METAL RUSSIAN-GERMAN ON THE WAY TO A PRODUCT

I. Höser, G. Klimonow, A.Küstner and B. Rüdiger
Gesellschaft für multilinguale Systeme mbH
Glinkastr.26, D-10117 Berlin
Fax: 0049 30 2042048 e-mail: kli@gmsber.de

The MT system prototype METAL Russian-German,
first time presented to the public in April,
1994, is planned to be developed into a product
by the end of 1996. This includes complete cover-
ing of the syntactic phenomena of Russian rele-
vant for draft translations of documentations
and other technical and political texts, expand-
ing the Russian and Russian-German lexicons, pro-
viding users with helpful tools, increasing
robustness of the system, and reimplementation
of the system for running on PC.

INTRODUCTION

The METAL Russian-German MT prototype was presented at last
year's Cranfield Conference "Machine Translation Ten Years On",
but only in form of a paper, as we missed the chance of demon-
strating it. Now we will not only describe the progressing work
on our system but also demonstrate the current state of it.

The development of the Russian-German MT system started in
1991 and reached its first stage with the presentation of a
prototype in 1994.

From May, 1994, to June, 1995, work mainly concentrated on
expanding the Russian and Russian-German lexicons with regard to
general, common technical, and special vocabularies (subject
areas: nuclear power and safety problems, law, railway
transportation, banking and others). Meanwhile, the size of the
Russian and Russian-German lexicons has reached more than 32,000
entries, each. For the next four years basic research on the
project of Russian-German and German-Russian MT will be
sponsored by the Federal Ministry of Education and Research
(BMBF).

The development of the system into a product shall be
completed by the end of 1996; after that the reverse option
German-Russian will be developed.

WHAT DO WE UNDERSTAND BY "ON THE WAY TO A PRODUCT"?

The METAL system had been originally developed at Austin university. Only later on it came into the hands of developers who tried to give it a commercial outfit. But there was a hard race between the large software package and the enormous progress in hardware and base software. Thus the system was ported from FORTRAN (running on main frame computers) to Symbolics Common LISP running on LISP machines.

Although these machines had hightech quality, they turned out to be technological outsiders. A quick port to a computer type more frequently used than Symbolics LISP machines and having a chance to keep pace with future developments was necessary. Thus METAL was ported to SUN workstations. The port of the Russian-German prototype followed immediately. Some efforts had to be made to teach the available software how to handle the Cyrillic character code.

However, the development of hardware and base software is quickly progressing, and the demands of the customers are increasing. SUN workstations belong to the standard inventory of universities and larger companies, but they cannot be afforded by the average customer. Besides, those who want to use MT wish to use it in their familiar environments which nowadays means Windows on a PC.

As a first step, a network version with a so-called PC-client for text handling was introduced. The actual translation job had to be sent from the client to the network server. To overcome this division of functions between server and PC-client it is necessary to port the whole system. Therefore, the next port is under way: by 1996 the translation process itself shall run on PC, based on C++ software. Different product levels will be offered:

• a PC standard version for the less ambitious user
• a PC expert version which offers sophisticated coding tools and other facilities
• a High-End version where the dictionary database is managed on a central UNIX server (SUN or e.g. Windows NT), but the translation itself is done on PC.

The latter version is intended for larger translation offices.

So far we mentioned some hardware and software prerequisites of a product, there are still other/properties necessary to make a system marketable, as robustness, user friendliness, portability.

Robustness

What makes an MT system robust?

- A grammar which is as complete as possible but which is also able to handle text types such as headlines, lists, incorrect or incomplete clauses.

- Huge dictionaries which cover nearly all words belonging to the common and common technical vocabulary which may occur in texts.

- A mechanism to handle any words which are not yet included in the dictionary.

During the last 5-6 years, the Russian language has undergone enormous changes, especially with respect to vocabulary, so that lexikographers can hardly keep pace with this development. Therefore, the systems have to be able to treat unknown words. The prototype METAL Russian-German includes a first version of a so-called lemmatizer or on-line defaulter. This is a tool which processes words unknown to the system with the aim of reducing inflected forms to their basic forms and of anticipating grammatical information about the unknown words. This enables the system to recognize the grammatical structure of the sentence, to solve other ambiguities and to translate the rest of the sentence as correct as possible. Furthermore, one or more hypotheses of the canonical form of the unknown word are offered which (along with hypotheses about grammatical information) shall be used as input for lexicon coding.

