

Types of Paraphrase Rules in Practice. German Paraphrases of a Russian Text

Robert Zangenfeind

Centrum für Informations- und Sprachverarbeitung (CIS)
Ludwig-Maximilians-Universität
Oettingenstraße 67
80538 München
Germany
R.Zangenfeind@lmu.de

Abstract

Paraphrases can be described at various levels of utterance representation in the Meaning-Text Model (MTM). We analyze different kinds of paraphrases that are found in a parallel corpus of German translations of a Russian novel and assign them to the different representation levels. Doing so, some rules of the deep syntactic paraphrasing system of MTM are modified. For paraphrases that need a deeper analysis to be recognized as such a division into two groups is proposed: paraphrasings that can be described at the semantic level and those that can be described at the transition between the semantic and the deep syntactic level. Some difficulties in the formal recognition of paraphrases will be discussed and the frequency of different types of paraphrase rules will be evaluated.

1 Introduction

Paraphrasing and synonymy of natural language texts can be observed at every level of utterance representation in the MTM. There is synonymy which can be described at the level of morphological representation,¹ of surface syntactic representation, of deep syntactic and of semantic representation, cf. (Apresjan & Cinman, 2002:104ff.) and (Milićević, 2007a:128–138).

In this paper we will examine to what extent paraphrase rules at these different levels are applied in practice, i.e. what kind of paraphrases are to be found when analyzing real texts. As a basis for this research we need a parallel corpus of texts having the same meaning. Such a corpus can be found e.g. in different translations of an original text which are especially available of classical novels. Thus, the original text represents a given meaning which is then expressed in different translations that represent paraphrases of each other. (Wirth, 1996) provides such a corpus by 22 German translations of the first ten sentences of Lev N. Tolstoj's novel *Anna Karenina*. The essential basis for what follows was a comparison of the 21 main predicates of this opening part of the novel, i.e. predicative nouns, adjectives and adverbs in support verb constructions, full verbs, and idioms. Some of the paraphrases of these predicates will be presented here. Besides that, paraphrases of some attributes and actants of these predicates and some pragmatic aspects will be considered. All linguistic examples are taken from (Wirth, 1996) unless noted otherwise.

¹ We will not distinguish between the deep and the surface morphological level in this paper because it is not necessary for our purpose.

2 Paraphrases at different levels of representation

A very important part of rules for paraphrasing natural language texts is described by means of lexical functions in the paraphrasing system which operates at the deep syntactic level of the MTM, cf. (Mel'čuk, 1974:149–176; Mel'čuk, 1992; Apresjan, 1974:316–345; Apresjan & Cinman 2002) and section 3 in this paper. Paraphrasing at the semantic level within the MTM has only recently been described, cf. (Milićević, 2005, 2007a, 2007b) and section 4 in this paper. A different kind of paraphrase – which has not been studied so far in detail – cannot be described within one single level but should be described at the transition between the semantic and the deep syntactic level, cf. section 5. Paraphrases at levels of representation that are closer to surface are briefly discussed in section 6. These include paraphrasings of surface syntactic representations which can be described e.g. by means of different realizations of connecting actants with predicates to their key word. Morphological paraphrases are realized by alternative spellings of a certain lexeme.

3 Paraphrases of predicates at the deep syntactic level

3.1 Exact paraphrases of predicates

First of all, we want to examine some paraphrases that can be described by means of lexical functions at the deep syntactic level.

The lexical rule 24 of (Mel'čuk 1992:39) which describes the paraphrase of a full verb into a support verb construction with a predicative adjective, $C_{0(V)} \Leftrightarrow A_1(C_0) \leftarrow \text{II-Oper}_1(A_1(C_0))$, is applied e.g. in the paraphrase of the following predicate:

- (1) a. alle glücklichen Familien *ähneln* [C_0] einander 'all happy families resemble each other'
 $\Leftrightarrow$
b. alle glücklichen Familien *sind* $\text{Oper}_1(A_1(C_0))$ einander *ähnlich* [$A_1(C_0)$] 'all happy families are similar to each other'

This rule and further rules for support verb constructions with predicative nouns as well as rules for synonyms, antonyms, and conversives which are applied quite often in our corpus, are described completely within the paraphrasing system of MTM.

