story funny and amusing, I had, more often than not, to invent, not even translate!

At the same time, being a translator is a great thing for a machine translator. I always understand very well my colleagues' requirements, and I am ready to hear that a translator is disappointed by the numerous childish mistakes PARS makes. I also think that very often translators prefer extensive electronic dictionaries to an MT program [8]. That's why I am always pleased when my system is praised by a professional translator. And the paradox lies in the fact that it's the most skilled and

and machine-aided translation systems as well as computer-based dictionaries. At present, it is *Lingvistica* experienced translators who find PARS useful in their everyday work.

One of the brightest examples of a translator's attitude I have ever experienced was PARS presentation at the Antonov Aviation Plant in Kiev. I was surrounded by a group of brilliant professionals who were watching, somewhat skeptically, the computer screen while PARS was busy translating a technical text from Russian into English. They analyzed the result attentively, and I asked one of them: «What do you think about it?» What he said amazed me: «Well, it translates like a student». «A fresher or a senior?» I asked him. He thought a little and said, smiling: «Like a sophomore» What he meant was that the translation was quite understandable but rather primitive. «You are flattering me!» I replied. «The student is human, while the computer is not».

Well, I did use my translator's experience when I designed PARS. The peculiarities of our systems consist in the service options rather than in the translation algorithms, and the former were introduced because I am a translator.

## 2 Technological

Speaking of a machine translation system, as I understand the problem, one may describe it from three angles:

- how the system looks, that is what you can see when you run the program on your computer;
- how the system works, that is what kind of translations it comes up with;
- how it does what it does, that is what algorithm the system uses for translating texts.

From the technological point of view, this paper is focused on dictionary support and updating, which is PARS's most characteristic feature. Other aspects have been discussed elsewhere [1-6].

It would hardly be an exaggeration to say that one of the most important criteria of evaluating a commercial MT system is its dictionary support subsystem the easier it is to extend dictionaries supplied with the system as well as to create user's dictionaries, the better the system is in general.

1) Dictionary entries in MT systems by *Lingvistica '98* remind those in traditional dictionaries.

2) Dictionaries in *Lingvistica '98* systems are bidirectional. For example, if the user enters an English word with its Russian translation into a PARS dictionary, the system automatically sets the opposite correspondence, Russian-English. Accordingly, any dictionary can be browsed and edited by any part, for example, English-Russian or Russian-English. At the same time, there is an option that lets the user disable bidirectional correspondence wherever necessary.

3) A word/phrase can have a practically limitless number of translations, which permitted us to implement choosing translation variants in the target text. The customer may use the one-keystroke transposition option in the dictionary entry assigning a higher priority to the translation which is considered the most likely one for the subject area the dictionary relates to.

4) The systems feature automatic indexing (tagging) of Slavic words being entered into the dictionary: the dictionary support program automatically assigns grammatical characteristics to them, such as part of speech, declension, conjugation, subclass characteristics (such as gender). If the program doubts how to tag a word, the user makes a choice out of two or more options.

PARS includes a large spectrum of English-Russian-English specialist dictionaries, the subject areas being technology, business, medicine, space engineering, electronics, mathematics, chemistry, automobile building, etc., as well as an exceptionally large general dictionary. The total number of translations as of June 1999 was above 1,000,000 words and phrases in each part - English-Russian and Russian-English.

And the last but not the least *Lingvistica '98* dictionary officers make use of our proprietary batch tagging technology, which ensures assigning grammatical characteristics to Russian and Ukrainian words in a batch mode within very short periods of time. For example, it takes one 2-3 hours to 'encode' an English-Russian dictionary of 50,000 translations.

An important technological aspect is practical usage of a system. Speaking of PARS users, I can subdivide them into the following groups:

#### **Individual users**

A very numerous subgroup is made up by students who need their diplomas and other kinds of papers to be translated from Russian into Ukrainian. We hope to meet their requirements with our cheap KOPERNIK CD.

Some people want to communicate with people living abroad. PARS/U is bought, in particular, by Americans and Canadians wishing to communicate with their friends and relatives residing in Ukraine. Usually these people tell me. «They speak Ukrainian, and I speak English. The only way to communicate is to use a computer program». I wonder if one of the international pen pal organizations might be interested in using PARSes for communication purposes. It would certainly require serious modifications in order to take into account peculiarities of the 'friendly communication' style, but the idea itself seems to me rather promising.