- Stable software: Although daily system crashes of the Symbolics machines belong to the past, it is necessary to find out various possibilities of incorrect operations of the users. The system should be able either to ignore these operations or to react to them by simple instructions how to continue. This constitutes an important aspect of user friendliness.

Translation quality

The developers are aware of the fact that translation quality is one of the key factors for a customer when choosing a special product.

METAL is designed as a system with a refined analysis and a high modularity which allows to freely combine lingware rules and procedures, and to adapt the structure of the dictionary entries to the requirements of the actual language. This is a good basis for the lingware developers to achieve an acceptable^ translation quality. But nevertheless new approaches must be

looked for as e.g. an effective integration of a translation
memory.

Prices

      Up to now METAL was one of the quite expensive systems with
respect to the necessary hardware and software. Porting the
system from Symbolics to SUN workstations did not yet lead to
the necessary break-through, the same holds for the PC-clients
within a network. At present, translation services in public
network are being tested which are already low-priced. But only
the future PC version will achieve a good ratio of price and
performance so that the system will have a chance to meet with a
ready sale on the market.

In the following, we want to show how special characteristic
Russian linguistic phenomena are treated by our prototype.

STRUCTURAL PECULIARITIES OF RUSSIAN

Russian is the first Slavic language among the METAL languages
treated so far, and it is in many aspects different from them.
Some structural divergences between Russian and German are the
following:

Morphology

      The Russian morphology has a rich inflectional system,
e.g.:
- six cases of nouns and adjectives for each number;
- cardinals inflect like nouns;

- verb forms in the past tense vary in gender and number:
    он видел_ – er sah (he saw)
    она видела – sie sah (she saw)
    они видели – sie sahen (they saw) ;

- in case of hyphenated compounds both head and specifier can
  inflect, e.g.
    ракета-носитель – Trägerrakete (carrier rocket)
    вес ракеты-носителя – Gewicht der T. (weight of carrier rocket)

- Russian verbs often have an alteration of consonants in their
  stems (e.g., наградить - награжу - награжден auszeichnen (to award). In
  the dictionary only one stem (награ) is coded which is supplied
  with information about possible continuations (ж-д-жд)together
  with the inflectional endings and participial and gerund

suffixes. The same principle is used for verbs without such alterations in order to restrict verb coding to a minimum of effort.

Syntax

There are various syntactico-semantic peculiarities:

- <u>Category of Verbal Aspect</u>. Russian has a grammatico-semantic category of verbal aspect for expressing the temporality of events from the point of view of the speaker (in German such a category is missing). Thus the perfective and imperfective aspect forms respectively influence the choice of the verbal tense and the conjunction of the subclause, e.g.:

  Как только он ушел, стало скучно.
  -Kaum daß er gegangen war, wurde es langweilig.
  -(As soon as he had left (the party) it became boring.)
  vs. Как только он уходил, становилось скучно.
  -Sobald (immer wenn) er ging, wurde es langweilig.
  - (Whenever he left (the party) it became boring.)

- <u>Definiteness/Indefiniteness</u>. In Russian articles are missing. Therefore, it is very difficult to generate correct articles in German. Russian, however, has other means of expressing definiteness or indefiniteness, e.g. verbal aspect, use of partitive genitive or marked word order reflecting theme-rheme relations. Inversion of surface subject, followed or not followed by relative clause, and predicate points to indefiniteness of the subject noun phrase.
  Example:
  Проводились работы по снижению воздушного загрязнения.
  -Arbeiten zur Senkung der Luftverschmutzung wurden durchgeführt.
  (Some work with the aim of reducing air pollution was done.)

The extraction and processing of the above kinds of information requires further investigation.