3.2 Approximate paraphrases of predicates

The lexical rule 1 of (Mel'čuk 1992:37), $C_0 \Leftrightarrow \text{Syn}(C_0)$, is applied very often with the modification that not a genuine synonym but a partial synonym is used in the paraphrase, as in (2):

- (2) a. alle glücklichen Familien *ähneln* [C_0] einander 'all happy families resemble each other'
 $\Leftrightarrow$
b. alle glücklichen Familien *gleichen* [$\text{Syn}_\subset(C_0)$] einander 'all happy families equal each other'

The formal notation of this rule could be rule 1':

Rule 1' $C_0 \equiv \text{Syn}_\subset(C_0) \equiv \text{Syn}_\supset(C_0) \equiv \text{Syn}_\cap(C_0)$

This rule can be used, of course, not only for predicates but also for actants and attributes.

There are some rules which are part of the semantic implications in Mel'čuk's (1992) system, namely rules 53 and 54, $\text{Incep}X \Rightarrow X$ and $\text{Cont}X \Rightarrow X$. When the application of rule 53 is combined with a change of tense and rule 54 is varied into rule 54' (cf. below) they describe a semantic implication in both directions and thus can form a class of approximate paraphrases. This can be illustrated in (3) and (4) respectively:

- (3) a. der Mann *war* [C_0] nicht zu Hause 'the husband was not at home'
 $\Leftrightarrow$

- b. der Mann *war* nicht nach Hause *gekommen* [Incep(C₀)] ‘the husband had not come home’

The German verb *sein* ‘be’ in its Imperfekt form *war* ‘was’ can be paraphrased very well into the verb *kommen* ‘come’ which is an Incep(*sein*) when used in its Plusquamperfekt form² *war gekommen* ‘had come’ and the other way round. So, a variation of rule 53 could be rule 53’:

Rule 53’ $C_{0(V, Imperfekt)} \cong \text{Incep}(C_0)_{\text{Plusquamperfekt}}$ ³

The rule 54 could be varied into rule 54’ by adding a temporal adverb on both sides of the rule because the presence of a temporal adverb in both phrases produces fairly good paraphrases in both directions, like in (4).

Rule 54’ $C_{0(V)} [+ \text{Adv}_{\text{temp}}] \cong \text{Cont}(C_0) [+ \text{Adv}_{\text{temp}}]$

- (4) a. Ivan *war* [C₀] drei Tage in Odessa ‘Ivan was in Odessa for three days’
 $\Leftrightarrow$
 b. Ivan *blieb* [Cont(C₀)] drei Tage in Odessa ‘Ivan stayed in Odessa for three days’⁴

In a similar way these approximate rules are applied in our corpus to different kinds of support verb constructions. Examples (5) and (6) may serve as an illustration for this:

- (5) a. die englische Gouvernante *zankte sich* [C₀] ... ‘the English governess quarreled ...’
 $\Leftrightarrow$
 b. die Erzieherin *war in Zank* [S₀(C₀)] *geraten* [IncepOper₁(S₀(C₀))] ... ‘the governess had got into a quarrel ...’⁵
- (6) a. Es war nun schon der dritte Tag, dass diese Situation [C₀] *bestand* [Func₀(C₀)] ‘it was the third day that this situation existed’
 $\Leftrightarrow$
 b. Es war schon der dritte Tag, dass diese gespannte Situation [C₀] *fortdauerte* [ContFunc₀(C₀)] ‘it was the third day that this tense situation continued’

A lot more examples of this kind of deep syntactic paraphrasing are analyzed in (Zangenfeind, 2009).

