

LIMITS OF A SENTENCE BASED PROCEDURAL APPROACH FOR ASPECT CHOICE IN GERMAN-RUSSIAN MT

Bianka BUSCHBECK, Renate HENSCHER, Iris HÖSER, Gerda KLIMONOW, Andreas KÜSTNER, Ingrid STARKE

Zentralinstitut für Sprachwissenschaft, Berlin
Prenzlauer Promenade 149-152
O-1100 Berlin

ABSTRACT

In this paper we discuss some problems arising in German-Russian Machine Translation with regard to tense and aspect. Since the formal category of aspect is missing in German the information required for generating Russian aspect forms has to be extracted from different representation levels. A sentence based procedure for aspect choice in the MT system VIRTEX is presented which takes lexical, morphological and semantic criteria into account. The limits of this approach are shown. To overcome these difficulties a human interaction component is proposed.

INTRODUCTION

Aspect is considered to be a grammatico-semantic category for expressing various temporal references in relation to the speech act moment. Regardless of the great number of special meanings that can be expressed by the perfective or imperfective aspect (p.asp./i.asp.), there are two oppositions representing the systematic or basic aspectual meanings, namely +TOTALITY/+LIMITEDNESS versus -TOTALITY/-LIMITEDNESS (see Bondarko 1990).

In this paper we will discuss the transfer of tense and aspect, a problem which arises immediately in Machine Translation and differs from language pair to language pair. This mainly depends on how aspect is expressed in the particular languages concerned.

It is obvious that aspect in several languages has a rather heterogeneous formal reflection in the verb system. Aspect and tense are closely connected with each other. In English, e.g., the two aspect constructions perfective and progressive can be seen as realizing the basic contrast of the action viewed as complete or as incomplete (for details see van Eynde 1988).

All Slavic languages on the other hand have a well-formed aspect system where verbs have a perfective and an imperfective aspect derived from the verbal stem by affixation. The translation of verbal groups from English into Russian, for example, seems to be possible by formulating rules which assign concrete Russian aspect forms to several combinations of tense and aspect in English, e.g.

has been giving (present perfect continuous)

→ *давал* (past, imperfective aspect)

has given (present perfect)

→ *дал* (past, perfective aspect)

(cf. Apresjan 1989: 154).

In contrast to the languages mentioned above, aspect meaning in German, which doubtlessly exists, has no explicit formal expression. Therefore, aspect information required for translation into Russian has to be extracted from different levels of text representation. This is necessary since without the correct choice of Russian aspect serious translation errors in the target language could occur. In our German-Russian MT project VIRTEX we have approached this problem by constructing a hierarchic procedure for aspect choice (presented in the next paragraph) which takes a complex of contextual, morphological and semantic criteria into account. If the aspect choice algorithm fails to select one of the two aspect forms, wider context (beyond the boundaries of sentence) or background knowledge must be taken into consideration. To meet this difficulty VIRTEX is provided with a system of inquiries. If necessary, human interaction is entered to make a final decision (in the sense of Personal MT, see Boitet 1990). A more perfect solution can only be reached by a more sophisticated text and knowledge representation including aspectual characteristics.

**A SET OF FORMAL CRITERIA
USED BY VIRTEX
FOR DETERMINING ASPECT AND TENSE**

The MT system VIRTEX is made to translate simple German main clauses into Russian including the decision of appropriate aspect forms for simple and complex verbal groups. We distinguish five different types of criteria all of them operating on the level of a syntactic surface structure enriched by semantic features:

1. Lexical Information

German verbs which in every context denote non-resultative activities are always translated by a Russian verb in imperfective aspect form, e.g. *arbeiten* 'to work' --> *работать*.

A contrasting class of verbs (*siegen* 'to win', *erreichen* 'to achieve') which represents achievements (see Vendler 1967) can be translated in an analogous way into perfective aspect forms unless the context suggests iterativity.

2. Valency Frames

Some verbs allow different readings concerning their semantics. These may be distinguished by the occurrence of certain verbal complements:

(a) *Er schrieb an einem Brief.*

'He was writing a letter.'

--> *Он писал письмо.* (i.asp.)

(b) *Er schrieb einen Brief.*

'He was writing/wrote/has written a letter.'