Professional free-lance translators in ex-Union make up another subgroup, though less numerous. Their language pairs are mainly English, German, French, Italian to and from Russian. Some of them like MT systems, some prefer MAT software (electronic dictionaries such as PG-PARS [8]), while others buy both. My opinion is, however, that the majority of this group are still our potential clients. The fact is that the foreign languages departments of Ukrainian universities train people who are good at languages but have no idea of the computer as translator's everyday tool. Introducing elements of language engineering at such departments

would contribute a lot to expanding the circle of our conscientious customers!

There is a group of individual users who require Russian to English translation of scientific texts. Here is an example. A scientist asked me to translate his medical paper for submitting it to a serious British journal. When I looked through the text, there was only one thing I understood: I could not do without PARS because the paper was abundant in «awful» medical terms. I faced a dilemma: either to translate the text manually looking every second or third word up in a Russian-English medical dictionary, or to let PARS make a draft translation and post-edit it. I chose the latter variant, and the paper was accepted.

#### **Corporate users**

MT and MAT systems seem to be rather popular with corporate users.

Generally speaking, all kinds of organizations, both state-owned and private, use PARS/RU for translating official documentation, including that of financial, scientific, and technical nature, between Russian and Ukrainian.

Many Ukrainian banks use PARS/RU for translating financial documentation, such as official instructions, between Russian and Ukrainian. Here is another example. In 1997, I installed PARS/RU at one of the banks in the town of Saki, the Crimea. They use it to translate megabytes of instructions they receive electronically from the Ukrainian National Bank. Those texts are written in Ukrainian, the country's state language, but the problem is that many people in the Southern and Eastern parts of Ukraine doesn't even understand Ukrainian, to say nothing of speaking it.

A tendency that gains popularity is making MT systems part of integrated products, such as PRAVO, a system very well-known in Ukraine. It is supplied on CD-ROM and comprises the full set of Ukrainian laws and decrees, with a retrieval system and our Ukrainian to Russian, English, and German translation modules.

I am especially proud that PARS was used for a long time for translating Russian medical abstracts into English for the *Medical Practice* journal published in Kharkov. I did it myself, first running the texts through PARS and then post-editing the raw translations. Using MT systems for translating abstracts in scientific journals may become a good tendency.

Large plants and design bureaus that export their products are among the users of the PARS/ER system. The Yangel Spacecraft Bureau in Dnepropetrovsk is among them. We supplied PARS/Aviation to them, which includes the core Russian-English-Russian system and a number of terminological dictionaries on aviation, space, communications, etc. Their reaction is very important for me: they say that PARS is better for translating technical documentation, while the well-known *Stylus* system (designed by another group of Prof Piotrowski's pupils) is preferable for business correspondence.) Well, PARS will be trying to be up to the mark in both aspects!

A new tendency is the application of PARS to translate Russian textbooks and courses of lectures into English for foreign students coming to study at Ukrainian universities. A vivid example is described by my elder

daughter, Olga Bezhanova: we translated Russian texts on aviation for the Iranian students of Kharkov State Aviation University [7].

MT can and should also be used for purely academic purposes. An example is using PARSES at Kharkov State Polytechnic University in the course of machine translation at the Department of Intelligent Information Systems.

Generally speaking, our experience shows that post-editing PARS-made Russian to English translations is about 3 times more efficient than translating the same text manually.

PARS/RU translates texts in such a way that they are 70-80%, sometimes even 90% ready for publication, the quality of Russian-Ukrainian translation being somewhat higher than that of Ukrainian-Russian. As to the other PARSES, they are used to:

- let the user grasp the general idea of the document, for example, when browsing large databases,
- create a draft for subsequent polishing, i.e. for turning the draft into a translation.

### 3 Linguistic

It seems to me that a commercial MT system should only have minimum (or a little more than minimum) data attached to the words in the dictionary; it has to be a compromise between the system designers' desire to develop a powerful linguistic tool, with maximum information assigned to the words (which would lead to high-quality translations), and understanding that such a tool will be useless from the customer's point of view if working with it would require too much effort from the user when extending the dictionary. In other words, the system should be powerful and easy-to-operate and customize, this desire bearing a strong intrinsic contradiction.

The problem is that a powerful linguistic apparatus, which is a necessary prerequisite for obtaining high-quality output, requires a lot of semantic information in the system dictionary. That is, the words entered into the dictionary must acquire special semantic notions, which will make the system really «clever». In this case, the words will be described not only as parts of speech that can have such and such endings, but their senses will also be represented. Here is an example.