3.3 Syntactic paraphrase of an attributive adjective into a predicate one

There are also some simply syntactic paraphrases that don’t imply any lexical functions (similar to active – passive transformations) like e.g. in (7) where an attributive adjective is paraphrased into a predicative one in a relative clause:

- (7) a. alle *glücklichen* [C₀] Familien ‘all happy families’
 $\Leftrightarrow$
 b. alle Familien, *die glücklich* [C₀] *sind* [Copul(C₀)] ‘all families that are happy’

² In some contexts the Perfekt form also produces quite acceptable paraphrases.

³ Two points about this rule should be specified: firstly, we are aware that the rule in this form is not universal – in another language this rule must be adopted according to its tense system. In Russian e.g. an appropriate change of tense can be realized by a change from the imperfective aspect to the perfective aspect like in: он *спал* [C₀] $\cong$ он *заснул* [Incep(C₀)] ‘he slept $\cong$ he had fallen asleep’. More research is needed to formulate a more general rule. Nevertheless, even if we don’t have a universal rule now, we have to take into account that a considerable number of paraphrases of this type are found in real texts. Secondly, in practice this rule is sometimes applied even without change of tense which, of course, makes the paraphrase less exact and more approximate.

⁴ This example is not taken from Anna Karenina but contributed by the author.

⁵ In this paraphrase a synonym for this 1st actant is used, *Erzieherin* = Syn(*Gouvernante*) ‘governess’, and the attribute *englisch* ‘English’ is omitted, cf. about this problem below.

The deep syntactic rule for this paraphrase is a general one and can be noted as in figure 1:

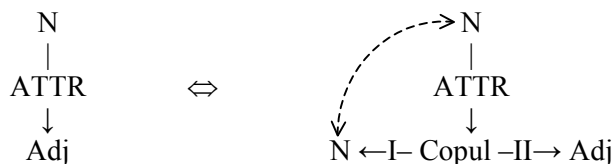


Figure 1. Syntactic rule for paraphrasing an attributive adjective into a predicative one.

It would be possible as well to describe the paraphrase in (7) as a transition rule between the semantic representation and the deep syntactic representation, like some others, as will be shown below.

3.4 Paraphrases of idioms

A lot of paraphrases by means of an idiom can be described at the deep syntactic level because they correspond to a single node of the deep syntactic structure. This is possible e.g. with the paraphrase in (8):

- (8) a. die Frau hatte *erfahren*, dass ... ‘the wife had learned that ...’
 $\Leftrightarrow$
 b. die Frau hatte *Wind bekommen* von ... ‘the wife had got wind of ...’

The idiom *Wind bekommen* ‘get wind’ in (8b) corresponds to the single node *erfahren* ‘learn’ which can be noted in the dictionary as $\text{Syn}(\text{Wind bekommen}) = \text{erfahren}$. However, this kind of notation is not possible for all idioms. An example to illustrate this is given in (9):

- (9) bei jemandem ist Hopfen und Malz verloren ‘*lit.* hops and malt are lost on somebody’⁶

This idiom roughly means ‘it is not worth the trouble trying to help or reform somebody because it seems useless’. This meaning cannot be expressed in German by a single lexeme, and thus cannot be noted comfortably in terms of the lexical function Syn. As a consequence, paraphrases like this must be described at the transition between the semantic and the deep syntactic level, cf. below.

This raises the question whether the paraphrasing of idioms should formally be described in two different ways. Of course, to avoid this, all paraphrasings by an idiom in principle could be described at the transition between the semantic and the deep syntactic level – which would be awkward in all cases where it is possible to note them in terms of a synonym (which is possible in most cases). But that is a problem that cannot be dealt with sufficiently here.