--> *Он писал/написал письмо.*

(both aspect forms are possible)

Furthermore, there are German verbs which include several sememes differing with regard to their terminative/aterminative usage (cf. Mehlig 1988). Such a verb is, e. g., the verb *sprechen* 'to speak'. For translating the terminative reading of the verb - *sprechen mit jmdm.* 'to talk with sb.' - in Russian both aspect forms can be used: *говорить/поговорить с кем*. The aterminative reading of *sprechen* does not occur in connection with the preposition *mit* 'with'. In Russian the imperfective aspect must be chosen:

Er sprach (vor Studenten) über Werkstoffe.

'He spoke (to the students) about materials.'

--> *Он говорил (*поговорил) (перед студентами) о материалах.*

Such temporal distinctions of verb readings make it to some extent possible to choose the appropriate aspect form already with the help of the dictionary only.

3. Adverbial Semantics

Various types of adverbials may help to arrive at a decision. In cooccurrence with durative, iterative or intensity adverbials (e.g. *den ganzen Tag lang* 'all day long', *häufig* 'frequently', *mehr und mehr* 'more and more'), the imperfective aspect is chosen. If there are adverbials of punctual meaning (*plötzlich* 'suddenly', date, time) or of future events (*demnächst* 'soon') and no adverbial of the former class, the perfective aspect is preferred. Within the aspect choice algorithm (see fig. 1) these two classes of adverbs were named ADV-I and ADV-P.

4. Tense

If none of the aforesaid criteria applies some German tenses determine the aspect choice:

- Past perfect is translated to perfective aspect form,
- in the case of the present tense (praesens futuri excluded) the imperfective aspect is preferred.
- Future perfect is translated into future using the perfective aspect if there is no indicator of subjunctive meaning which is expressed in Russian by the preterite form *an* and insertion of *вероятно* 'probably' (see the symbol PRT+VEROJATNO in fig.1).

5. Aktionsart Type and Additional Conditions

In the case of the remaining tense forms (not listed in 4.), choice of aspect depends on the verbal semantics. There are distinctions between durative verbs (*warten* 'to wait', *diskutieren* 'to discuss'), verbs with a resultative meaning (*erhöhen* 'to raise', *definieren* 'to define') or verbs such as *aufzählen* 'to enumerate', *produzieren* 'to produce', which are characterized by such properties as limitedness, repeatability, general factitive meaning, named LIM+ITER in fig. 1. In these cases the existence of a direct object, its number and definiteness (N4 PLUR, N4 DET in fig. 1) must be taken into consideration.

For details see figure 1 showing the aspect choice algorithm for active voice sentences implemented in VIRTEX. Some of the strict decisions in this algorithm are preferential ones as will be discussed in the next paragraph. In the case of the passive voice or of modal constructions, different sequences of conditions are