There are, however, also cases where article insertion depends only on the structure of the respective noun phrase. Proper names of persons normally have no article, but if they are preceded by an attributive adjective the definite article has to be inserted:

  Елизавета - Elisabeth
  vs. Святая Елизавета - die heilige Elisabeth (the holy Elisabeth)

Titles followed by a name do not get an article, whereas titles alone or with a following genitive phrase must have an article:
директор Иванов - Direktor Ivanov

vs. `директор машинного завода` - der Direktor der Maschinenfabrik
(the director of the machine factory)

If the subject of a sentence dominates an attributive conjunctional clause introduced by `что` (that) , the subject gets the definite article:

`возникает впечатление, что...` - es entsteht der Eindruck, daß ...
(one gets the impression that ...)

- <u>Restricted Anaphora Resolution</u>. Besides the possessive pronouns existing in other West European languages, Russian has a second type of pronoun `(свой)` which characterizes the relation to the subject of the sentence.
Example:
<u>Я</u> послала <u>своему</u> брату компьютер, а <u>вы</u> послали ему <u>свои</u> поздравления.
- <u>Ich</u> schickte <u>meinem</u> Bruder einen Computer, und <u>Sie</u> sandten ihm <u>Ihre</u> Glückwünsche.
(I <u>sent</u> my brother a computer, and <u>you</u> sent him <u>your</u> congratulations.)

In various types of relative clauses with the relative pronouns `который`, `чей` and `что` the antecedents are correctly identified.
Example:
Завод, руководство <u>которым</u> было поручено Иванову, разработал новый тип автомашины.
-Das Werk, <u>dessen</u> Leitung Ivanov übertragen wurde, entwickelte <u>einen</u> neuen Fahrzeugtyp.
(The company the management <u>of which</u> was handed over to Ivanov developed a new type of car.)

- <u>Double and Multiple Negation</u>. In Russian sentences several negated constituents (marked as NEG in the examples below) can occur which all but one have to be switched to their affirmative antonyms during transfer.

Example for double negation:
Никто (NEG) не (NEG) спит.
-Niemand schläft.
(Nobody is sleeping, vs. *Nobody is not sleeping.)

Example for multiple negation:
Никто (NEG) из нас никогда (NEG) ни о чем (NEG) с ним не (NEG) говорил.
- Niemand von uns sprach jemals über irgendetwas mit ihm.
(None of us ever talked to him about anything.
vs. *None of us never not talked to him about nothing.)

REFERENCES

1. Апресян, Ю. Д. et al., 1989. Лингвистическое обеспечение системы ЭТАП-2, Москва.

2. Höser, I., and Klimonow, G., 1994. Russisch-Deutsch: ein
    neues METAL-Sprachpaar. In: Sprache und Datenverarbeitung,
    Vol. 18, 1/1994, pp. 53-64.

3. Höser, I., and Rüdiger, B., 1994. Russian - a challenge to
    METAL? In: MACHINE TRANSLATION TEN YEARS ON, 12-14 November
    1994, ed. by D. Clarke and A. Vella, The British Computer
    Society, Cranfield, pp. 18-1 - 18-13.

4. Iomdin, L. I., 1994. Automatic Syntactic Analysis of Russian
    in the CAT2 MT System, IAI Working Papers No. 32,
    Saarbrücken.

5. METAL Russisch-Deutsch. Prototyp eines maschinellen
    Übersetzungssystems, GMS Berlin, Internal Documentation,
    April 1994, 600 pp.

6. Мельчук, И. А., 1985. Поверхностный синтаксис русских числовых выражений,
    Wiener Slawistischer Almanach, Sonderband 16 (Linguist.
    Reihe, hrsg. v. T. Reuther), Wien, 509 pp.

7. Schubert, K., 1994. Sietec's Metal (Also) Does Russian. In:
    Language Industry Monitor, Amsterdam, May-June 1994, Issue
    Number 21, pp. 10-11.

8. Thurmair, G., 1990. METAL Computer Integrated Translation.
    In: J. McNaught (ed.), Proceedings of the SALT Workshop
    1990, Manchester.

9 . Зализняк, А. А., 1980. Грамматический словарь русского языка. Москва: Русский
    Язык, 890 pp.