4 Paraphrases at the semantic level

4.1 Approximate paraphrases

A lot of paraphrases in the translations of Anna Karenina are rather complicated and the only reasonable way to describe them could be with rules at the semantic level because they have different semantic representations. These rules are complicated in themselves, cf. some rules in (Milićević, 2007a, 2007b) but what makes them additionally complicated, is the fact that in practice these paraphrases are sometimes only very approximate paraphrases, like the following:

- (10) a. alle glücklichen Familien *gleichen* [C₀] einander ‘all happy families equal each other’
 $\Leftrightarrow$

⁶ Example by the author.

- b. *die äußeren Formen*, in denen das sogenannte glückliche Familienleben sich abzuspielen pflegt, *sind* überall *die gleichen* ‘the external forms in which the so called happy family life usually takes place are the same everywhere’

To describe paraphrases like these in a formal way and recognize them automatically seems hardly possible because, other than in paraphrases which Milićević (2007a) treats as propositional quasi-equivalences, the semantic relation between paraphrases like in (10) is quite irregular. Of course, one may ask whether (10b) is really a well-formed paraphrase of (10a) (which is a good translation of the Russian original phrase) but, nevertheless, it is a paraphrase that is found in a real text.

4.2 Paraphrases concerning pragmatics

In (11b) an introductory expression attracts the attention of the reader in a pragmatic way without giving additional meaning to the sentence concerning the extralinguistic situation. Nevertheless, the two phrases of (11) have different semantic representations because of the speaker’s attitude that is communicated in (11b) and must be described in the modal frame⁷ which is roughly ‘P, and the speaker thinks that P is important’:

- (11) a. die Frau des Hauses hatte erklärt ... ‘the hostess had announced ...’
 ⇔
 b. *die Sache war die*, dass die Frau des Hauses erklärt hatte ... ‘the point was that the hostess had announced ...’

In the phrases of (12) a change of communicative organization takes place by means of a change of word order. This does not affect the semantic structure of the phrases but only the distribution of the semantic theme and rheme, cf. (Mel’čuk, 2001:61f.):

- (12) a. im Hause der Oblonskijs ging alles drunter und drüber ‘at Oblonskijs’ house everything was at sixes and sevens’
 ⇔
 b. alles ging drunter und drüber im Hause der Oblonskijs ‘everything was at sixes and sevens at Oblonskijs’ house’

Another change of pragmatics occurs in the paraphrase of (13) by the addition of the particle *mehr* ‘more’ which must be described again in the modal frame of the semantic representation:

- (13) a. der Mann war seit drei Tagen nicht nach Hause gekommen ‘the husband had not come home for three days’
 ⇔
 b. der Mann war seit drei Tagen nicht *mehr* nach Hause gekommen ‘the husband had not come home again for three days’

The phrase (13a) is more neutral than (13b) which includes a presupposition of the speaker who thinks that three days is a lot. The propositional meaning that is described in both phrases, i.e. ‘the man was not at home for three days’, is not affected by this.

5 Paraphrases of predicates at the transition between the semantic and the deep syntactic level

There are some paraphrases which cannot be described sufficiently at one single level of representation. The reason for this is that on the one hand they have different deep syntactic representations but (more or

⁷ Cf. e.g. Apresjan (1974:67).

less) the same semantic representation,⁸ which means that no paraphrase at the semantic level is necessarily required. On the other hand their different deep syntactic representations cannot be paraphrased into each other by means of lexical functions because they do not belong to the same so called basic deep syntactic structure (базовая ГСС, cf. Mel'čuk, 1974:177).⁹ The only way how phrases of this kind can be recognized as paraphrases is to analyze them down to their semantic representation. This kind of paraphrases has not yet been discussed in detail, as far as we know.