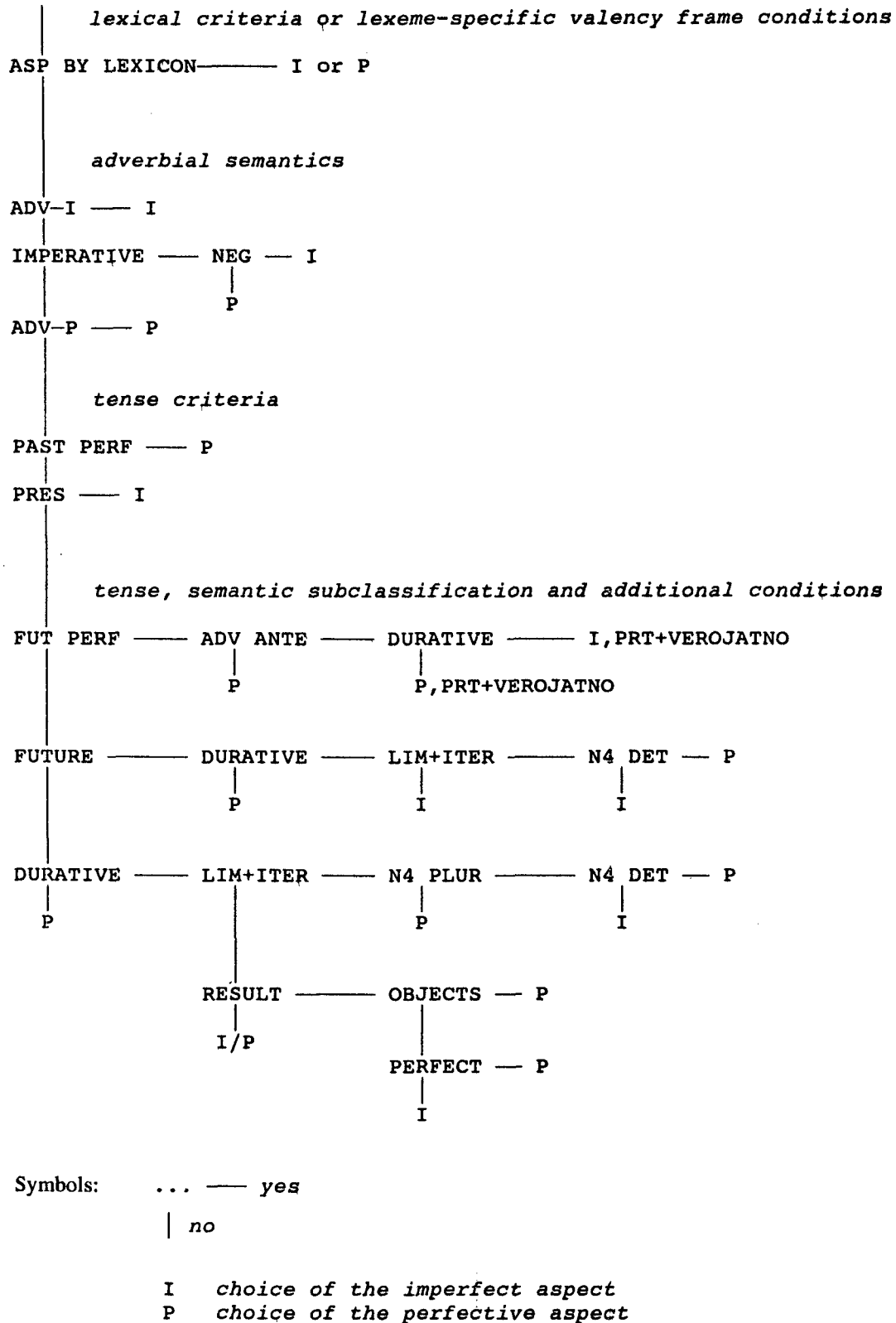


Figure 1. The VIRTEX aspect choice algorithm for active voice

checked in combination with the operations of passive to active transformation (if necessary) or structural transfer for certain modal constructions.

THE ROLE OF CONTEXT

When translating isolated sentences into Russian the absence of information about how to interpret the verbal meaning from an aspectual point of view causes major problems. Often the sentence is too short to find indicators allowing for a decision between several possible interpretations (cf. Somers 1990) which would lead to different results of aspect choice. In such cases it is obvious that by using formal criteria an unambiguous solution is not possible. In other words: the rigid aspect choice algorithm implemented in VIRTEX at first compelled us to make preferential decisions although we have been aware of the fact that sometimes another interpretation of the sentence to be translated would not be captured.

In the following we shall show with five examples how certain contexts help us to clarify the intended interpretation of the given sentence in order to choose the proper aspect form. Here the term 'context' refers to what is expressed in the text surrounding the sentence to be translated or to the user's background knowledge about the text. As long as this kind of knowledge is not accessible, it shall be introduced by means of a dialogue component.

Current Process / Result

(1) *Der Student schrieb einen Brief.*

(1a) *Студент написал письмо.* (p.asp.)
'The student wrote/has written a letter.'

(1b) *Студент писал письмо.* (i.asp.)
'The student was writing a letter.'

In the first version of VIRTEX designed without a user dialogue we preferred the interpretation by which the denoted action is assumed to be completed and consequently the perfective aspect is chosen (see (1a)). For verifying this reading a suitable context criterion could be, e. g., whether another action follows (sequence of predicates): "*Der Student schrieb einen Brief. Danach brachte er ihn zur Post.*" 'The student wrote a letter. After that he took it to the post office.'

Variant (1b) is a good translation if the sentence can be related to a parallel situation or to an action going on simultaneously: "*Es war spät am Abend. Der Student schrieb einen Brief.*" 'It was late in the evening. The student was writing a letter.'