The minimum information for the word *dog* is that it is a noun, and its plural is *dogs*. Besides, it's animate. However, much more can be said about it, for instance, that a dog is an animal. The latter is also important. Let's take a nice sentence provided by one Japanese linguist: *I saw a dog with a telescope*. The program will not understand this sentence if it doesn't know that a dog is not human and that's why dogs can't use telescopes.

It's not easy to develop such a dictionary. Mind that the system doesn't simply exist the way it was purchased. No, many users will necessarily want to customize it by entering new words into the existing dictionaries and creating new dictionaries. In this case, the user will have to know as much about semantics as the linguist who developed the system. And this may make the system

practically useless, for example, for an engineer who needs translations of texts in his/her subject area, but knows nothing of semantic categories of the words entered into the dictionary. And, as the experience and common sense show, it's hardly possible to write an algorithm for automatic assigning semantic categories to the words being entered into system dictionaries, so it would be up to the user!

On the other hand, too little information is as dangerous as too much of it since the system should not only be convenient, but also practically useful. Bearing on too little linguistic data, the system will not be able to give comprehensible translations in some cases, no matter how easy it may be to enter new words into it.

That is why the authors of an MT system have to determine the scope of linguistic information the system will really not do without, balancing between «too little» and «too much».

So, what translation philosophy is laid in the foundation of PARS?

First I wanted to make use of the «almost-classic» definition of three translation approaches: direct, transfer-based, and interlingua-based. But the more I was thinking about all this, the more convinced I became that it would hardly be possible to use this definition **practically** since it is very hard to draw a demarcation line between the three approaches. The most important reason is that the champions of this definition consider what they call «direct translation» quite fruitless, while, on the other side, systems translating «directly» are sold and, what is more important, bought throughout the world, giving their developers honestly earned profits, the latter being sometimes rather high. Or maybe we have to admit that no «pure direct» translation systems really exist, and each system is a combination of two or three philosophies, so a different kind of terminology should be suggested.

In each of our PARSES, the translation program first generates a word-for-word translation, and then brushes it up intensively, making it look as natural as it (the program) can. That's why I call our approach FTA - «first-translate-then-analyze».

Generally speaking, FTA is usually resorted to if system developers don't want to view the sentence as a single structural entity, considering it as a linear sequence of lexical units and analyzing some syntactic and semantic relations for disambiguation purposes only.

On the contrary, a system may first analyze the source text, and then translate it, using the results of this analysis, thus working according to the FAT - «first-analyze-then-translate» principle. Traditionally, the FAT-type systems consider the whole sentence as a syntactic (or even semantic-syntactic) unit, the basic idea being that the more information you use in your analysis, the better results you will obtain.

It should be taken into consideration, however, that mistakes in the target text are practically inevitable, or at least highly probable in each case, that is «when she (the translation algorithm) is good, she is very, very good, but when she is bad, she is horrid»: mistakes made in analyzing as complicated entity as a sentence will cause translation mistakes.

This situation is but very well known to practical developers of language engineering systems, who constantly face the «noise/recall(completeness)» dilemma. From time to time, we come across the typical situation: too much analysis causes poorer translation quality than no analysis at all. My opponents may contradict that «too much analysis» means «too little analysis», but have you ever seen enough analysis in real-life MT systems?

PARSes bear on hundreds of rules to analyze the source text and synthesize the target one, some of the rules being rather sophisticated, such as disambiguation of *-ed* forms in English. However, the programs don't dare to view the sentence as a structural unit. The program only analyzes a word if it is grammatically ambiguous. At the same time, the set of rules is constantly extended in the system «growing» process: we analyze translation results, and if a mistake is typical, i.e. a certain ambiguity type is come across regularly, we think of a rule to eliminate the ambiguity. And sometimes we do find such a rule.

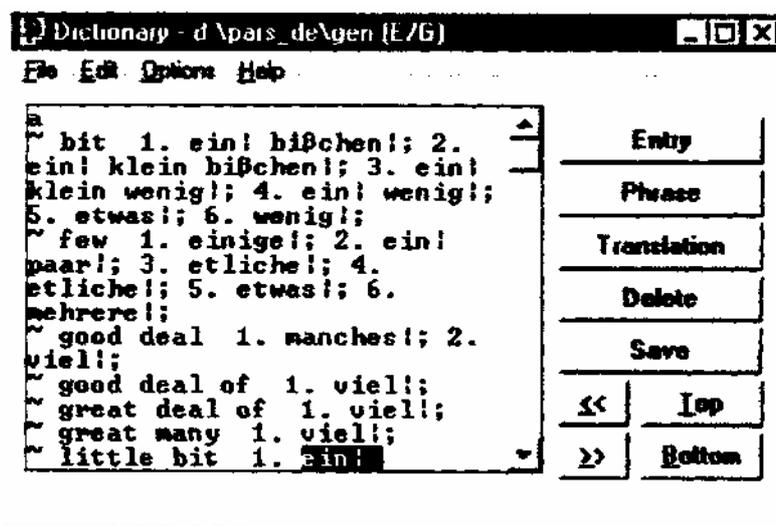
So, here is a general outline of the translation procedure in PARSes.