The two phrases in (14) also have more or less the same semantic representation (maybe with a hint to slightly different stylistic specifications in the rhetorical structure). Their paraphrasing implies a substitution of a predicate which at first glance could be expressed by a simple lexical rule with a lexical function, namely $C_{0(V)} \Leftrightarrow S_0(C_0)$. Nevertheless, this paraphrase is more complex and should be formalized at the transition between the semantic and the deep syntactic level when the nodes of the semantic net are replaced by deep syntactic units. This is because it is rather complicated to describe the relation of the semantic actants of the different predicates involved in the paraphrase:

- (14) a. die englische Erzieherin hatte einer Freundin einen Brief geschrieben, *in dem sie dieselbe bat* [C_0], ihr eine neue Stelle zu besorgen '*lit.* the English governess had written a letter to a friend in which she asked her to get her a new job'
- $\Leftrightarrow$
- b. die englische Gouvernante hatte einer Freundin einen Brief geschrieben *mit der Bitte* [$S_0(C_0)$], ihr eine neue Stellung zu verschaffen '*lit.* the English governess had written a letter to a friend with the request to get her a new job'

The 1st semantic actant of *Brief* 'letter' (*die englische Erzieherin/Gouvernante* 'the English governess') is realized in both phrases as the syntactic actant I of *schreiben* 'write' which serves as Oper₁(*Brief*). This actant is also the 1st semantic actant of *bitten* 'ask' in (14a) and of *Bitte* 'request' in (14b). But, while this is explicated by the pronoun *sie* 'she' in (14a), it is only implicitly known in (14b). Similarly, the 2nd semantic actant of the request (*Freundin* 'friend') is explicitly realized in (14a) by the pronoun *dieselbe* 'the same' and only implicitly known in (14b). These relations are clear in the semantic representation, and the way they can be realized or omitted in the text should be formally noted in a rule of the semantic component of the MTM, i.e. at the transition from the semantic to the deep syntactic representation. This rule must describe the following fact: When some kind of information (like e.g. a request, a promise, news) is transported by a "bearer/medium of the information" (which can be e.g. a letter, a telegram or an agreement) the 1st actant of the information is identical with the 1st actant of the medium of the information. The same applies to the counter agent of the information which is identical in such cases with the counter agent of the medium of the information.

The paraphrase in (15) shows two phrases with the same request as in (14); in (15a) the request is explicitly expressed by means of the full verb *bitten* 'ask', similarly to (14a), in (15b) it is expressed by the modal verb *mögen* 'may' in a slightly different syntactic construction:

- (15) a. die englische Erzieherin hatte einer Freundin geschrieben, sie *bitte* [C_0] sie, sich nach einer anderen Stellung für sie umzutun '*lit.* the English governess had written to a friend she would ask her to look for a new job for her'
- $\Leftrightarrow$
- b. die englische Erzieherin hatte an eine Freundin geschrieben, sie *möchte* sich nach einer anderen Stelle für sie umsehen '*lit.* the English governess had written to a friend she [the friend] may look for a new job for her'

The use of *mögen* 'may' in this context of a request should also be described as a rule of the semantic component of the MTM because a paraphrase of *bitten* 'ask' by *mögen* 'may' is not possible in every con-

⁸ In this respect they are similar to conversives which have the same semantic structure but different deep syntactic structures (and different communicative structures), cf. (Mel'čuk, 2001:120f.).

⁹ In this respect they are different from conversives which do belong to the same basic deep syntactic structure.

text. In a phrase like (16) where the 3rd actant of the request is not realized by a verbal expression it would produce an ungrammatical sentence:

- (16) a. die Erzieherin hatte einer Freundin geschrieben, sie *bitte* den Hausherrn um eine Stelle für Ivan '*lit.* the governess had written to a friend she would ask the landlord for a job for Ivan',¹⁰
 ⇔
 b. *die Erzieherin hatte einer Freundin geschrieben, der Hausherr *möchte* eine Stelle für Ivan '**lit.* the governess had written to a friend the landlord may a job for Ivan'

6 Paraphrases at the surface syntactic level and at the morphological level

In the government pattern of predicates different forms of connecting actants to a keyword are described. This allows the formal description of paraphrases at the surface syntactic level, like e.g. in (17). Unlike the paraphrases above this is not a paraphrase of the predicate itself:

