

VIRTEX - a German-Russian Translation Experiment

B. Buschbeck, R. Henschel, I. Höser, G. Klimonow, A. Küstner, I. Starke

Zentralinstitut für Sprachwissenschaft
Akademie der Wissenschaften der DDR
Prenzlauer Promenade 149-152, Berlin, 1100, DDR

Machine Translation from German into Russian did not gain the attention it deserves so far. Especially such typical Russian phenomena like the aspect of verbs complicate the translation process. The known MT approaches do not deal with such problems in a satisfactory way. To overcome this lack, we started the German-Russian translation experiment VIRTEX. We focussed our attention on the crucial point - a high quality translation of verbal groups (VIRTEX = Verbs-into-Russian Translation Experiment). That means, VIRTEX is made to solve on principle the linguistic difficulties connected with verb translation:

- recognition of German analytical tense forms, passive constructions, reflexive forms and verbs with detached prefixes
- analysis of complex German verb groups (modal and modal-like constructions)
- selection of Russian equivalents according to valency frames including semantic features of verb complements
- choice of the appropriate Russian aspect form
- passive-active transformation in case of missing Russian passive form
- adequate translation of modal constructions.

The used vocabulary, above 200 sample verbs, is selected to represent these phenomena completely.

To translate verbal groups adequately, it is necessary to have some context. Therefore we accept as input simple sentences with verb complements and adverbials. That means we have also a set of

nouns, pronouns, adverbs and some other words. The restricted syntax model is made for short main clauses; it excludes relative and other subordinated clauses, attributes and infinitive clauses with complements. VIRTEX can be classified as an MT system between second and third generation with a syntactic interface structure completed by some semantic information. The system's architecture is the classical one. Analysis, Transfer and Synthesis are sequentially processed separate routines. For each of them an extra lexicon has been implemented. The input-work is supported by special tools including an automatic generation of the inflectional types. While the procedures are written in LISP, the dictionaries and their editors are made in dBASE. The format restrictions of dBASE are compensated by a special management of the entries and a corresponding way of LISP access to them. VIRTEX runs on PC-XT.

Some steps of the system will shortly be illustrated.

ANALYSIS

The lexical entries for German syntactic analysis are taken from the analysis dictionary which is combined with a system of inflection lists; an analysis of inflectional endings is embedded in the dictionary access. A word form entry contains word class information, morphological features and values, some semantic information and the name of the lexeme. While the system is running, the lists of information and features are converted

into numbers.

For syntactic parsing an ATN-grammar is used to recognize nominal and prepositional phrases, adverbials and verbs or verbal groups. A special embedded program analyzes the verb constituents. Homonymous morphological and word class features are disambiguated as far as possible throughout the parsing process.

TRANSFER

All transfer rules are written within our tree transformation language GRACOLI. This is a universal software tool developed to handle MT transfer problems. GRACOLI manages the input and processing of tree rewriting rules. One rule consists of a left side and a right side tree. According to the usual notation of syntactic trees, every tree node can be labelled by a finite set of attribute-value-pairs. So we have got a universal rule format for structure changing with assignment of new labels depending on the occurrence of the left side structure and left side label conditions. It includes all simpler transfer actions like label tests, structure tests, lexical transfer etc. The compilation of GRACOLI rule sequences in GRACOLI grammars allows high evidence and easy correction of implemented transfer algorithms.

The transfer lexicon comprises all lexeme-dependent translation information. As for the verbs in case of lexical ambiguity, the equivalents are listed in an order suitable for selecting the appropriate one by checking

- (1) characteristics of valency frames,
- (2) semantic features of verb complements or
- (3) other conditions.

ad 1) E.g., the verb eintreten may be disambiguated by the prepositions:

eintreten für - vystupatj za
'to stand up for';
eintreten in - vstupatj v
'to join sth.'

ad 2) One or more semantic features are attributed to nouns. They may be used for disambiguation, e.g.:

abschließen (accusative inanimated).