Stage 1. The system makes what is called word-for-word and phrase-for-phrase translation of the source text, recognizing phrases and single words and extracting the corresponding grammatical data from the dictionary. This is done using the morphological analysis rules. For example, when analyzing English texts, a table of irregular verbs is made use of, as well as a set of rules for recognizing noun plural forms; the German morphological analysis is based on the rules of linking German separable prefixes to their corresponding verbs; Slavic words are recognized in the source text due to special tables of Russian and Ukrainian paradigms.

Stage 3. The system generates the target text. The task consists in making the target sentences look as natural as possible. The system tries to insert articles (which is even a difficult task for some humans, to say nothing of an algorithm), changes word order, etc.

Let's call a spade a spade: if the grammatical structures of the source and target languages are not so much alike as, for example, those in Russian and Ukrainian (although the Russian and Ukrainian grammars have a lot of differences), the output texts are very far from those made by qualified translators. When I hear or read that an MT system ensures «80-90-percent accuracy», I am inclined to consider such a statement a mere advertising trick, especially speaking of such different languages as Germanic and Slavic. Yes, machine grammars are being constantly improved, but, being a professional language engineer, I can hardly imagine that computer programs will ever be able to compete with qualified humans. Or maybe I am mistaken? People used to think that a computer would never be «cleverer» than a human chess player, but *Deep Blue* beat Kasparov...

An important peculiarity of MT systems by *Lingvistica '98* is what we call distant phrases, an idea I gained from my unforgettable hours-on-end talks with one of my teachers, Dr Boris Pevzner. He considered it unrealistic to list all possible phrases in a dictionary, no matter how large the dictionary may be. It would be more reasonable, he said, to enter a typical (model) phrase and a rule for making substitutions so that the system could generate phrases similar to the model. It means that a phrase in this case will not be something fixed, but rather a flexible unity similar to many other phrases. Unlike fixed phrases, the elements of which are always adjacent, such as *in order to*, *door handle*, etc, a distant phrase may have a 'gap', for example, *pay...attention*, we can see



Word entry in the English-German-English PARS

Stage 2. The system analyzes the resulting text and makes its best to eliminate as many ambiguities as it can. When doing so, it makes use of special contextual rules for grammatical and semantic disambiguation.

definite words in real-life texts instead of three dots, such as *pay great attention*, *pay extraordinarily serious attention*, etc.

So far, all our MT systems only recognize one kind of distant phrases. A distant phrase is considered to be a 2-

word source-language phrase having a 2-word translation.

Two 'gap' types are distinguished:

- positional: in a text, not more than 5 words may appear between the left and the right words;
- grammatical: no other part of speech may occur between them but an article, one or more adjectives and/or adverbs.

So, as we have seen, the words in the dictionary acquire grammatical characteristics. Some semantic information is also assigned, such as *Time*, *Geographic notion*, *Name*, etc., which makes it possible for the program to determine the right meaning of a word in context.

However, semantics is something an ordinary end-user can hardly cope with when extending the dictionary or compiling one of his/her own. That is why our systems will even translate if no semantic marks are present, though maybe a little worse.

And now, just to illustrate the above, here is how three PARSes translated the same text, the Ukrainian Declaration on the State Sovereignty. In fact, the original text was in Ukrainian (just all right for PARS/U and PARS/DU), and I also had it translated by PARS/RU into Russian and edited the translation manually, after which the Russian text was translated into English by PARS. Mind that no translation variants, although provided by the systems, are not displayed in the illustrations.

Ukrainian to German:

Die Deklaration über die staatlichen Souveränität der Ukraine.