- (17) a. die englische Erzieherin hatte *einer Freundin* geschrieben 'the English governess had written a friend'
 ⇔
 b. die englische Erzieherin hatte *an eine Freundin* geschrieben 'the English governess had written to a friend'

Here the verb *schreiben* 'write' is used as a full verb. In its government pattern the following two possibilities for a connection of its 3rd actant, the counter agent, are described: a noun in the dative case and a noun in the accusative case via the preposition *an* 'to' (or formally: D3.1 N_{dat} and D3.2 *an* N_{acc}).¹¹

The change of word order in (18) is a paraphrase without change of communicative organization at the semantic level, other than in (12). The two phrases of (18) have identical surface syntactic structures but different morphological structures and thus can be described formally at the transition between the surface syntactic and the morphological level.¹²

- (18) a. die Hausfrau hatte erklärt, sie könne nicht länger *mit ihm unter einem Dache* wohnen 'the hostess had announced that she could not live any longer with him under one roof'
 ⇔
 b. die Frau hatte erklärt, sie könne nicht mehr *unter einem Dache mit ihm* wohnen 'the hostess had announced that she could not live any longer under one roof with him'

Paraphrases at the morphological level concern e.g. alternative inflections and spellings of lexemes which must be noted in the morphological zone of the dictionary entry; an example of this is shown in (19) where the lexeme *Dach* 'roof' is realized in two different dative forms:

- (19) a. unter einem *Dach* 'under one roof'
 ⇔
 b. unter einem *Dache* 'under one roof'

In principle, paraphrases like these are possible, of course, with predicates as well as with actants and attributes like the one in (19). They can be identified without difficulty because of their formal description in the dictionary.

¹⁰ Example by the author.

¹¹ More types of paraphrases at the surface syntactic level are described in (Milićević, 2007:135f.). All these paraphrases don't concern predicates directly.

¹² (Milićević, 2007:136f.) describes a similar paraphrase with a change of linear word order at the deep morphological level. Nevertheless, as long as there are no explicit rules for paraphrases by means of word order at the deep morphological level, an analysis to the surface syntactic level is necessary to recognize paraphrases like these because the information for different forms of linearization of an utterance is available only in the surface syntactic representation.

A special problem in this context, however, can be caused by the different spellings of transcribed Russian names and other expressions because they often do not have dictionary entries. The name of the family in Tolstoj's novel e.g. has three different spellings in the translations: *Oblonskij*, *Oblonski* and *Oblonsky*.

7 Two difficulties in the recognition of paraphrases

When examining the German paraphrases of Anna Karenina, two difficulties are met quite often: Firstly, in a lot of paraphrases that can be described by lexical rules of the deep syntactic paraphrasing system of the MTM two or even more lexical rules are applied at the same time for one single predicate. Secondly, as already has been stated above, a lot of paraphrasings in practice are only very approximate paraphrases at the semantic level or they are based on only very vague lexical synonyms, some of which are not to be found in the dictionary.

An example for a simple combination of two rules is shown in (20):

- (20) a. alle *fühlten* ... 'everybody felt'
 $\Leftrightarrow$
 b. alle *hatten das Empfinden* ... 'everybody had the emotion'

Here the full verb *fühlen* 'feel' is paraphrased by a support verb construction with an $Oper_1$ and a predicative noun. At the same time the predicative noun S_0 (*fühlen* 'feel') – which is *Gefühl* 'feeling' – is paraphrased by Syn (*Gefühl* 'feeling') which is *Empfinden* 'emotion'.