-> zapidatj 'to lock'

abschließen (accusative process)

-> zakanchivatj 'to terminate'

ad 3) Concrete noun lexemes can also be used as fillers of verb slots, e.g.

abschließen (accusative Vertrag

'contract')

-> zakluchatj (dogovor)

'to conclude'.

The form of lexical entries allows to adapt them as GRACOLI rules. With the help of the transfer lexicon lexical and valency frame transfer is managed simultaneously. Passive sentences get a special pre-processing before the lexicon is consulted. After this, two distinguished GRACOLI procedures for active and passive voice select the aspect form according to a hierarchy of criteria. Three groups of such criteria are distinguished:

- (1) a lexical criterion,
- (2) syntactic-semantic criteria and
- (3) German tense criteria.

ad 1) For certain German verbs, the transfer lexicon specifies which of the two aspects of the Russian equivalent has to be chosen, e.g.

Prüfung bestehen

-> sdatj äkzamen (perfective aspect)

'to pass an examination'

sich bemühen um

-> dobivatjsja (imperfective aspect)

'to try hard'.

ad 2) In some cases verbal aspect depends on the valency frame of the German verb or/and on semantic features of complements:

Er schrieb an einem Buch

-> On pisal knigu 'he wrote a book'

(preposition an, case dative

--> imperfective aspect).

The cooccurrence of verbs with certain types of adverbials influences the verbal aspect choice, too. Adverbs like jahrelang

'for years', häufig 'often', zunehmend 'more and more' demand the imperfective aspect, but plötzlich 'suddenly', unerwartet 'unexpected', demnächst 'soon' the perfective one. If there occur adverbials of both types within the sentence, the choice of imperfective aspect is preferred.

ad 3) If none of the aforesaid criteria applies, the tense of the finite German verb determines the verbal aspect:

future, perfect, pluperfect

--> perfective aspect

present --> imperfective aspect

In case of preterite, perfect and future semantic verb subclassification as well as definiteness of the direct object are additionally taken into account.

A separate GRACOLI grammar is also used to carry out the structural transfer of modal constructions because the description of the complex conditions of these constructions does not fit in the frame of the lexical entries of the transfer dictionary. This shall be demonstrated by the translation of the German verb können 'can', 'to be able to'. It is necessary to select a limited number of translation patterns because of the wide variety of corresponding Russian constructions. Thus two Russian constructions were taken into consideration. Their choice depends mainly on the kind of the subject and the voice of the infinitive.

(1) The impersonal construction with the predicative mozhno 'it is possible' is selected if one of the following conditions to the subject is fulfilled:

- indefinite-personal pronoun man or
- a nounphrase which belongs to the subclass 'process', or
- a formal es, or
- there is no subject,

for example:

Es kann (ihm) geholfen werden.

'It is possible to help him.'

-> Emu možhno pomochi.

(2) The personal construction with a fi-

nite form of the modal verb mochi 'can', 'to be able to' is preferred if:

- none of the above given conditions is applicable, and if

- the subject is a personal one,

for example:

Er kann die Aufgabe lösen

'He is able to solve the task'

-> On možhet reshitj zadachu.

SYNTHESIS

The transfer result is a constituent structure containing the Russian equivalents and information for syntactic and morphological synthesis. Since the valency frame transfer is already done, only some tasks remain for syntactic synthesis. The Russian verbs must be expanded if they demand analytical forms, that holds for future, subjunctive and passive constructions. Adverbial groups which are handled as only one constituent will get their explicit form now. Gender agreement for Russian preterite and reflexive forms is formed here. Furthermore, Russian word order processing is done.

Input of the morphological synthesis are the results of the previous steps. Information of the Russian dictionary entries is combined with the inflectional lists.

The experimental system VIRTEX proposes a solution for the aspect problem, and is therefore not only a new one in the long list of existing MT systems, but a substantial contribution to MT research.

VIRTEX will be demonstrated at COLING '90.