To solve this ambiguity by dialogue the user should be asked whether a continuous process or a completed action is meant. This may be done by inserting an adverb into the sentence and asking the user whether the meaning remains unchanged. The following question should be asked: "*Ist der Satz so gemeint: 'Der Student schrieb gerade einen Brief'?* (j/n)" 'Does the sentence mean: The student was just writing a letter? (y/n)'. If the user says no, reading (1b) is excluded.

Praesens Futuri / Habitual Action

Depending on context, German present tense can be used to express future events. That holds for every kind of verb. Indicators like adverbs help in recognizing the future meaning ("*Er kommt morgen.*" 'He will come tomorrow'). Even if the sentence lacks such adverbs, a future interpretation may be possible but we neglect this fact for the time being. Only if the German sentence contains an achievement verb (the achievement verbs form a subclass of the non-durative ones), the future interpretation seems to have a higher probability because this class of verbs cannot be used to denote a currently ongoing action:

(2) *Er löst die Aufgaben rechtzeitig.*

(2a) *Он решает задачи во-время.* (p. asp.)
'He will solve the tasks in time.'

(2b) *Он решает задачи во-время.* (i. asp.)
'He solves the tasks in time.'

An indicator for the praesens futuri interpretation leading to the translation (2a) would be a context like "*Morgen muß der Student die Arbeit abgeben. Ich bin sicher: Er löst die Aufgaben rechtzeitig.*" 'Tomorrow the student has to submit the paper. I am sure: he will solve the tasks in time.' In this case the perfective aspect is necessary. But it is also possible to assign the sentence an iterative/habitual interpretation leading to sentence (2b). Then we have in mind rather a certain property than a concrete action of the person specified in the subject position. A context suggesting this reading could be a characterization of the student.

To test whether this reading is meant the user is invited to compare the original sentence with "Er löst die Aufgaben in der Regel rechtzeitig." 'As a rule he solves the tasks in time.' If the insertion is possible without changing the sentence meaning, the imperfective aspect of the verb will be chosen, otherwise we assume that the future interpretation holds, which is expressed by the perfective aspect.

Type / Token

Another class of verbs (such as *herstellen* 'to produce', *exportieren* 'to export', *verkaufen* 'to sell') causes a type of ambiguity as shown in (3):

- (3) *Der Trabant wurde in der DDR verkauft.*
 (3a) *Трабант был продан в ГДР.* (p.asp.)
 'The Trabant car was sold in the GDR.'
 (3b) *Трабант продавался в ГДР.* (i.asp.)
 'The Trabant car was sold in the GDR.'

In a context like "Außerhalb des Landes stieß der Trabant auf Absatzschwierigkeiten." 'Abroad the Trabant car met with sales resistance.' sentence (3) describes a frequentative process. In another context a single event of *verkaufen* 'to sell' could be meant: "Die Polizei befaßt sich noch immer mit dem Unfallauto. Es ist jetzt sicher: Der Trabant wurde in der DDR verkauft." 'The police is still investigating the car damaged in the accident. Now it is clear: the Trabant car was sold in the GDR.'

You may observe in our example that the aspectual ambiguity is interrelated with an ambiguity of the semantic object: whereas in the first reading it refers to a set of objects, *Trabant* is type, in the second reading it denotes one concrete individual - *Trabant* is token. The distinction between type and token requires deeper semantic analysis which is impossible without contextual knowledge.

In order to avoid the terms 'type' and 'token' within the dialogue, two sentences are offered to the user. He must decide which of them is more suitable to be used as a paraphrase of the original sentence. With our example, he must select between "Dieses Objekt wurde in der DDR verkauft" 'This object was sold in the GDR' and "Die Objekte wurden in der DDR verkauft" 'The objects were sold in the GDR'. If the user prefers the first paraphrase, the Russian perfective aspect will be used, otherwise the imperfective one.

General Factitive Meaning / Concrete Action

- (4) *Er hat Pläne ausgearbeitet.*
 (4a) *Он разрабатывал планы.* (i.asp.)
 'He has worked out plans.'
 (4b) *Он разработал планы.* (p.asp.)
 'He has worked out plans.'

The imperfective meaning (see (4a)) is inherent in the source sentence when it is interpreted in the following way: a person has gained some experience in working out plans, maybe it was his professional task. Such a translation underlines the general factitive meaning which can be emphasized by using the adverbials *irgendwann einmal*, *eine Zeitlang* 'some time (during his life)': "Er hat irgendwann einmal / eine Zeitlang Pläne ausgearbeitet." 'Some time he worked out plans.' This is the preferred reading in the VIRTEX aspect choice algorithm.