A quite complex combination of rules in just one phrase that is part of a relative clause in reported speech is shown in (21):

- (21) a. ... dass die in einer Herberge zufällig sich begegnenden Gäste [E1] *mehr* [$C_{0-2,kompar}$] miteinander *verbunden* [$A_1(C_{0-1})$] *seien* [$Oper_1(A_1(S_{0-1}))$] als sie [die Familienmitglieder = E2] '... that the guests that meet accidentally in a hostel are connected more with each other than they [the members of the family] are'
 $\Leftrightarrow$
 b. ... dass *kaum* noch *so viel* [$Anti(C_{0-2,kompar})$] *Gemeinsames* [= $Syn_\cap(C_{0-1})$] zwischen ihnen [E2] *war* [$Func_0(Syn(S_{0-1}))$], wie etwa zwischen den zufälligen Gästen [E1] eines Hotels '... that there was hardly as much common ground between them anymore as, for instance, between the accidental guests of a hotel'

There are two predicates that are essential for a consideration of this paraphrase. In (21a) one predicate is realized by the lexeme *verbunden* 'connected' [= $A_1(C_{0-1})$], an adjective that describes the property of the first actant of the situation C_0 which could be represented by *Verbindung* 'connection'. This adjective is part of a support verb construction with the $Oper_1$ *sein* 'be' in its subjunctive form *seien*. The other predicate is the lexeme *mehr* 'more' [= $C_{0-2,kompar}$], a comparative adverb. The actants of the comparison are *Gäste* 'guests' [= E1] and *sie* 'they' [= E2] (a pronoun that represents the members of the family).

In (21b) the first predicate is realized by a partial synonym of C_0 , namely *Gemeinsames* 'common ground' [= $Syn_\cap(C_{0-1})$] which is used in a support verb construction with $Func_0$. The second predicate is realized now as an antonym by means of *kaum so viel* 'hardly as much as' [$Anti(C_{0-2,kompar})$], i.e. 'less' which entails an exchange of its actants *Gäste* 'guests' and *sie* 'they' in its dative form *ihnen*.

So, the rules for a formal description of the paraphrase of these two predicates are a combination of rules 20 and 24 from (Mel'čuk, 1992), $A_1(C_0) \leftarrow II- Oper_1(A_1(C_0)) \Leftrightarrow S_0(C_0) \leftarrow I- Func_0(S_0(C_0))$, the modified rule 1', $C_0 \equiv Syn_\cap(C_0)$ (cf. above), and a rule from (Apresjan & Cinman, 2002), $E1 + C_{0,kompar} + E2 \Leftrightarrow E2 + Anti_3(C_{0,kompar}) + E1$. Written in one rule this would be (shown here without syntactic relations): $A_1(C_{0-1}) + Oper_1(A_1(C_{0-1})) + E1 + C_{0-2,kompar} + E2 \Leftrightarrow Syn_\cap(C_{0-1}) + Func_0(Syn_\cap(C_{0-1})) + E2 + Anti(C_{0-2}) + E1$.

But the paraphrase of (21) is even more complex, and here we come to the second difficulty of recognizing paraphrases in practice. The first actant, E1, is specified in (21a) by the expression *die in einer Herberge zufällig sich begegnenden [Gäste]* 'the guests that meet accidentally in a hostel' which is

considerably different from the specification in (21b): *die zufälligen [Gäste] eines Hotels* ‘the accidental guests of a hotel’. The lexeme *Herberge* ‘hostel’ does not really mean the same as *Hotel* ‘hotel’ and the attribute *zufällig sich belegend* ‘that meet accidentally’ is more precise than simply *zufällig* ‘accidental’. Furthermore, in (21b) there is a modal frame by the lexeme *noch* ‘anymore’ and the parenthesis *etwa* ‘for instance’ which make the paraphrase still more approximate.

In other paraphrases some actants are even more different from each other than they are in (21). So, in some translations instead of *Gäste* ‘guests’ the very distant synonym *Leute* ‘people’ is used. And there are predicates which are used in paraphrases but are only very vague synonyms, like e.g.:

- (22) a. *Sinn* ‘sense’
 ⇔
 b. *höherer Gedanke* ‘higher idea’
 ⇔
 c. *innerer Zusammenhang* ‘internal connection’
 ⇔
 d. *Boden* ‘ground’

It is fairly obvious that paraphrases like these are really hard to recognize automatically.