Der Oberste Rat Ukrainischen SSR, aussprechend den Willen des Volkes der Ukraine, strebend schaffen die demokratischen Gesellschaft, hervorgehend vom den Bedarfen der allseitigen Versorgung der Rechte und den Freiheiten dem Mann, verehrend die nationalen Rechte der allen Völker, sorgend um hochwertigen politischen, ökonomischen, sozialen und die geistigen Entwicklung des Volkes der Ukraine, bekennend die Notwendigkeit des Aufbaues des Rechtsstaat, habend Ziel aufnehmen die Souveränität und die Selbstverwaltung des Volkes der Ukraine, proklamiert die staatlichen Souveränität der Ukraine als die Vorherrschaft, die Selbständigkeit, die Fülle und die Unteilbarkeit der Gewalt der Republik innerhalb ihren Gelände und die Unabhängigkeit und die Gleichberechtigung in dem äußerliche Verkehr

Ukrainian to English:

Declaration about the state sovereignty of Ukraine.

Supreme rada Ukrainian SSR expressing freedom the nation of Ukraine seeking to create democratic society proceeding from the needs of the all-round provision of rights and the freedoms of man respecting national right all nations caring about complete political, economic, social and spiritual development the nation of Ukraine accepting the necessity of the construction of legal

country having aim to affirm sovereignty and self-governance the nation of Ukraine, declares the state sovereignty of Ukraine as supremacy, independence, amplitude and the indivisibility of the authority of Republic within it territory and self-support and equality into foreign communion.

Russian to English:

Declaration about the state sovereignty of Ukraine.

Supreme Soviet Ukrainian SSR expressing the will of the nation of Ukraine, aiming to create democratic society, based on needs of the all-round provisioning of rights and freedoms man respecting the national rights of all nations attending to full-value political, economic, social and spiritual development of the nation of Ukraine recognising the necessity of the building of legal state having purpose to affirm sovereignty and the autonomy of the nation of Ukraine, proclaims the state sovereignty of Ukraine as supremacy, independence, completeness and indivisibility the power of Republic within its territory and independence and equal rights into exterior relations.

Well, we did put our love into those systems, and that is why the love story will last very long.

#### REFERENCES

- 1 M.Blekhman How Lingvistica '93 MT Systems are Used (1991-1993) - MT News International, No7, 1994.
2. M.Blekhman. Through the Looking Glass to the PARS Land. - Machine Translation: Ten Years On. Proceedings of the International Conference, Cranfield University, 1994.
3. M.Blekhman News from Lingvistica '93 Co. MT News International, No 12, 1995.
4. M.Blekhman PARS and PARS/U Machine Translation Systems: English-Russian-English and English-Ukrainian-English. - Expanding MT Horizons. Proceedings of the Second Conference of the Association for Machine Translation in the Americas 2-5 October, 1996, Montreal, Quebec, Canada.
5. M.Blekhman. Language Engineering Systems by Lingvistica '93 and ETS Publishers Ltd.: English, Russian, Ukrainian - Machine Translation Review, No. 5, 1997, UK.
6. M.Blekhman et al. PARS/U: The World's First Commercial Ukrainian-English Machine Translation System. - Proceedings of MT Summit VI, 1997, San-Diego, Ca.
7. O.Bezhanova. Translating Scientific Texts Using MT and MAT Systems: Practical Experience of a Professional Translator - Proceedings of MT Summit VI, 1997, San-Diego, Ca.
8. M.Blekhman et al. Machine-Aided Translation Tools for Slavonic Languages - Machine Translation Review, April 1998.

Short stories about family, including brothers, sisters, fathers and sons, mothers and daughters, fathers and daughters, mothers and sons & grandparents. Looking for love and attention at home, she engages in petty crimes, which escalates to her running away. Read "How I Contemplated" Oliver's Evolution | John Updike. He relates some experiences with her during her later years, including how they would go hunting in the neighbourhood for glass fragments and old jewelry. Read "The Jade Peony". The Medicine Bag | Virginia Driving Hawk Sneve. "Love Story" is a song written and recorded by American singer-songwriter Taylor Swift. She also produced the track with Nathan Chapman. It was released on September 12, 2008 by Big Machine Records, as the lead single from Swift's second studio album Fearless (2008). Considered as her signature song, it was written about a love interest of Swift's who was not popular among Swift's family and friends. Because of the scenario, Swift related to the plot of William Shakespeare's Romeo and Juliet (1597) A peak into Dolly Parton's 50-year love story. June 9, 2016. 756 Views. Marley McMillan. 0. Awww this is sweet. Take a look at Dolly Parton's 50-year love story. She and her husband Carl Dean just renewed their vows. Happy anniversary! Inside @DollyParton's 50-year love story: <https://t.co/111MrWkcPZ> pic.twitter.com/QYvxGW5NUA. " Vogue Magazine (@voguemagazine) June 9, 2016. 0. Dolly Parton. Marley.