Nevertheless, the sentence also can suggest a concrete, completed action, e. g., if the context refers to the result of this action as in "Er hat Pläne ausgearbeitet. Sie liegen zur Ansicht aus." 'He has elaborated plans. They are open to inspection.' In this case the translation must use the perfective aspect.

To test which of the two readings is the appropriate one, the system offers a sentence with the inserted adverbs as mentioned above, and the user is requested to compare its meaning with that of the sentence to be translated.

The preference of (4a) to (4b) assumed by VIRTEX would be the converse if the direct object were definite.

Further types of aspectual ambiguity may occur. In addition, within one aspect form it may become necessary to resolve temporal ambiguities, e.g.:

Future Perfect / Subjunctive Meaning

- (5) *Der Student wird die Prüfung abgelegt haben.*
 (5a) *Студент сдаст экзамен.*
 'The student will have passed the exam.'
 (5b) *Студент, вероятно, сдаст экзамен.*
 'The student probably passed the exam.'

Sentences (5a) and (5b) exemplify that future perfect in German does not only express future events but more often expresses a presumption with regard to events, actions, etc. which took place in the past. The latter interpretation could be indicated by adverbs which

semantically contradict the future interpretation. These are adverbs of anteriority denoting spans or points of time in the past such as *gestern* 'yesterday', *eben / gerade* 'just' or *letztes Jahr* 'last year'. In this case the choice of the proper aspect form depends on the semantic subclass of the associated verb. For non-durative verbs the perfective aspect must be chosen, for durative verbs - the imperfective one. On the other hand, adverbs of posteriority underline the future tense interpretation. Without such adverbials the sentence remains ambiguous. Adverbs of simultaneity and those deictic adverbs which can express simultaneity as well as anteriority and posteriority do not contribute to disambiguating future perfect sentences because they allow for both interpretations.

To solve the ambiguity in example (5) the inquiry might be: "*Nehmen Sie an, daß das bereits erfolgt ist?*" 'Do you think that it already happened?'

When formulating the inquiries of the dialogue component, we followed the principle that the questions to be answered by the user should be made as precise and simple as possible and should not presuppose any special knowledge in linguistics.

CONCLUSIONS

The above examples show the necessity of taking wider context into account if the sentences are too short to make a wellfounded choice of aspect and tense. As a preliminary solution the integration of inquiries into the system was proposed. For practical use such inquiries may be very helpful because they allow us to improve the translation of isolated sentences and, moreover, of sentences taken from texts. Nevertheless, from the linguistic point of view there has to be further investigation in the field of semantics for the automatic generation of the appropriate aspect forms.

In future we plan to treat aspect and tense by expressing them in a deep semantic representation. This forces us to include wider context beyond sentence boundaries or extralinguistic knowledge, e.g. style and text typology. This can be done either in an interactive way as proposed in this paper or by means of knowledge based MT.

REFERENCES

- Apresjan, Juri D. et al. 1989: *Linguističeskoe obespečenie sistemy ETAP 2*. Moskva.
- Boitet, Christian 1990: *Towards Personal MT: general design, dialogue structure, potential role of speech*. In: Proceedings of COLING-90, Helsinki, Vol.3:30-35.
- Bondarko, Aleksandr V. 1990: *O značenijach vidov russkogo glagola* ('Aspect Meanings of Russian Verbs'). In: *Voprosy jazykoznanija*, No. 4:5-24.
- Buschbeck, B., R. Henschel, I. Höser, G. Klimonow, A. Küstner, and I. Starke 1990: *VIRTEX - a German-Russian Translation Experiment*. Proceedings of COLING-90, Helsinki, Vol.3:321-323.
- Mehlig, Hans R. 1988: *Verbalaspekt und Determination*. In: *Slavistische Beiträge*, München, Vol.230:245-296.
- Somers, Harold L. 1990: *Current Research in Machine Translation*. In: The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language, 11-13 June 1990, University of Texas, Austin.
- van Eynde, Frank 1988: *The Analysis of Tense and Aspect in Eurotra*. In: Proceedings of COLING-88, Budapest, Vol.2:699-704.
- Vendler, Zeno 1967: *Linguistics in Philosophy*. Cornell University Press, Ithaca, N.Y., 97-121.