In practice, as we have seen in (21), the two discussed difficulties are met quite often at the same time in one single phrase. In fact, it is hard to find phrases in which just one single rule is applied. This makes the practice of recognizing paraphrases in real texts even more complicated.

8 Evaluation of the frequency of rules

When considering the 21 predicates of the first ten sentences in the German translations of Tolstoj’s novel we find about 46% are no paraphrases of each other but use identical predicates – these are the direct translations of the predicates in the Russian original text. About 19% of the predicates are paraphrases by means of synonyms, including idioms (5%), about 14% of the paraphrases can be described by exact rules for support verb constructions, 7% of the paraphrases must be described at the semantic level, 5% at the transition between the semantic and the deep syntactic level, for a little more than 3% of the paraphrases approximate rules for support verb constructions are applied, for a little less than 3% approximate rules for full verbs, and 2% are paraphrasings by an antonym or converse.

Leaving aside direct translations and taking into consideration only paraphrases of these, this leads to the following distribution: Almost 79% of all paraphrases which are found in the 22 translations of Anna Karenina can be described by lexical rules of the deep syntactic paraphrasing system of the MTM. Most interesting is that a big part of this, i.e. more than 32% (including approximate paraphrases) of all rules applied, are rules for support verb constructions – that is almost as much as the 36% paraphrases by synonyms. 12% of the predicates are paraphrased by rules that are applied at the semantic level and 9% can be described by rules applied at the transition between this level and the deep syntactic representation.

These figures show very impressively the importance of the deep syntactic paraphrasing system of the MTM, especially of the rules for support verb constructions. Yet, the considerable numbers of semantic paraphrasings and of those that are described by transition rules show that it is important too, on the other hand, to develop further these parts of the MTM.

Acknowledgements

Many thanks to Franz Guenther for the discussions during the preparation of this paper and to the two anonymous reviewers for their comments on the pre-final version of this paper.

References

Mel’čuk, Igor A. 1992. Paraphrase et lexique. In: Igor A. Mel’čuk et al., *Dictionnaire explicatif et combinatoire du français contemporain: Recherches lexico-sémantiques*, vol. III. Montréal: Presses de l’Univ. de Montréal, 9–58.

- Mel'čuk, Igor A. 2001. *Communicative Organization in Natural Language: The Semantic-Communicative Structure of Sentences*. Amsterdam: John Benjamins.
- Milićević, Jasmina 2005. Semantic Paraphrasing in Meaning-Text Linguistic Models. In: Юрий Д. Апресян (ред.), *Восток – Запад*. Москва: Языки славянской культуры, 316–329.
- Milićević, Jasmina 2007a. *La paraphrase. Modélisation de la paraphrase langagière*. Bern: Peter Lang.
- Milićević, Jasmina 2007b. Semantic Equivalence Rules in Meaning-Text Paraphrasing. In: Leo Wanner (ed.), *Selected Lexical and Grammatical Issues in the Meaning-Text Theory*. Amsterdam: John Benjamins, 267–296.
- Wirth, Dieter 1996. *Paraphrase und Übersetzung in einem Inhalt ↔ Text-Modell*. Tübingen: Niemeyer.
- Zangenfeind, Robert 2009. *Das Bedeutung-Text-Modell und dessen Paraphrasierungssystem unter besonderer Berücksichtigung von Stützverbgefügen*. [To appear]
- Апресян, Юрий Д. 1974. *Лексическая семантика. Синонимические средства языка*. Москва: Наука.
- Апресян, Юрий Д. & Леонид Л. Цинман 2002. Формальная модель перифразирования предложений для систем переработки текстов на естественных языках. In: *Русский язык в научном освещении*, № 2 (4), 102–146.
- Мельчук, Игорь А. 1974. *Опыт теории лингвистических моделей «Смысл ↔ Текст»*. Семантика, синтаксис. Москва: Наука.